# NUCLEASE-MEDIATED TARGETING WITH LARGE TARGETING VECTORS

## CROSS REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims priority to U.S. Provisional Application No. 61/638,267 filed 25 April 2012, which is incorporated by reference herein in its entirety.

## FIELD

**[0002]** Nucleases and DNA constructs, including targeting vectors (*e.g.*, large targeting vectors, "LTVEC") for achieving homologous recombination at a target genomic locus. Compositions and methods for making a targeted genetic modification via homologous recombination that is facilitated by a single or double-strand break at or near a target genomic locus. Compositions and methods for promoting efficiency of homologous recombination between an LTVEC and a target genomic locus in prokaryotic or eukaryotic cells employing engineered nucleases.

## BACKGROUND

**[0003]** Homologous recombination using targeting vectors that are specifically designed to add, delete, or replace a particular nucleic acid sequence at a genomic locus is a popular approach to achieving a desired genomic modification in non-human animals. A nuclease that is specifically engineered to introduce a singe or a double-strand break at or near a target genomic locus can be used together with a targeting vector to enhance efficiency of homologous recombination at the target genomic locus.

**[0004]** Although the art of genome modification through homologous recombination has advanced considerably over the last two decades, difficulties still remain with achieving an acceptable targeting frequency using very large targeting vectors, LTVECs, in many circumstances, for example, when a large portion of a rodent genome is replaced with a large human genomic fragment, or difficulties targeting certain cell types, *e.g.*, fibroblasts or other somatic cells. There is a need in the art for further and improved methods for modifying large genomic loci of a eukaryotic genome using LTVECs.

## SUMMARY

**[0005]** Compositions and methods are provided for modifying a genomic locus of interest using a large targeting vector (LTVEC) in combination with a nuclease agent, which allow efficient deletion, addition (*e.g.*, insertion), and/or replacement of a large nucleic acid sequence at the genomic locus of interest.

**[0006]** Compositions and methods are provided for modifying a target genomic locus of

a mammal in a prokaryotic cell using an LTVEC and a nuclease agent via bacterial homologous recombination (BHR), wherein the BHR is facilitated by a single or double-strand cleavage at or near the target genomic locus created by the nuclease agent. Prokaryotic cells are provided comprising an LTVEC and a nuclease agent that, upon expression, is capable of introducing a single or double-strand cleavage at or near a target site. Compositions and methods are provided for replacing a large genomic locus of a non-human animal with an exogenous nucleic acid sequence, *e.g.*, homologous or orthologous human genomic nucleic acid sequences, in a recombinogenic prokaryotic cell by employing various LTVECs and nucleases as described herein.

**[0007]** Compositions and methods are provided for modifying a genomic locus of interest using an LTVEC and a nuclease agent in various pluripotent mammalian cells. Compositions and methods are provided for replacing a large genomic locus of a non-human animal with exogenous nucleic acid sequence, *e.g.,* homologous or orthologous human genomic nucleic acid sequences, in a pluripotent cell of the non-human animal by employing various LTVECs and nucleases as described herein.

**[0008]** Pluripotent cells of a non-human animal are provided comprising various LTVECs and nuclease agents described herein.

**[0009]** Compositions and methods are provided for generating a genetically modified non-human animal comprising one or more targeted genetic modifications as described herein.

**[00010]** In one aspect, a prokaryotic cell is provided, comprising a large targeting vector (LTVEC) comprising homology arms directed to a target locus, and a nucleic acid sequence encoding a nuclease agent that makes a single- or double-strand break at or near the target locus.

**[00011]** In one embodiment, the prokaryotic cell is capable of expressing a recombinase that mediates bacterial homologous recombination (BHR). In one embodiment, the prokaryotic cell is a recombination-competent strain of *E. coli*.

**[00012]** In one embodiment, the LTVEC ranges from about 50kb to about 300kb. In one embodiment, the LTVEC ranges from about 50kb to about 75kb. In one embodiment, the LTVEC ranges from about 75kb to about 100kb. In one embodiment, the LTVEC ranges from about 100kb to 125kb. In one embodiment, the LTVEC ranges from about 125kb to about 150kb. In one embodiment, the LTVEC ranges from about 150kb to about 175kb. In one embodiment, the LTVEC ranges from about 175kb to about 200kb. In one embodiment,

the LTVEC ranges from about 200kb to about 225kb. In one embodiment, the LTVEC ranges from about 225kb to about 250kb. In one embodiment, the LTVEC ranges from about 250kb to about 275kb. In one embodiment, the LTVEC ranges from about 275kb to about 300kb.

**[00013]** In one embodiment, the homology arms of the targeting vector are derived from a BAC library, a cosmid library, or a P1 phage library. In one embodiment, the homology arms are derived from a genomic locus of the non-human animal that is not targetable using a conventional method. In one embodiment, the homology arms are derived from a synthetic DNA.

**[00014]** In one embodiment, a sum total of the upstream homology arm and the downstream homology arm is at least 10kb. In one embodiment, the upstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the downstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the upstream and the downstream homology arms range from about 5kb to about 10kb. In one embodiment, the upstream and the downstream homology arms range from about 10kb to about 20kb. In one embodiment, the upstream and the downstream homology arms range from about 20kb to about 30kb. In one embodiment, the upstream and the downstream homology arms range from about 30kb to about 40kb. In one embodiment, the upstream and the downstream homology arms range from about 40kb to about 50kb. In one embodiment, the upstream and the downstream homology arms range from about 50kb to about 60kb. In one embodiment, the upstream and the downstream homology arms range from about 60kb to about 70kb. In one embodiment, the upstream and the downstream homology arms range from about 70kb to about 80kb. In one embodiment, the upstream and the downstream homology arms range from about 80kb to about 90kb. In one embodiment, the upstream and the downstream homology arms range from about 90kb to about 100kb. In one embodiment, the upstream and the downstream homology arms range from about 100kb to about 110kb. In one embodiment, the upstream and the downstream homology arms range from about 110kb to about 120kb. In one embodiment, the upstream and the downstream homology arms range from about 120kb to about 130kb. In one embodiment, the upstream and the downstream homology arms range from about 130kb to about 140kb. In one embodiment, the upstream and the downstream homology arms range from about 140kb to about 150kb. In one embodiment, the upstream and the downstream homology arms range from about 150kb to about 160kb. In one embodiment, the upstream and the downstream homology arms range from about 160kb to about 170kb. In one embodiment, the upstream and the downstream homology arms range from about 170kb to about 180kb. In one embodiment, the upstream

and the downstream homology arms range from about 180kb to about 190kb. In one embodiment, the upstream and the downstream homology arms range from about 190kb to about 200kb.

**[00015]** In one embodiment, the LTVEC comprises a selection cassette. In one embodiment, the selection cassette comprises a nucleic acid sequence encoding a selection marker, wherein the nucleic acid sequence is operably linked to a promoter. In one embodiment, the promoter is active in a prokaryotic cell. In one embodiment, the promoter is active both in prokaryotic and eukaryotic cells. In one embodiment, the selection marker is selected from neomycin phosphotransferase ($neo^r$), hygromycin B phosphotransferase ($hyg^r$), puromycin-N-acetyltransferase ($puro^r$), blasticidin S deaminase ($bsr^r$), xanthine/guanine phosphoribosyl transferase (gpt), and herpes simplex virus thymidine kinase (HSV-k), and a combination thereof.

**[00016]** In one embodiment, the LTVEC comprises an insert nucleic acid ranging from about 5kb to about 200kb. In one embodiment, the insert nucleic acid is from about 5kb to about 10kb. In one embodiment, the insert nucleic acid is from about 10kb to about 20kb. In one embodiment, the insert nucleic acid is from about 20kb to about 30kb. In one embodiment, the insert nucleic acid is from about 30kb to about 40kb. In one embodiment, the insert nucleic acid is from about 40kb to about 50kb. In one embodiment, the insert nucleic acid is from about 60kb to about 70kb. In one embodiment, the insert nucleic acid is from about 80kb to about 90kb. In one embodiment, the insert nucleic acid is from about 90kb to about 100kb. In one embodiment, the insert nucleic acid is from about 100kb to about 110kb. In one embodiment, the insert nucleic acid is from about 120kb to about 130kb. In one embodiment, the insert nucleic acid is from about 130kb to about 140kb. In one embodiment, the insert nucleic acid is from about 140kb to about 150kb. In one embodiment, the insert nucleic acid is from about 150kb to about 160kb. In one embodiment, the insert nucleic acid is from about 160kb to about 170kb. In one embodiment, the insert nucleic acid is from about 170kb to about 180kb. In one embodiment, the insert nucleic acid is from about 180kb to about 190kb. In one embodiment, the insert nucleic acid is from about 190kb to about 200kb.

**[00017]** In one embodiment, the LTVEC comprises an insert nucleic acid comprising a nucleic acid flanked with site-specific recombination target sequences. In one embodiment, the nucleic acid comprises a genomic nucleic acid. In one embodiment, the genomic nucleic acid is derived from a mouse, a human, or a combination thereof. In one embodiment, the site-specific recombination target sequences are selected from the group consisting of loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox, and

a combination thereof.

**[00018]** In one embodiment, the LTVEC comprises an insert nucleic acid comprising a conditional allele. In one embodiment, the conditional allele is a multifunctional allele, as described in US 2011/0104799, which is incorporated by reference in its entirety. In one embodiment, the conditional allele comprises: (a) an actuating sequence in sense orientation with respect to transcription of a target gene, and a drug selection cassette in sense or antisense orientation; (b) in antisense orientation a nucleotide sequence of interest (NSI) and a conditional by inversion module (COIN, which utilizes an exon-splitting intron and an invertible genetrap-like module; see, for example, US 2011/0104799, which is incorporated by reference in its entirety); and (c) recombinable units that recombine upon exposure to a first recombinase to form a conditional allele that (i) lacks the actuating sequence and the DSC, and (ii) contains the NSI in sense orientation and the COIN in antisense orientation.

**[00019]** In one embodiment, the LTVEC comprises an insert nucleic acid comprising a selection cassette. In one embodiment, the selection cassette comprises a nucleic acid sequence encoding a selection marker, wherein the nucleic acid sequence is operably linked to a promoter. In one embodiment, the promoter is active in a prokaryotic cell. In one embodiment, the nucleic acid is active in both prokaryotic and eukaryotic cells. In one embodiment, the selection cassette is flanked with site-specific recombination target sequences. In one embodiment, the selection marker is selected from neomycin phosphotransferase ($neo^r$), hygromycin B phosphotransferase ($hyg^r$), puromycin-N-acetyltransferase ($puro^r$), blasticidin S deaminase ($bsr^r$), xanthine/guanine phosphoribosyl transferase (gpt), and herpes simplex virus thymidine kinase (HSV-k), and a combination thereof.

**[00020]** In one embodiment, the LTVEC comprises an insert nucleic acid comprising a reporter gene operably linked to a promoter, wherein the reporter gene encodes a reporter protein selected from the group consisting of LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, enhanced yellow fluorescent protein (EYFP), Emerald, enhanced green fluorescent protein (EGFP), CyPet, cyan fluorescent protein (CFP), Cerulean, T-Sapphire, luciferase, alkaline phosphatase, and a combination thereof. In one embodiment, the reporter gene is expressed under a control of an inducible promoter. In one embodiment, the reporter gene is expressed under a control of an endogenous promoter. In one embodiment, the reporter gene is expressed under a control of an exogenous promoter. In one embodiment, the reporter gene is expressed in a specific cell type. In one embodiment, the reporter gene is expressed in a tissue-specific

manner. In one embodiment, the reporter gene is expressed in a developmental stage-specific manner.

**[00021]** In one aspect, a eukaryotic cell is provided, comprising a large targeting vector comprising homology arms directed to a target locus within the genome of the eukaryotic cell, and a nucleic acid sequence encoding a nuclease agent that makes a single- or double-stranded break at or near the target locus.

**[00022]** In one embodiment, the eukaryotic cell is a pluripotent cell. In one embodiment, the pluripotent cell is an embryonic stem (ES) cell. In one embodiment, the pluripotent cell is a non-human ES cell. In one embodiment, the pluripotent cell is an induced pluripotent stem (iPS) cell. In one embodiment, the induced pluripotent (iPS) cell is derived from a fibroblast. In one embodiment, the induced pluripotent (iPS) cell is derived from a human fibroblast. In one embodiment, the pluripotent cell is a hematopoietic stem cell (HSC). In one embodiment, the pluripotent cell is a neuronal stem cell (NSC). In one embodiment, the pluripotent cell is an epiblast stem cell. In one embodiment, the pluripotent cell is a developmentally restricted progenitor cell. In one embodiment, the pluripotent cell is a rodent pluripotent cell. In one embodiment, the rodent pluripotent cell is a rat pluripotent cell. In one embodiment, the rat pluripotent cell is a rat ES cell. In one embodiment, the rodent pluripotent cell is a mouse pluripotent cell. In one embodiment, the pluripotent cell is a mouse embryonic stem (ES) cell.

**[00023]** In one embodiment, the eukaryotic cell is an immortalized mouse or rat cell. In one embodiment, the eukaryotic cell is an immortalized human cell. In one embodiment, the eukaryotic cell is a human fibroblast. In one embodiment, the eukaryotic cell is a cancer cell. In one embodiment, the eukaryotic cell is a human cancer cell.

**[00024]** In one embodiment, the LTVEC ranges from about 50kb to about 300kb. In one embodiment, the LTVEC ranges from about 50kb to about 75kb. In one embodiment, the LTVEC ranges from about 75kb to about 100kb. In one embodiment, the LTVEC ranges from about 100kb to 125kb. In one embodiment, the LTVEC ranges from about 125kb to about 150kb. In one embodiment, the LTVEC ranges from about 150kb to about 175kb. In one embodiment, the LTVEC ranges from about 175kb to about 200kb. In one embodiment, the LTVEC ranges from about 200kb to about 225kb. In one embodiment, the LTVEC ranges from about 225kb to about 250kb. In one embodiment, the LTVEC ranges from about 250kb to about 275kb. In one embodiment, the LTVEC ranges from about 275kb to about 300kb.

**[00025]** In one embodiment, the homology arms of the targeting vector are derived from a BAC library, a cosmid library, or a P1 phage library. In one embodiment, the homology arms are derived from a genomic locus of the non-human animal that is not targetable using a conventional method. In one embodiment, the homology arms are derived from a synthetic DNA.

**[00026]** In one embodiment, a sum total of the upstream homology arm and the downstream homology arm is at least 10kb. In one embodiment, the upstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the downstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the upstream and the downstream homology arms range from about 5kb to about 10kb. In one embodiment, the upstream and the downstream homology arms range from about 10kb to about 20kb. In one embodiment, the upstream and the downstream homology arms range from about 20kb to about 30kb. In one embodiment, the upstream and the downstream homology arms range from about 30kb to about 40kb. In one embodiment, the upstream and the downstream homology arms range from about 40kb to about 50kb. In one embodiment, the upstream and the downstream homology arms range from about 50kb to about 60kb. In one embodiment, the upstream and the downstream homology arms range from about 60kb to about 70kb. In one embodiment, the upstream and the downstream homology arms range from about 70kb to about 80kb. In one embodiment, the upstream and the downstream homology arms range from about 80kb to about 90kb. In one embodiment, the upstream and the downstream homology arms range from about 90kb to about 100kb. In one embodiment, the upstream and the downstream homology arms range from about 100kb to about 110kb. In one embodiment, the upstream and the downstream homology arms range from about 110kb to about 120kb. In one embodiment, the upstream and the downstream homology arms range from about 120kb to about 130kb. In one embodiment, the upstream and the downstream homology arms range from about 130kb to about 140kb. In one embodiment, the upstream and the downstream homology arms range from about 140kb to about 150kb. In one embodiment, the upstream and the downstream homology arms range from about 150kb to about 160kb. In one embodiment, the upstream and the downstream homology arms range from about 160kb to about 170kb. In one embodiment, the upstream and the downstream homology arms range from about 170kb to about 180kb. In one embodiment, the upstream and the downstream homology arms range from about 180kb to about 190kb. In one embodiment, the upstream and the downstream homology arms range from about 190kb to about 200kb.

**[00027]** In one embodiment, the LTVEC comprises a selection cassette. In one

embodiment, the selection cassette comprises a nucleic acid sequence encoding a selection marker, wherein the nucleic acid sequence is operably linked to a promoter. In one embodiment, the promoter is active in a prokaryotic cell. In one embodiment, the promoter is active both in prokaryotic and eukaryotic cells. In one embodiment, the selection marker is selected from neomycin phosphotransferase ($neo^r$), hygromycin B phosphotransferase ($hyg^r$), puromycin-N-acetyltransferase ($puro^r$), blasticidin S deaminase ($bsr^r$), xanthine/guanine phosphoribosyl transferase (gpt), and herpes simplex virus thymidine kinase (HSV-k), and a combination thereof.

**[00028]** In one embodiment, the LTVEC comprises an insert nucleic acid ranging from about 5kb to about 200kb. In one embodiment, the insert nucleic acid is from about 5kb to about 10kb. In one embodiment, the insert nucleic acid is from about 10kb to about 20kb. In one embodiment, the insert nucleic acid is from about 20kb to about 30kb. In one embodiment, the insert nucleic acid is from about 30kb to about 40kb. In one embodiment, the insert nucleic acid is from about 40kb to about 50kb. In one embodiment, the insert nucleic acid is from about 60kb to about 70kb. In one embodiment, the insert nucleic acid is from about 80kb to about 90kb. In one embodiment, the insert nucleic acid is from about 90kb to about 100kb. In one embodiment, the insert nucleic acid is from about 100kb to about 110kb. In one embodiment, the insert nucleic acid is from about 120kb to about 130kb. In one embodiment, the insert nucleic acid is from about 130kb to about 140kb. In one embodiment, the insert nucleic acid is from about 140kb to about 150kb. In one embodiment, the insert nucleic acid is from about 150kb to about 160kb. In one embodiment, the insert nucleic acid is from about 160kb to about 170kb. In one embodiment, the insert nucleic acid is from about 170kb to about 180kb. In one embodiment, the insert nucleic acid is from about 180kb to about 190kb. In one embodiment, the insert nucleic acid is from about 190kb to about 200kb.

**[00029]** In one embodiment, the LTVEC comprises an insert nucleic acid comprising a nucleic acid flanked with site-specific recombination target sequences. In one embodiment, the nucleic acid comprises a genomic nucleic acid. In one embodiment, the genomic nucleic acid is derived from a mouse, a human, or a combination thereof. In one embodiment, the site-specific recombination target sequences are selected from the group consisting of loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox, and a combination thereof.

**[00030]** In one embodiment, the LTVEC comprises an insert nucleic acid comprising a conditional allele. In one embodiment, the conditional allele is a multifunctional allele, as described in US 2011/0104799, which is incorporated by reference in its entirety. In one

embodiment, the conditional allele comprises: (a) an actuating sequence in sense orientation with respect to transcription of a target gene, and a drug selection cassette in sense or antisense orientation; (b) in antisense orientation a nucleotide sequence of interest (NSI) and a conditional by inversion module (COIN, which utilizes an exon-splitting intron and an invertible genetrap-like module; see, for example, US 2011/0104799, which is incorporated by reference in its entirety); and (c) recombinable units that recombine upon exposure to a first recombinase to form a conditional allele that (i) lacks the actuating sequence and the DSC, and (ii) contains the NSI in sense orientation and the COIN in antisense orientation.

**[00031]** In one embodiment, the LTVEC comprises an insert nucleic acid comprising a selection cassette. In one embodiment, the selection cassette comprises a nucleic acid sequence encoding a selection marker, wherein the nucleic acid sequence is operably linked to a promoter. In one embodiment, the promoter is active in a prokaryotic cell. In one embodiment, the nucleic acid is active in both prokaryotic and eukaryotic cells. In one embodiment, the selection cassette is flanked with site-specific recombination target sequences. In one embodiment, the selection marker is selected from neomycin phosphotransferase ($neo^r$), hygromycin B phosphotransferase ($hyg^r$), puromycin-N-acetyltransferase ($puro^r$), blasticidin S deaminase ($bsr^r$), xanthine/guanine phosphoribosyl transferase (gpt), and herpes simplex virus thymidine kinase (HSV-k), and a combination thereof.

**[00032]** In one embodiment, the LTVEC comprises an insert nucleic acid comprising a reporter gene operably linked to a promoter, wherein the reporter gene encodes a reporter protein selected from the group consisting of LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, enhanced yellow fluorescent protein (EYFP), Emerald, enhanced green fluorescent protein (EGFP), CyPet, cyan fluorescent protein (CFP), Cerulean, T-Sapphire, luciferase, alkaline phosphatase, and a combination thereof. In one embodiment, the reporter gene is expressed under a control of an inducible promoter. In one embodiment, the reporter gene is expressed under a control of an endogenous promoter. In one embodiment, the reporter gene is expressed under a control of an exogenous promoter. In one embodiment, the reporter gene is expressed in a specific cell type. In one embodiment, the reporter gene is expressed in a tissue-specific manner. In one embodiment, the reporter gene is expressed in a developmental stage-specific manner.

**[00033]** In one aspect, a method for modifying a target genomic locus of a mammalian cell via bacterial homologous recombination (BHR) in a prokaryotic cell is provided,

comprising:

(a) introducing into a prokaryotic cell comprising a target genomic locus of a mammal:

(i) a targeting vector comprising an insert nucleic acid flanked with a first upstream homology arm and a first downstream homology arm, and

(ii) a nuclease agent that makes a single or double-strand break at or near the target genomic locus, and

(i) selecting a targeted prokaryotic cell comprising the insert nucleic acid,

wherein the prokaryotic cell is capable of expressing a recombinase that mediates the BHR.

**[00034]** In one embodiment, the target genomic locus is selected from an FcER1a locus, a TLR4 locus, a PRLR locus, a Notch4 locus, an Accn2 locus, an Adamts5 locus, a TRPA1 locus, FolH1 locus, an LRP5 locus, and an ERBB4 locus.

**[00035]** In one embodiment, the target genomic locus is present in a large targeting vector (LTVEC) comprising a second upstream homology arm and a second downstream homology arm. In one embodiment, a sum total of the second upstream homology arm and the second downstream homology arm is at least 10kb. In one embodiment, the second upstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the second downstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the second upstream and the downstream homology arms range from about 5kb to about 10kb. In one embodiment, the second upstream and the downstream homology arms range from about 10kb to about 20kb. In one embodiment, the second upstream and the downstream homology arms range from about 20kb to about 30kb. In one embodiment, the second upstream and the downstream homology arms range from about 30kb to about 40kb. In one embodiment, the second upstream and the downstream homology arms range from about 40kb to about 50kb. In one embodiment, the second upstream and the downstream homology arms range from about 50kb to about 60kb. In one embodiment, the second upstream and the downstream homology arms range from about 60kb to about 70kb. In one embodiment, the second upstream and the downstream homology arms range from about 70kb to about 80kb. In one embodiment, the second upstream and the downstream homology arms range from about 80kb to about 90kb. In one embodiment, the second upstream and the downstream homology arms range from about 90kb to about

100kb. In one embodiment, the second upstream and the downstream homology arms range from about 100kb to about 110kb. In one embodiment, the second upstream and the downstream homology arms range from about 110kb to about 120kb. In one embodiment, the second upstream and the downstream homology arms range from about 120kb to about 130kb. In one embodiment, the second upstream and the downstream homology arms range from about 130kb to about 140kb. In one embodiment, the second upstream and the downstream homology arms range from about 140kb to about 150kb. In one embodiment, the second upstream and the downstream homology arms range from about 150kb to about 160kb. In one embodiment, the second upstream and the downstream homology arms range from about 160kb to about 170kb. In one embodiment, the second upstream and the downstream homology arms range from about 170kb to about 180kb. In one embodiment, the second upstream and the downstream homology arms range from about 180kb to about 190kb. In one embodiment, the second upstream and the downstream homology arms range from about 190kb to about 200kb.

**[00036]** In one embodiment, the mammal is a human and the targeting is of an *ex vivo* human cell. In one embodiment, the mammal is a rodent. In one embodiment, the rodent is selected from a mouse, a rat, and a hamster.

**[00037]** In one embodiment, the nuclease agent is introduced together with the targeting vector. In one embodiment, the nuclease agent is introduced separately from the targeting vector over a period of time. In one embodiment, the nuclease agent is introduced prior to the introduction of the targeting vector. In one embodiment, the nuclease agent is introduced following introduction of the targeting vector.

**[00038]** In one embodiment, combined use of the targeting vector with the nuclease agent results in an increased targeting efficiency compared to use of the targeting vector alone. In one embodiment, when the targeting vector is used in conjunction with the nuclease agent, targeting efficiency of the targeting vector is increased at least by two-fold compared to when the targeting vector is used alone. In one embodiment, when the targeting vector is used in conjunction with the nuclease agent, targeting efficiency of the targeting vector is increased at least by three-fold compared to when the targeting vector is used alone. In one embodiment, when the targeting vector is used in conjunction with the nuclease agent, targeting efficiency of the targeting vector is increased at least by four-fold compared to when the targeting vector is used alone.

**[00039]** In one embodiment, the prokaryotic cell is a recombination-competent strain of *E. coli*. In one embodiment, the prokaryotic cell comprises a nucleic acid that encodes the

recombinase. In one embodiment, the prokaryotic cell does not comprise the nucleic acid that encodes the recombinase, and the nucleic acid encoding the recombinase is introduced into the prokaryotic cell. In one embodiment, the nucleic acid comprises a DNA or an mRNA encoding the recombinase. In one embodiment the nucleic acid encoding the recombinase is pABG. In one embodiment, the recombinase is expressed under the control of an inducible promoter. In one embodiment, expression of the recombinase is controlled by arabinose.

**[00040]** In one embodiment, the nuclease agent is an expression construct comprising a nucleic acid sequence encoding a nuclease, wherein the nucleic acid sequence is operably linked to a promoter. In one embodiment, the promoter is a constitutively active promoter. In one embodiment, the promoter is an inducible promoter. In one embodiment, the promoter is active in the prokaryotic cell. In one embodiment, the nuclease agent is an mRNA encoding an endonuclease.

**[00041]** In one embodiment, the nuclease agent is a zinc-finger nuclease (ZFN). In one embodiment, each monomer of the ZFN comprises 3 or more zinc finger-based DNA binding domains, wherein each zinc finger-based DNA binding domain binds to a 3bp subsite. In one embodiment, the ZFN is a chimeric protein comprising a zinc finger-based DNA binding domain operably linked to an independent nuclease. In one embodiment, the independent endonuclease is a FokI endonuclease. In one embodiment, the nuclease agent comprises a first ZFN and a second ZFN, wherein each of the first ZFN and the second ZFN is operably linked to a FokI nuclease, wherein the first and the second ZFN recognize two contiguous target DNA sequences in each strand of the target DNA sequence separated by about 6bp to about 40bp cleavage site, and wherein the FokI nucleases dimerize and make a double strand break.

**[00042]** In one embodiment, the nuclease agent is a Transcription Activator-Like Effector Nuclease (TALEN). In one embodiment, each monomer of the TALEN comprises 12-25 TAL repeats, wherein each TAL repeat binds a 1bp subsite. In one embodiment, the nuclease agent is a chimeric protein comprising a TAL repeat-based DNA binding domain operably linked to an independent nuclease. In one embodiment, the independent nuclease is a FokI endonuclease. In one embodiment, the nuclease agent comprises a first TAL-repeat-based DNA binding domain and a second TAL-repeat-based DNA binding domain, wherein each of the first and the second TAL-repeat-based DNA binding domain is operably linked to a FokI nuclease, wherein the first and the second TAL-repeat-based DNA binding domain recognize two contiguous target DNA sequences in each strand of the target DNA sequence separated by about 6bp to about 40bp cleavage site, and wherein the FokI nucleases

dimerize and make a double strand break at a target sequence.

**[00043]** In one embodiment, each monomer of the nuclease recognizes a target sequence of at least 9 nucleotides. In one embodiment, the target sequence is from about 9 to about 12 nucleotides in length. In one embodiment, the target sequence is from about 12 to about 15 nucleotides in length. In one embodiment, the target sequence is from about 15 to about 18 nucleotides in length. In one embodiment, the target sequence is from about 18 to about 21 nucleotides in length.

**[00044]** In one embodiment, a target sequence of the nuclease agent is located in an intron. In one embodiment, the target sequence is located in an exon. In one embodiment, the target sequence is located in a promoter. In one embodiment, the target sequence is in a non-protein-coding region. In one embodiment, the non-protein-coding region is a regulatory region. In one embodiment, the target sequence is located in a promoter regulatory region. In one embodiment, the target sequence is located in an enhancer region.

**[00045]** In one embodiment, the nuclease agent is a meganuclease. In one embodiment, the meganuclease recognizes double-stranded DNA sequences of 12 to 40 base pairs. In one embodiment, the meganuclease recognizes one perfectly matched target sequence in the genome. In one embodiment, the meganuclease is a homing nuclease. In one embodiment, the homing nuclease is a LAGLIDADG family of homing nuclease. In one embodiment, the LAGLIDADG family of homing nuclease is selected from I-SceI, I-CreI, and I-DmoI.

**[00046]** In one embodiment, the targeting vector is a large targeting vector (LTVEC).

**[00047]** In one embodiment, the LTVEC ranges from about 50kb to about 300kb. In one embodiment, the LTVEC ranges from about 50kb to about 75kb. In one embodiment, the LTVEC ranges from about 75kb to about 100kb. In one embodiment, the LTVEC ranges from about 100kb to 125kb. In one embodiment, the LTVEC ranges from about 125kb to about 150kb. In one embodiment, the LTVEC ranges from about 150kb to about 175kb. In one embodiment, the LTVEC ranges from about 175kb to about 200kb. In one embodiment, the LTVEC ranges from about 200kb to about 225kb. In one embodiment, the LTVEC ranges from about 225kb to about 250kb. In one embodiment, the LTVEC ranges from about 250kb to about 275kb. In one embodiment, the LTVEC ranges from about 275kb to about 300kb.

**[00048]** In one embodiment, the homology arms of the targeting vector are derived from a BAC library, a cosmid library, or a P1 phage library. In one embodiment, the homology

arms are derived from a genomic locus of the non-human animal that is not targetable using a conventional method. In one embodiment, the homology arms are derived from a synthetic DNA.

**[00049]** In one embodiment, a sum total of the upstream homology arm and the downstream homology arm is at least 10kb. In one embodiment, the upstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the downstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the upstream and the downstream homology arms range from about 5kb to about 10kb. In one embodiment, the upstream and the downstream homology arms range from about 10kb to about 20kb. In one embodiment, the upstream and the downstream homology arms range from about 20kb to about 30kb. In one embodiment, the upstream and the downstream homology arms range from about 30kb to about 40kb. In one embodiment, the upstream and the downstream homology arms range from about 40kb to about 50kb. In one embodiment, the upstream and the downstream homology arms range from about 50kb to about 60kb. In one embodiment, the upstream and the downstream homology arms range from about 60kb to about 70kb. In one embodiment, the upstream and the downstream homology arms range from about 70kb to about 80kb. In one embodiment, the upstream and the downstream homology arms range from about 80kb to about 90kb. In one embodiment, the upstream and the downstream homology arms range from about 90kb to about 100kb. In one embodiment, the upstream and the downstream homology arms range from about 100kb to about 110kb. In one embodiment, the upstream and the downstream homology arms range from about 110kb to about 120kb. In one embodiment, the upstream and the downstream homology arms range from about 120kb to about 130kb. In one embodiment, the upstream and the downstream homology arms range from about 130kb to about 140kb. In one embodiment, the upstream and the downstream homology arms range from about 140kb to about 150kb. In one embodiment, the upstream and the downstream homology arms range from about 150kb to about 160kb. In one embodiment, the upstream and the downstream homology arms range from about 160kb to about 170kb. In one embodiment, the upstream and the downstream homology arms range from about 170kb to about 180kb. In one embodiment, the upstream and the downstream homology arms range from about 180kb to about 190kb. In one embodiment, the upstream and the downstream homology arms range from about 190kb to about 200kb.

**[00050]** In one embodiment, the targeting vector comprises a selection cassette. In one embodiment, the selection cassette comprises a nucleic acid sequence encoding a selection marker, wherein the nucleic acid sequence is operably linked to a promoter. In one

embodiment, the promoter is active in a prokaryotic cell. In one embodiment, the promoter is active both in prokaryotic and eukaryotic cells. In one embodiment, the selection marker is selected from neomycin phosphotransferase ($neo^r$), hygromycin B phosphotransferase ($hyg^r$), puromycin-N-acetyltransferase ($puro^r$), blasticidin S deaminase ($bsr^r$), xanthine/guanine phosphoribosyl transferase (gpt), and herpes simplex virus thymidine kinase (HSV-k), and a combination thereof.

**[00051]** In one embodiment, the insert nucleic acid is from about 5kb to about 200kb. In one embodiment, the insert nucleic acid is from about 5kb to about 10kb. In one embodiment, the insert nucleic acid is from about 10kb to about 20kb. In one embodiment, the insert nucleic acid is from about 20kb to about 30kb. In one embodiment, the insert nucleic acid is from about 30kb to about 40kb. In one embodiment, the insert nucleic acid is from about 40kb to about 50kb. In one embodiment, the insert nucleic acid is from about 60kb to about 70kb. In one embodiment, the insert nucleic acid is from about 80kb to about 90kb. In one embodiment, the insert nucleic acid is from about 90kb to about 100kb. In one embodiment, the insert nucleic acid is from about 100kb to about 110kb. In one embodiment, the insert nucleic acid is from about 120kb to about 130kb. In one embodiment, the insert nucleic acid is from about 130kb to about 140kb. In one embodiment, the insert nucleic acid is from about 140kb to about 150kb. In one embodiment, the insert nucleic acid is from about 150kb to about 160kb. In one embodiment, the insert nucleic acid is from about 160kb to about 170kb. In one embodiment, the insert nucleic acid is from about 170kb to about 180kb. In one embodiment, the insert nucleic acid is from about 180kb to about 190kb. In one embodiment, the insert nucleic acid is from about 190kb to about 200kb.

**[00052]** In one embodiment, the insert nucleic acid comprises a nucleic acid flanked with site-specific recombination target sequences. In one embodiment, the nucleic acid comprises a genomic nucleic acid. In one embodiment, the genomic nucleic acid is derived from a mouse, a human, or a combination thereof. In one embodiment, the site-specific recombination target sequences are selected from the group consisting of loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox, and a combination thereof.

**[00053]** In one embodiment, the insert nucleic acid comprises a conditional allele. In one embodiment, the conditional allele is a multifunctional allele, as described in US 2011/0104799, which is incorporated by reference in its entirety. In one embodiment, the conditional allele comprises: (a) an actuating sequence in sense orientation with respect to transcription of a target gene, and a drug selection cassette in sense or antisense

orientation; (b) in antisense orientation a nucleotide sequence of interest (NSI) and a conditional by inversion module (COIN, which utilizes an exon-splitting intron and an invertible genetrap-like module; see, for example, US 2011/0104799, which is incorporated by reference in its entirety); and (c) recombinable units that recombine upon exposure to a first recombinase to form a conditional allele that (i) lacks the actuating sequence and the DSC, and (ii) contains the NSI in sense orientation and the COIN in antisense orientation.

**[00054]** In one embodiment, the insert nucleic acid comprises a selection cassette. In one embodiment, the selection cassette comprises a nucleic acid sequence encoding a selection marker, wherein the nucleic acid sequence is operably linked to a promoter. In one embodiment, the promoter is active in a prokaryotic cell. In one embodiment, the nucleic acid is active in both prokaryotic and eukaryotic cells. In one embodiment, the selection cassette is flanked with site-specific recombination target sequences. In one embodiment, the selection marker is selected from neomycin phosphotransferase ($neo^r$), hygromycin B phosphotransferase ($hyg^r$), puromycin-N-acetyltransferase ($puro^r$), blasticidin S deaminase ($bsr^r$), xanthine/guanine phosphoribosyl transferase (gpt), and herpes simplex virus thymidine kinase (HSV-k), and a combination thereof.

**[00055]** In one embodiment, the insert nucleic acid comprises a reporter gene operably linked to a promoter, wherein the reporter gene encodes a reporter protein selected from the group consisting of LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, enhanced yellow fluorescent protein (EYFP), Emerald, enhanced green fluorescent protein (EGFP), CyPet, cyan fluorescent protein (CFP), Cerulean, T-Sapphire, luciferase, alkaline phosphatase, and a combination thereof. In one embodiment, the reporter gene is expressed under a control of an inducible promoter. In one embodiment, the reporter gene is expressed under a control of an endogenous promoter. In one embodiment, the reporter gene is expressed under a control of an exogenous promoter. In one embodiment, the reporter gene is expressed in a specific cell type. In one embodiment, the reporter gene is expressed in a tissue-specific manner. In one embodiment, the reporter gene is expressed in a developmental stage-specific manner.

**[00056]** In one embodiment, integration of the insert nucleic acid into the target genomic locus introduces one or more genetic modifications as described herein. In one embodiment, the genetic modification is a deletion of an endogenous nucleic acid sequence. In one embodiment, the genetic modification is an addition of an exogenous nucleic acid sequence into the target genomic locus. In one embodiment, the genetic modification is a replacement of an endogenous nucleic acid sequence with an exogenous nucleic acid sequence at the target genomic locus. In one embodiment, the exogenous nucleic acid sequence is a non-

mouse nucleic acid sequence. In one embodiment, the exogenous nucleic acid sequence is a human nucleic acid sequence. In one embodiment, the genetic modification is a knockout, a deletion, an insertion, a replacement ("knock-in"), a point mutation, a domain swap, an exon swap, an intron swap, a regulatory sequence swap, a gene swap, or a combination thereof.

**[00057]** In one embodiment, the insert nucleic acid is homologous to a mouse nucleic acid sequence. In one embodiment, the insert nucleic acid is a human nucleic acid. In one embodiment, the insert nucleic acid is a fragment of a genomic nucleic acid. In one embodiment, the genomic nucleic acid is a mouse genomic nucleic acid, a human genomic nucleic acid, or a combination thereof. In one embodiment, the insert nucleic acid ranges from about 5kb to about 200kb as described above.

**[00058]** In one embodiment, the insert nucleic acid is orthologous to a mouse nucleic acid sequence. In one embodiment, the insert nucleic acid is a human nucleic acid. In one embodiment, the insert nucleic acid is a fragment of a genomic nucleic acid. In one embodiment, the genomic nucleic acid is a mouse genomic nucleic acid, a human genomic nucleic acid, or a combination thereof. In one embodiment, the insert nucleic acid ranges from about 5kb to about 200kb as described above.

**[00059]** In one embodiment, the insert nucleic acid comprises a genomic locus that encodes a protein expressed in the nervous system, the skeletal system, the digestive system, the circulatory system, the muscular system, the respiratory system, the cardiovascular system, the lymphatic system, the endocrine system, the urinary system, the reproductive system, or a combination thereof. In one embodiment, the insert nucleic acid comprises a genomic locus that encodes a protein expressed in a bone marrow or a bone marrow-derived cell. In one embodiment, the insert nucleic acid comprises a genomic locus that encodes a protein expressed in a spleen cell. In one embodiment, the genomic locus comprises a mouse genomic nucleic acid sequence, a human genomic nucleic acid sequence, or a combination thereof. In one embodiment, the nucleic acid comprises a genomic locus that encodes a protein expressed in a B cell. In one embodiment, the nucleic acid comprises a genomic locus that encodes a protein expressed in an immature B cell. In one embodiment, the nucleic acid comprises a genomic locus that encodes a protein expressed in a mature B cell.

**[00060]** In one embodiment, the insert nucleic acid comprises a genomic nucleic acid sequence that encodes a human immunoglobulin heavy chain variable region amino acid sequence.

**[00061]** In one embodiment, the genomic nucleic acid sequence comprises an unrearranged human immunoglobulin heavy chain variable region nucleic acid sequence operably linked to an immunoglobulin heavy chain constant region nucleic acid sequence. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is a mouse immunoglobulin heavy chain constant region nucleic acid sequence or human immunoglobulin heavy chain constant region nucleic acid sequence, or a combination thereof. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is selected from a $C_H1$, a hinge, a $C_H2$, a $C_H3$, and a combination thereof. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a $C_H1$-hinge-$C_H2$-$C_H3$. In one embodiment, the genomic nucleic acid sequence comprises a rearranged human immunoglobulin heavy chain variable region nucleic acid sequence operably linked to an immunoglobulin heavy chain constant region nucleic acid sequence. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is a mouse immunoglobulin heavy chain constant region nucleic acid sequence or a human immunoglobulin heavy chain constant region nucleic acid sequence, or a combination thereof. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is selected from a $C_H1$, a hinge, a $C_H2$, a $C_H3$, and a combination thereof. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a $C_H1$-hinge-$C_H2$-$C_H3$.

**[00062]** In one embodiment, the insert nucleic acid comprises a genomic nucleic acid sequence that encodes a human immunoglobulin light chain variable region amino acid sequence. In one embodiment, the genomic nucleic acid sequence comprises an unrearranged human λ and/or κ light chain variable region nucleic acid sequence. In one embodiment, the genomic nucleic acid sequence comprises a rearranged human λ and/or κ light chain variable region nucleic acid sequence. In one embodiment, the unrearranged or rearranged λ and/or κ light chain variable region nucleic acid sequence is operably linked to a mouse, rat, or human immunoglobulin light chain constant region nucleic acid sequence selected from a λ light chain constant region nucleic acid sequence and a κ light chain constant region nucleic acid sequence.

**[00063]** In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence. In one embodiment, the human nucleic acid sequence encodes an extracellular protein. In one embodiment, the human nucleic acid sequence encodes a ligand for a receptor. In one embodiment, the ligand is a cytokine. In one embodiment, the cytokine is a chemokine selected from CCL, CXCL, CX3CL, and XCL. In one embodiment, the cytokine is a tumor necrosis factor (TNF). In one embodiment, the cytokine is an interleukin (IL). In one

embodiment, the interleukin is selected from IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IL-34, IL-35, and IL-36. In one embodiment, the interleukin is IL-2. In one embodiment, the human genomic nucleic acid sequence encodes a cytoplasmic protein. In one embodiment, the human genomic nucleic acid sequence encodes a membrane protein. In one embodiment, the membrane protein is a receptor. In one embodiment, the receptor is a cytokine receptor. In one embodiment, the cytokine receptor is an interleukin receptor. In one embodiment, the interleukin receptor is an interleukin 2 receptor alpha. In one embodiment, the interleukin receptor is an interleukin 2 receptor beta. In one embodiment, the interleukin receptor is an interleukin 2 receptor gamma. In one embodiment, the human genomic nucleic acid sequence encodes a nuclear protein. In one embodiment, the nuclear protein is a nuclear receptor.

**[00064]** In one embodiment, the insert nucleic acid comprises a genetic modification in a coding sequence. In one embodiment, the genetic modification comprises a deletion mutation of a coding sequence. In one embodiment, the genetic modification comprises a fusion of two endogenous coding sequences.

**[00065]** In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence encoding a mutant human protein. In one embodiment, the mutant human protein is characterized by an altered binding characteristic, altered localization, altered expression, and/or altered expression pattern. In one embodiment, the human nucleic acid sequence comprises at least one human disease allele. In one embodiment, the human disease allele is an allele of a neurological disease. In one embodiment, the human disease allele is an allele of a cardiovascular disease. In one embodiment, the human disease allele is an allele of a kidney disease. In one embodiment, the human disease allele is an allele of a muscle disease. In one embodiment, the human disease allele is an allele of a blood disease. In one embodiment, the human disease allele is an allele of a cancer-causing gene. In one embodiment, the human disease allele is an allele of an immune system disease. In one embodiment, the human disease allele is a dominant allele. In one embodiment, the human disease allele is a recessive allele. In one embodiment, the human disease allele comprises a single nucleotide polymorphism (SNP) allele.

**[00066]** In one embodiment, the insert nucleic acid comprises a regulatory sequence. In one embodiment, the regulatory sequence is a promoter sequence. In one embodiment, the regulatory sequence is an enhancer sequence. In one embodiment, the regulatory sequence is a transcriptional repressor-binding sequence. In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence, wherein the human nucleic acid

sequence comprises a deletion of a non-protein-coding sequence, but does not comprise a deletion of a protein-coding sequence. In one embodiment, the deletion of the non-protein-coding sequence comprises a deletion of a regulatory sequence. In one embodiment, the deletion of the regulatory element comprises a deletion of a promoter sequence. In one embodiment, the deletion of the regulatory element comprises a deletion of an enhancer sequence.

**[00067]** In one aspect, a method for modifying a target genomic locus in a mammalian cell is provided, comprising introducing into a mammalian cell: (i) a nuclease agent that makes a singe or double-strand break at or near the target genomic locus, and (ii) a large targeting vector (LTVEC) comprising an insert nucleic acid flanked with an upstream homology arm and a downstream homology arm.

**[00068]** In one embodiment, the mammalian cell is a pluripotent cell. In one embodiment, the pluripotent cell is an embryonic stem (ES) cell. In one embodiment, the pluripotent cell is a non-human ES cell. In one embodiment, the pluripotent cell is an induced pluripotent stem (iPS) cell. In one embodiment, the induced pluripotent (iPS) cell is derived from a fibroblast. In one embodiment, the induced pluripotent (iPS) cell is derived from a human fibroblast. In one embodiment, the pluripotent cell is a hematopoietic stem cell (HSC). In one embodiment, the pluripotent cell is a neuronal stem cell (NSC). In one embodiment, the pluripotent cell is an epiblast stem cell. In one embodiment, the pluripotent cell is a developmentally restricted progenitor cell.

**[00069]** In one embodiment, the pluripotent cell is a rodent pluripotent cell. In one embodiment, the rodent pluripotent cell is a rat pluripotent cell. In one embodiment, the rat pluripotent cell is a rat ES cell. In one embodiment, the rodent pluripotent cell is a mouse pluripotent cell. In one embodiment, the pluripotent cell is a mouse embryonic stem (ES) cell.

**[00070]** In one embodiment, the mammalian cell is an immortalized mouse or rat cell. In one embodiment, the mammalian cell is an immortalized human cell. In one embodiment, the mammalian cell is a human fibroblast. In one embodiment, the mammalian cell is a cancer cell. In one embodiment, the mammalian cell is a human cancer cell.

**[00071]** In one embodiment, the mammalian cell is a human cell isolated from a patient having a disease. In one embodiment, the human cell comprises a human nucleic acid sequence encoding a mutant protein. In one embodiment, the mutant human protein is characterized by an altered binding characteristic, altered localization, altered expression,

and/or altered expression pattern. In one embodiment, the human nucleic acid sequence comprises at least one human disease allele. In one embodiment, the human nucleic acid sequence comprises at least one human disease allele. In one embodiment, the human disease allele is an allele of a neurological disease. In one embodiment, the human disease allele is an allele of a cardiovascular disease. In one embodiment, the human disease allele is an allele of a kidney disease. In one embodiment, the human disease allele is an allele of a muscle disease. In one embodiment, the human disease allele is an allele of a blood disease. In one embodiment, the human disease allele is an allele of a cancer-causing gene. In one embodiment, the human disease allele is an allele of an immune system disease. In one embodiment, the human disease allele is a dominant allele. In one embodiment, the human disease allele is a recessive allele. In one embodiment, the human disease allele comprises a single nucleotide polymorphism (SNP) allele.

**[00072]** In one embodiment, the target genomic locus is selected from an FcER1a locus, a TLR4 locus, a PRLR locus, a Notch4 locus, an Accn2 locus, an Adamts5 locus, a TRPA1 locus, FolH1 locus, an LRP5 locus, and an ERBB4 locus.

**[00073]** In one embodiment, the target genomic locus comprises a human genomic sequence. In one embodiment, the target genomic locus comprises a genomic nucleic acid sequence of non-human animal. In one embodiment, the non-human animal is a rodent. In one embodiment, the rodent is selected from a mouse, a rat, and a hamster.

**[00074]** In one embodiment, the at least one human disease allele described above that is located at the target genomic locus is replaced with the insert nucleic acid. In one embodiment, the replacement of the human disease allele is mediated by a knockout, a deletion, an insertion, a replacement ("knock-in"), a domain swap, an exon swap, an intron swap, a regulatory sequence swap, a gene swap, or a combination thereof.

**[00075]** In one embodiment, the nuclease agent is introduced together with the large targeting vector (LTVEC). In one embodiment, the nuclease agent is introduced separately from the LTVEC over a period of time. In one embodiment, the nuclease agent is introduced prior to the introduction of the LTVEC. In one embodiment, the nuclease agent is introduced following introduction of the LTVEC.

**[00076]** In one embodiment, combined use of the LTVEC with the nuclease agent results in an increased targeting efficiency compared to use of the LTVEC alone. In one embodiment, when the LTVEC is used in conjunction with the nuclease agent, targeting efficiency of the LTVEC is increased at least by two-fold compared to when the LTVEC is

used alone. In one embodiment, when the LTVEC is used in conjunction with the nuclease agent, targeting efficiency of the LTVEC is increased at least by three-fold compared to when the LTVEC is used alone. In one embodiment, when the LTVEC is used in conjunction with the nuclease agent, targeting efficiency of the LTVEC is increased at least by four-fold compared to when the LTVEC is used alone.

**[00077]** In one embodiment, the nuclease agent is an expression construct comprising a nucleic acid sequence encoding a nuclease, wherein the nucleic acid sequence is operably linked to a promoter. In one embodiment, the promoter is a constitutively-active promoter. In one embodiment, the promoter is an inducible promoter. In one embodiment, the promoter is active in the mammalian cell. In one embodiment, the nuclease agent is an mRNA encoding an endonuclease.

**[00078]** In one embodiment, the nuclease agent is a zinc-finger nuclease (ZFN). In one embodiment, each monomer of the ZFN comprises 3 or more zinc finger-based DNA binding domains, wherein each zinc finger-based DNA binding domain binds to a 3bp subsite. In one embodiment, the ZFN is a chimeric protein comprising a zinc finger-based DNA binding domain operably linked to an independent nuclease. In one embodiment, the independent endonuclease is a FokI endonuclease. In one embodiment, the nuclease agent comprises a first ZFN and a second ZFN, wherein each of the first ZFN and the second ZFN is operably linked to a FokI nuclease, wherein the first and the second ZFN recognize two contiguous target DNA sequences in each strand of the target DNA sequence separated by about 6bp to about 40bp cleavage site, and wherein the FokI nucleases dimerize and make a double strand break.

**[00079]** In one embodiment, the nuclease agent is a Transcription Activator-Like Effector Nuclease (TALEN). In one embodiment, each monomer of the TALEN comprises 12-25 TAL repeats, wherein each TAL repeat binds a 1bp subsite. In one embodiment, the nuclease agent is a chimeric protein comprising a TAL repeat-based DNA binding domain operably linked to an independent nuclease. In one embodiment, the independent nuclease is a FokI endonuclease. In one embodiment, the nuclease agent comprises a first TAL-repeat-based DNA binding domain and a second TAL-repeat-based DNA binding domain, wherein each of the first and the second TAL-repeat-based DNA binding domain is operably linked to a FokI nuclease, wherein the first and the second TAL-repeat-based DNA binding domain recognize two contiguous target DNA sequences in each strand of the target DNA sequence separated by about 6bp to about 40bp cleavage site, and wherein the FokI nucleases dimerize and make a double strand break.

**[00080]** In one embodiment, each monomer of the nuclease recognizes a target sequence of at least 9 nucleotides. In one embodiment, the target sequence is from about 9 to about 12 nucleotides in length. In one embodiment, the target sequence is from about 12 to about 15 nucleotides in length. In one embodiment, the target sequence is from about 15 to about 18 nucleotides in length. In one embodiment, the target sequence is from about 18 to about 21 nucleotides in length.

**[00081]** In one embodiment, a target nucleic acid sequence of the nuclease agent is located in an intron. In one embodiment, the target nucleic acid sequence is located in an exon. In one embodiment, the target nucleic acid sequence is located in a promoter. In one embodiment, the target nucleic acid sequence is in a non-protein-coding region. In one embodiment, the non-protein-coding region is a regulatory region. In one embodiment, the target nucleic acid sequence is located in a promoter regulatory region. In one embodiment, the target nucleic acid sequence is located in an enhancer region.

**[00082]** In one embodiment, the nuclease agent is a meganuclease. In one embodiment, the meganuclease recognizes double-stranded DNA sequences of 12 to 40 base pairs. In one embodiment, the meganuclease recognize one perfectly matched target sequence in the genome. In one embodiment, the meganuclease is a homing nuclease. In one embodiment, the homing nuclease is a LAGLIDADG family of homing nuclease. In one embodiment, the LAGLIDADG family of homing nuclease is selected from I-SceI, I-CreI, and I-DmoI.

**[00083]** In one embodiment, the LTVEC ranges from about 50kb to about 300kb. In one embodiment, the LTVEC ranges from about 50kb to about 75kb. In one embodiment, the LTVEC ranges from about 75kb to about 100kb. In one embodiment, the LTVEC ranges from about 100kb to 125kb. In one embodiment, the LTVEC ranges from about 125kb to about 150kb. In one embodiment, the LTVEC ranges from about 150kb to about 175kb. In one embodiment, the LTVEC ranges from about 175kb to about 200kb. In one embodiment, the LTVEC ranges from about 200kb to about 225kb. In one embodiment, the LTVEC ranges from about 225kb to about 250kb. In one embodiment, the LTVEC ranges from about 250kb to about 275kb. In one embodiment, the LTVEC ranges from about 275kb to about 300kb.

**[00084]** In one embodiment, the homology arms of the LTVEC are derived from a BAC library, a cosmid library, or a P1 phage library. In one embodiment, the homology arms are derived from a genomic locus of the non-human animal that is not targetable using a conventional method. In one embodiment, the homology arms are derived from a synthetic

DNA.

**[00085]** In one embodiment, a sum total of the upstream homology arm and the downstream homology arm is at least 10kb. In one embodiment, the upstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the downstream homology arm ranges from about 5kb to about 100kb. In one embodiment, the upstream and the downstream homology arms range from about 5kb to about 10kb. In one embodiment, the upstream and the downstream homology arms range from about 10kb to about 20kb. In one embodiment, the upstream and the downstream homology arms range from about 20kb to about 30kb. In one embodiment, the upstream and the downstream homology arms range from about 30kb to about 40kb. In one embodiment, the upstream and the downstream homology arms range from about 40kb to about 50kb. In one embodiment, the upstream and the downstream homology arms range from about 50kb to about 60kb. In one embodiment, the upstream and the downstream homology arms range from about 60kb to about 70kb. In one embodiment, the upstream and the downstream homology arms range from about 70kb to about 80kb. In one embodiment, the upstream and the downstream homology arms range from about 80kb to about 90kb. In one embodiment, the upstream and the downstream homology arms range from about 90kb to about 100kb. In one embodiment, the upstream and the downstream homology arms range from about 100kb to about 110kb. In one embodiment, the upstream and the downstream homology arms range from about 110kb to about 120kb. In one embodiment, the upstream and the downstream homology arms range from about 120kb to about 130kb. In one embodiment, the upstream and the downstream homology arms range from about 130kb to about 140kb. In one embodiment, the upstream and the downstream homology arms range from about 140kb to about 150kb. In one embodiment, the upstream and the downstream homology arms range from about 150kb to about 160kb. In one embodiment, the upstream and the downstream homology arms range from about 160kb to about 170kb. In one embodiment, the upstream and the downstream homology arms range from about 170kb to about 180kb. In one embodiment, the upstream and the downstream homology arms range from about 180kb to about 190kb. In one embodiment, the upstream and the downstream homology arms range from about 190kb to about 200kb.

**[00086]** In one embodiment, the targeting vector comprises a selection cassette. In one embodiment, the selection cassette comprises a nucleic acid sequence encoding a selection marker, wherein the nucleic acid sequence is operably linked to a promoter. In one embodiment, the promoter is active in a mammalian cell. In one embodiment, the promoter is active both in prokaryotic and eukaryotic cells. In one embodiment, the selection marker is

selected from neomycin phosphotransferase ($neo^r$), hygromycin B phosphotransferase ($hyg^r$), puromycin-N-acetyltransferase ($puro^r$), blasticidin S deaminase ($bsr^r$), xanthine/guanine phosphoribosyl transferase (gpt), and herpes simplex virus thymidine kinase (HSV-k), and a combination thereof.

**[00087]** In one embodiment, the insert nucleic acid is from about 5kb to about 200kb. In one embodiment, the insert nucleic acid is from about 5kb to about 10kb. In one embodiment, the insert nucleic acid is from about 10kb to about 20kb. In one embodiment, the insert nucleic acid is from about 20kb to about 30kb. In one embodiment, the insert nucleic acid is from about 30kb to about 40kb. In one embodiment, the insert nucleic acid is from about 40kb to about 50kb. In one embodiment, the insert nucleic acid is from about 60kb to about 70kb. In one embodiment, the insert nucleic acid is from about 80kb to about 90kb. In one embodiment, the insert nucleic acid is from about 90kb to about 100kb. In one embodiment, the insert nucleic acid is from about 100kb to about 110kb. In one embodiment, the insert nucleic acid is from about 120kb to about 130kb. In one embodiment, the insert nucleic acid is from about 130kb to about 140kb. In one embodiment, the insert nucleic acid is from about 140kb to about 150kb. In one embodiment, the insert nucleic acid is from about 150kb to about 160kb. In one embodiment, the insert nucleic acid is from about 160kb to about 170kb. In one embodiment, the insert nucleic acid is from about 170kb to about 180kb. In one embodiment, the insert nucleic acid is from about 180kb to about 190kb. In one embodiment, the insert nucleic acid is from about 190kb to about 200kb.

**[00088]** In one embodiment, the insert nucleic acid comprises a nucleic acid flanked with site-specific recombination target sequences. In one embodiment, the nucleic acid comprises a genomic nucleic acid. In one embodiment, the genomic nucleic acid is derived from a mouse, a human, or a combination thereof. In one embodiment, the site-specific recombination target sequences are selected from the group consisting of loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox, and a combination thereof.

**[00089]** In one embodiment, the insert nucleic acid comprises a conditional allele. In one embodiment, the conditional allele is a multifunctional allele, as described in US 2011/0104799, which is incorporated by reference in its entirety. In one embodiment, the conditional allele comprises: (a) an actuating sequence in sense orientation with respect to transcription of a target gene, and a drug selection cassette in sense or antisense orientation; (b) in antisense orientation a nucleotide sequence of interest (NSI) and a conditional by inversion module (COIN, which utilizes an exon-splitting intron and an

invertible genetrap-like module; see, for example, US 2011/0104799, which is incorporated by reference in its entirety); and (c) recombinable units that recombine upon exposure to a first recombinase to form a conditional allele that (i) lacks the actuating sequence and the DSC, and (ii) contains the NSI in sense orientation and the COIN in antisense orientation.

**[00090]** In one embodiment, the insert nucleic acid comprises a selection cassette. In one embodiment, the selection cassette comprises a nucleic acid sequence encoding a selection marker, wherein the nucleic acid sequence is operably linked to a promoter. In one embodiment, the promoter is active in a mammalian cell. In one embodiment, the nucleic acid is active in a eukaryotic cell. In one embodiment, the selection cassette is flanked with site-specific recombination target sequences. In one embodiment, the selection marker is selected from neomycin phosphotransferase ($neo^r$), hygromycin B phosphotransferase ($hyg^r$), puromycin-N-acetyltransferase ($puro^r$), blasticidin S deaminase ($bsr^r$), xanthine/guanine phosphoribosyl transferase (gpt), and herpes simplex virus thymidine kinase (HSV-k), and a combination thereof.

**[00091]** In one embodiment, the insert nucleic acid comprises a reporter gene operably linked to a promoter, wherein the reporter gene encodes a reporter protein selected from the group consisting of LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, enhanced yellow fluorescent protein (EYFP), Emerald, enhanced green fluorescent protein (EGFP), CyPet, cyan fluorescent protein (CFP), Cerulean, T-Sapphire, luciferase, alkaline phosphatase, and a combination thereof. In one embodiment, the reporter gene is expressed under a control of an inducible promoter. In one embodiment, the reporter gene is expressed under a control of an endogenous promoter. In one embodiment, the reporter gene is expressed under a control of an exogenous promoter. In one embodiment, the reporter gene is expressed in a specific cell type. In one embodiment, the reporter gene is expressed in a tissue-specific manner. In one embodiment, the reporter gene is expressed in a developmental stage-specific manner.

**[00092]** In one embodiment, integration of the insert nucleic acid into the target genomic locus introduces one or more genetic modifications as described herein. In one embodiment, the genetic modification is a deletion of an endogenous nucleic acid sequence. In one embodiment, the genetic modification is an addition of an exogenous nucleic acid sequence into the target genomic locus. In one embodiment, the genetic modification is a replacement of an endogenous nucleic acid sequence with an exogenous nucleic acid sequence at the target genomic locus. In one embodiment, the exogenous nucleic acid sequence is a non-mouse nucleic acid sequence. In one embodiment, the exogenous nucleic acid sequence is a human nucleic acid sequence. In one embodiment, the genetic modification is a knockout,

a deletion, an insertion, a replacement ("knock-in"), a point mutation, a domain swap, an exon swap, an intron swap, a regulatory sequence swap, a gene swap, or a combination thereof.

**[00093]** In one embodiment, the insert nucleic acid is homologous to a mouse nucleic acid sequence. In one embodiment, the insert nucleic acid is a human nucleic acid. In one embodiment, the insert nucleic acid is a fragment of a genomic nucleic acid. In one embodiment, the genomic nucleic acid is a mouse genomic nucleic acid, a human genomic nucleic acid, or a combination thereof. In one embodiment, the insert nucleic acid ranges from about 5kb to about 200kb as described above.

**[00094]** In one embodiment, the insert nucleic acid is orthologous to a mouse nucleic acid sequence. In one embodiment, the insert nucleic acid is a human nucleic acid. In one embodiment, the insert nucleic acid is a fragment of a genomic nucleic acid. In one embodiment, the genomic nucleic acid is a mouse genomic nucleic acid, a human genomic nucleic acid, or a combination thereof. In one embodiment, the insert nucleic acid ranges from about 5kb to about 200kb as described above.

**[00095]** In one embodiment, the insert nucleic acid comprises a genomic locus that encodes a protein expressed in the nervous system, the skeletal system, the digestive system, the circulatory system, the muscular system, the respiratory system, the cardiovascular system, the lymphatic system, the endocrine system, the urinary system, the reproductive system, or a combination thereof. In one embodiment, the insert nucleic acid comprises a genomic locus that encodes a protein expressed in a bone marrow or a bone marrow-derived cell. In one embodiment, the insert nucleic acid comprises a genomic locus that encodes a protein expressed in a spleen cell. In one embodiment, the genomic locus comprises a mouse genomic nucleic acid sequence, a human genomic nucleic acid sequence, or a combination thereof. In one embodiment, the nucleic acid comprises a genomic locus that encodes a protein expressed in a B cell. In one embodiment, the nucleic acid comprises a genomic locus that encodes a protein expressed in an immature B cell. In one embodiment, the nucleic acid comprises a genomic locus that encodes a protein expressed in a mature B cell.

**[00096]** In one embodiment, the insert nucleic acid comprises a genomic nucleic acid sequence that encodes a human immunoglobulin heavy chain variable region amino acid sequence.

**[00097]** In one embodiment, the genomic nucleic acid sequence comprises an

unrearranged human immunoglobulin heavy chain variable region nucleic acid sequence operably linked to an immunoglobulin heavy chain constant region nucleic acid sequence. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is a mouse immunoglobulin heavy chain constant region nucleic acid sequence or human immunoglobulin heavy chain constant region nucleic acid sequence, or a combination thereof. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is selected from a $C_H1$, a hinge, a $C_H2$, a $C_H3$, and a combination thereof. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a $C_H1$-hinge-$C_H2$-$C_H3$. In one embodiment, the genomic nucleic acid sequence comprises a rearranged human immunoglobulin heavy chain variable region nucleic acid sequence operably linked to an immunoglobulin heavy chain constant region nucleic acid sequence. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is a mouse immunoglobulin heavy chain constant region nucleic acid sequence or a human immunoglobulin heavy chain constant region nucleic acid sequence, or a combination thereof. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is selected from a $C_H1$, a hinge, a $C_H2$, a $C_H3$, and a combination thereof. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a $C_H1$-hinge-$C_H2$-$C_H3$.

**[00098]** In one embodiment, the insert nucleic acid comprises a genomic nucleic acid sequence that encodes a human immunoglobulin light chain variable region amino acid sequence. In one embodiment, the genomic nucleic acid sequence comprises an unrearranged human λ and/or κ light chain variable region nucleic acid sequence. In one embodiment, the genomic nucleic acid sequence comprises a rearranged human λ and/or κ light chain variable region nucleic acid sequence. In one embodiment, the unrearranged or rearranged λ and/or κ light chain variable region nucleic acid sequence is operably linked to a mouse or human immunoglobulin light chain constant region nucleic acid sequence selected from a λ light chain constant region nucleic acid sequence and a κ light chain constant region nucleic acid sequence.

**[00099]** In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence. In one embodiment, the human nucleic acid sequence encodes an extracellular protein. In one embodiment, the human nucleic acid sequence encodes a ligand for a receptor. In one embodiment, the ligand is a cytokine. In one embodiment, the cytokine is a chemokine selected from CCL, CXCL, CX3CL, and XCL. In one embodiment, the cytokine is a tumor necrosis factor (TNF). In one embodiment, the cytokine is an interleukin (IL). In one embodiment, the interleukin is selected from IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9,

IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IL-34, IL-35, and IL-36. In one embodiment, the interleukin is IL-2. In one embodiment, the human genomic nucleic acid sequence encodes a cytoplasmic protein. In one embodiment, the human genomic nucleic acid sequence encodes a membrane protein. In one embodiment, the membrane protein is a receptor. In one embodiment, the receptor is a cytokine receptor. In one embodiment, the cytokine receptor is an interleukin receptor. In one embodiment, the interleukin receptor is an interleukin 2 receptor alpha. In one embodiment, the interleukin receptor is an interleukin 2 receptor beta. In one embodiment, the interleukin receptor is an interleukin 2 receptor gamma. In one embodiment, the human genomic nucleic acid sequence encodes a nuclear protein. In one embodiment, the nuclear protein is a nuclear receptor.

**[000100]** In one embodiment, the insert nucleic acid comprises a genetic modification in a coding sequence. In one embodiment, the genetic modification comprises a deletion mutation of a coding sequence. In one embodiment, the genetic modification comprises a fusion of two endogenous coding sequences.

**[000101]** In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence encoding a mutant human protein. In one embodiment, the mutant human protein is characterized by an altered binding characteristic, altered localization, altered expression, and/or altered expression pattern. In one embodiment, the human nucleic acid sequence comprises at least one human disease allele. In one embodiment, the human disease allele is an allele of a neurological disease. In one embodiment, the human disease allele is an allele of a cardiovascular disease. In one embodiment, the human disease allele is an allele of a kidney disease. In one embodiment, the human disease allele is an allele of a muscle disease. In one embodiment, the human disease allele is an allele of a blood disease. In one embodiment, the human disease allele is an allele of a cancer-causing gene. In one embodiment, the human disease allele is an allele of an immune system disease. In one embodiment, the human disease allele is a dominant allele. In one embodiment, the human disease allele is a recessive allele. In one embodiment, the human disease allele comprises a single nucleotide polymorphism (SNP) allele.

**[000102]** In one embodiment, the insert nucleic acid comprises a regulatory sequence. In one embodiment, the regulatory sequence is a promoter sequence. In one embodiment, the regulatory sequence is an enhancer sequence. In one embodiment, the regulatory sequence is a transcriptional repressor-binding sequence. In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence, wherein the human nucleic acid sequence comprises a deletion of a non-protein-coding sequence, but does not comprise a

deletion of a protein-coding sequence. In one embodiment, the deletion of the non-protein-coding sequence comprises a deletion of a regulatory sequence. In one embodiment, the deletion of the regulatory element comprises a deletion of a promoter sequence. In one embodiment, the deletion of the regulatory element comprises a deletion of an enhancer sequence.

**[000103]** In one aspect, a mammalian cell made with a method as described herein is provided. In one embodiment, the mammalian cell comprises an insert nucleic acid comprising one or more genetic modifications as described herein at a target genomic locus.

**[000104]** In one embodiment, the mammalian cell is a pluripotent cell. In one embodiment, the pluripotent cell is an embryonic stem (ES) cell. In one embodiment, the pluripotent cell is an induced pluripotent stem (iPS) cell. In one embodiment, the induced pluripotent (iPS) cell is derived from a fibroblast. In one embodiment, the induced pluripotent (iPS) cell is derived from a human fibroblast. In one embodiment, the pluripotent cell is a hematopoietic stem cell (HSC). In one embodiment, the pluripotent cell is a neuronal stem cell (NSC). In one embodiment, the pluripotent cell is an epiblast stem cell. In one embodiment, the pluripotent cell is a developmentally restricted progenitor cell.

**[000105]** In one embodiment, the pluripotent cell is a mouse pluripotent cell. In one embodiment, the pluripotent cell is a mouse embryonic stem (ES) cell.

**[000106]** In one embodiment, the mammalian cell is an immortalized mouse or rat cell. In one embodiment, the mammalian cell is an immortalized human cell. In one embodiment, the mammalian cell is a human fibroblast. In one embodiment, the mammalian cell is a cancer cell. In one embodiment, the mammalian cell is a human cancer cell.

**[000107]** In one embodiment, the target genomic locus is selected from an FcER1a locus, a TLR4 locus, a PRLR locus, a Notch4 locus, an Accn2 locus, an Adamts5 locus, a TRPA1 locus, FolH1 locus, an LRP5 locus, and an ERBB4 locus.

**[000108]** In one embodiment, the target genomic locus comprises one or more genetic modifications as described herein. In one embodiment, the genetic modification is a deletion of an endogenous nucleic acid sequence. In one embodiment, the genetic modification is an addition of an exogenous nucleic acid sequence into the target genomic locus. In one embodiment, the genetic modification is a replacement of an endogenous nucleic acid sequence with an exogenous nucleic acid sequence at the target genomic locus. In one embodiment, the exogenous nucleic acid sequence is a non-mouse nucleic acid sequence. In one embodiment, the exogenous nucleic acid sequence is a human nucleic acid

sequence. In one embodiment, the target genomic locus comprises a modification selected from a knockout, a deletion, an insertion, a replacement ("knock-in"), a domain swap, an exon swap, an intron swap, a regulatory sequence swap, a gene swap, and a combination thereof.

**[000109]** In one embodiment, the target genomic locus comprises an insert nucleic acid that is homologous to a mouse nucleic acid sequence. In one embodiment, the insert nucleic acid is a human nucleic acid. In one embodiment, the insert nucleic acid is a fragment of a genomic nucleic acid. In one embodiment, the genomic nucleic acid is a mouse genomic nucleic acid, a human genomic nucleic acid, or a combination thereof. In one embodiment, the insert nucleic acid ranges from about 5kb to about 200kb as described above.

**[000110]** In one embodiment, the target genomic locus comprises an insert nucleic acid that is orthologous to a mouse nucleic acid sequence. In one embodiment, the insert nucleic acid is a human nucleic acid. In one embodiment, the insert nucleic acid is a fragment of a genomic nucleic acid. In one embodiment, the genomic nucleic acid is a mouse genomic nucleic acid, a human genomic nucleic acid, or a combination thereof. In one embodiment, the insert nucleic acid ranges from about 5kb to about 200kb as described above.

**[000111]** In one embodiment, the target genomic locus comprises a conditional allele. In one embodiment, the conditional allele is a multifunctional allele, as described in US 2011/0104799, which is incorporated by reference in its entirety. In one embodiment, the conditional allele comprises: (a) an actuating sequence in sense orientation with respect to transcription of a target gene, and a drug selection cassette in sense or antisense orientation; (b) in antisense orientation a nucleotide sequence of interest (NSI) and a conditional by inversion module (COIN, which utilizes an exon-splitting intron and an invertible genetrap-like module; see, for example, US 2011/0104799, which is incorporated by reference in its entirety); and (c) recombinable units that recombine upon exposure to a first recombinase to form a conditional allele that (i) lacks the actuating sequence and the DSC, and (ii) contains the NSI in sense orientation and the COIN in antisense orientation.

**[000112]** In one embodiment, the insert nucleic acid comprises a reporter gene operably linked to a promoter, wherein the reporter gene encodes a reporter protein selected from the group consisting of LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, Venus, YPet, enhanced yellow fluorescent protein (EYFP), Emerald, enhanced green fluorescent protein (EGFP), CyPet, cyan fluorescent protein (CFP), Cerulean, T-Sapphire, luciferase, alkaline phosphatase, and a combination thereof. In one embodiment, the reporter gene is expressed under a control of an inducible promoter.

In one embodiment, the reporter gene is expressed under a control of an endogenous promoter. In one embodiment, the reporter gene is expressed under a control of an exogenous promoter. In one embodiment, the reporter gene is expressed in a specific cell type. In one embodiment, the reporter gene is expressed in a tissue-specific manner. In one embodiment, the reporter gene is expressed in a developmental stage-specific manner.

**[000113]** In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence encoding a protein expressed in the nervous system, the skeletal system, the digestive system, the circulatory system, the muscular system, the respiratory system, the cardiovascular system, the lymphatic system, the endocrine system, the urinary system, the reproductive system, or a combination thereof. In one embodiment, the human nucleic acid sequence encodes a protein expressed in a bone marrow or a bone marrow-derived cell. In one embodiment, the genome of the mouse ES cell comprises a human genomic locus that encodes a protein expressed in a spleen cell. In one embodiment, the human nucleic acid encodes a protein expressed in a B cell. In one embodiment, the human nucleic acid encodes a protein expressed in an immature B cell. In one embodiment, the human nucleic acid encodes a protein expressed in a mature B cell.

**[000114]** In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence that encodes a human immunoglobulin heavy chain variable region amino acid sequence. In one embodiment, the human nucleic acid sequence comprises an unrearranged human immunoglobulin heavy chain variable region nucleic acid sequence. In one embodiment, the unrearranged human immunoglobulin heavy chain variable region nucleic acid sequence is operably linked to a mouse immunoglobulin heavy chain constant region nucleic acid sequence, a human immunoglobulin heavy chain constant region nucleic acid sequence, or a combination thereof. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is selected from a $C_H1$, a hinge, a $C_H2$, a $C_H3$, and a combination thereof. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a $C_H1$-hinge- $C_H2$- $C_H3$. In one embodiment, the human nucleic acid sequence comprises a rearranged human immunoglobulin heavy chain variable region nucleic acid sequence. In one embodiment, the human immunoglobulin heavy chain variable region nucleic acid sequence is operably linked to a mouse immunoglobulin heavy chain constant region nucleic acid sequence, a human immunoglobulin heavy chain constant region nucleic acid sequence, or a combination thereof. In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is selected from a $C_H1$, a hinge, a $C_H2$, a $C_H3$, and a combination thereof. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a $C_H1$-hinge- $C_H2$- $C_H3$.

**[000115]** In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence that encodes a human immunoglobulin light chain variable region amino acid sequence. In one embodiment, the human nucleic acid sequence comprises an unrearranged human λ and/or κ light chain variable region nucleic acid sequence. In one embodiment, the human nucleic acid sequence comprises a rearranged human λ and/or κ light chain variable region nucleic acid sequence. In one embodiment, the unrearranged λ and/or κ light chain variable region nucleic acid sequence is operably linked to a mouse or human immunoglobulin light chain constant region nucleic acid sequence selected from a λ light chain constant region nucleic acid sequence and a κ light chain constant region nucleic acid sequence.

**[000116]** In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence. In one embodiment, the human nucleic acid sequence encodes an extracellular protein. In one embodiment, the human nucleic acid sequence encodes a ligand for a receptor. In one embodiment, the ligand is a cytokine. In one embodiment, the cytokine is a chemokine selected from CCL, CXCL, CX3CL, and XCL. In one embodiment, the cytokine is a tumor necrosis factor (TNF). In one embodiment, the cytokine is an interleukin (IL). In one embodiment, the interleukin is selected from IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IL-34, IL-35, and IL-36. In one embodiment, the interleukin is IL-2. In one embodiment, the human genomic nucleic acid sequence encodes a cytoplasmic protein. In one embodiment, the human genomic nucleic acid sequence encodes a membrane protein. In one embodiment, the membrane protein is a receptor. In one embodiment, the receptor is a cytokine receptor. In one embodiment, the cytokine receptor is an interleukin receptor. In one embodiment, the interleukin receptor is an interleukin 2 receptor alpha. In one embodiment, the interleukin receptor is an interleukin 2 receptor beta. In one embodiment, the interleukin receptor is an interleukin 2 receptor gamma. In one embodiment, the human genomic nucleic acid sequence encodes a nuclear protein. In one embodiment, the nuclear protein is a nuclear receptor.

**[000117]** In one embodiment, the insert nucleic acid comprises a genetic modification in a coding sequence. In one embodiment, the genetic modification comprises a deletion mutation of a coding sequence. In one embodiment, the genetic modification comprises a fusion of two endogenous coding sequences.

**[000118]** In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence encoding a mutant human protein. In one embodiment, the mutant human protein is characterized by an altered binding characteristic, altered localization, altered expression,

and/or altered expression pattern. In one embodiment, the human nucleic acid sequence comprises at least one human disease allele. In one embodiment, the human disease allele is an allele of a neurological disease. In one embodiment, the human disease allele is an allele of a cardiovascular disease. In one embodiment, the human disease allele is an allele of a kidney disease. In one embodiment, the human disease allele is an allele of a muscle disease. In one embodiment, the human disease allele is an allele of a blood disease. In one embodiment, the human disease allele is an allele of a cancer-causing gene. In one embodiment, the human disease allele is an allele of an immune system disease. In one embodiment, the human disease allele is a dominant allele. In one embodiment, the human disease allele is a recessive allele. In one embodiment, the human disease allele comprises a single nucleotide polymorphism (SNP) allele.

**[000119]** In one embodiment, the insert nucleic acid comprises a regulatory sequence. In one embodiment, the regulatory sequence is a promoter sequence. In one embodiment, the regulatory sequence is an enhancer sequence. In one embodiment, the regulatory sequence is a transcriptional repressor-binding sequence. In one embodiment, the insert nucleic acid comprises a human nucleic acid sequence, wherein the human nucleic acid sequence comprises a deletion of a non-protein-coding sequence, but does not comprise a deletion of a protein-coding sequence. In one embodiment, the deletion of the non-protein-coding sequence comprises a deletion of a regulatory sequence. In one embodiment, the deletion of the regulatory element comprises a deletion of a promoter sequence. In one embodiment, the deletion of the regulatory element comprises a deletion of an enhancer sequence.

**[000120]** In one aspect, a method for making a non-human animal comprising in its germline one or more genetic modifications as described herein is provided, comprising:

**[000121]** (a) modifying a genomic locus of interest of a non-human animal in a prokaryotic cell employing a large targeting vector (LTVEC) and a nuclease agent that generates a single or double-strand break at or near the genomic locus of interest, wherein the LTVEC comprises an insert nucleic acid flanked with upstream and downstream homology arms, and the prokaryotic cell is capable of expressing a recombinase;

**[000122]** (b) selecting a modified prokaryotic cell comprising a genetically modified LTVEC;

**[000123]** (c) isolating the genetically modified LTVEC;

**[000124]** (d) introducing the genetically modified LTVEC into a pluripotent cell of the non-

human animal to generate a genetically modified pluripotent cell comprising the insert nucleic acid in the genomic locus of interest;

**[000125]** (e) selecting the genetically modified pluripotent cell;

**[000126]** (f) introducing the genetically modified pluripotent cell into a host embryo of the non-human animal at a pre-morula stage; and

**[000127]** (g) implanting the host embryo comprising the genetically modified pluripotent cell into a surrogate mother to generate an F0 generation derived from the genetically modified pluripotent cell.

**[000128]** In one embodiment, the non-human animal is a mammal. In one embodiment, the mammal is a rodent. In one embodiment, the rodent is selected form a mouse, a rat, and a hamster. In one embodiment, the non-human animal is a mouse, and the pluripotent cell is a mouse ES cell. In one embodiment, the non-human animal is a rat, and the pluripotent cell is a rat ES cell.

**[000129]** In one embodiment, the target genomic locus comprises one or more genetic modifications as described herein.

**[000130]** In one embodiment, isolating step (c) further comprises (c)' linearizing the genetically modified LTVEC.

**[000131]** In one embodiment, introducing step (d) further comprises (d)' introducing a nuclease agent as described herein into the pluripotent cell. In one embodiment, the nuclease agent is a zinc finger nuclease (ZFN). In one embodiment, the nuclease agent is a Transcription Activator-Like Effector Nuclease (TALEN).

**[000132]** In one embodiment, selecting steps (b) and/or (e) are carried out by applying a selectable agent as described herein to the prokaryotic cell or the pluripotent cell.

**[000133]** In one embodiment, selecting steps (b) and/or (e) are carried out via a modification of allele (MOA) assay as described herein.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[000134]** **FIG. 1** illustrates the Zinc Finger Nuclease (ZFN) cleavage site on the mouse *Il2rg* gene. The figure was not scaled proportionally. The boxed sequences in the bottom panel represent ZFN target sequences.

## DETAILED DESCRIPTION

**[000135]** This invention is not limited to particular methods, and experimental conditions described, as such methods and conditions may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention is defined by the claims.

**[000136]** Unless defined otherwise, all terms and phrases used herein include the meanings that the terms and phrases have attained in the art, unless the contrary is clearly indicated or clearly apparent from the context in which the term or phrase is used. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, particular methods and materials are now described. All publications mentioned are hereby incorporated by reference.

### Definitions

**[000137]** The term "embryonic stem cell" or "ES cell" as used herein includes an embryo-derived totipotent or pluripotent cell that is capable of undifferentiated proliferation *in vitro*, and is capable of contributing to any tissue of the developing embryo upon introduction into an embryo. The term "pluripotent cell" as used herein includes an undifferentiated cell that possesses the ability to develop into more than one differentiated cell type.

**[000138]** The term "germline" in reference to an immunoglobulin nucleic acid sequence includes a nucleic acid sequence that can be passed to progeny.

**[000139]** The phrase "heavy chain," or "immunoglobulin heavy chain" includes an immunoglobulin heavy chain sequence, including immunoglobulin heavy chain constant region sequence, from any organism. Heavy chain variable domains include three heavy chain CDRs and four framework (FR) regions, unless otherwise specified. Fragments of heavy chains include CDRs, CDRs and FRs, and combinations thereof. A typical heavy chain has, following the variable domain (from N-terminal to C-terminal), a $C_H1$ domain, a hinge, a $C_H2$ domain, and a $C_H3$ domain. A functional fragment of a heavy chain includes a fragment that is capable of specifically recognizing an epitope (e.g., recognizing the epitope with a $K_D$ in the micromolar, nanomolar, or picomolar range), that is capable of expressing and secreting from a cell, and that comprises at least one CDR. Heavy chain variable domains are encoded by variable region nucleotide sequence, which generally comprises $V_H$, $D_H$, and $J_H$ segments derived from a repertoire of $V_H$, $D_H$, and $J_H$ segments present in the germline. Sequences, locations and nomenclature for V, D, and J heavy chain segments for various organisms can be found in IMGT database, which is accessible via the internet on

the world wide web (www) at the URL "imgt.org."

**[000140]** The phrase "light chain" includes an immunoglobulin light chain sequence from any organism, and unless otherwise specified includes human kappa (κ) and lambda (λ) light chains and a VpreB, as well as surrogate light chains. Light chain variable domains typically include three light chain CDRs and four FRs, unless otherwise specified. Generally, a full-length light chain includes, from amino terminus to carboxyl terminus, a variable domain that includes FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4, and a light chain constant region amino acid sequence. Light chain variable domains are encoded by the light chain variable region nucleotide sequence, which generally comprises light chain $V_L$ and light chain $J_L$ gene segments, derived from a repertoire of light chain V and J gene segments present in the germline. Sequences, locations and nomenclature for light chain V and J gene segments for various organisms can be found in IMGT database, which is accessible via the internet on the world wide web (www) at the URL "imgt.org." Light chains include those, *e.g.*, that do not selectively bind either a first or a second epitope selectively bound by the epitope-binding protein in which they appear. Light chains also include those that bind and recognize, or assist the heavy chain with binding and recognizing, one or more epitopes selectively bound by the epitope-binding protein in which they appear.

**[000141]** The term "homologous nucleic acid" as used herein includes a nucleic acid sequence that is either identical or substantially similar to a known reference sequence. In one embodiment, the term "homologous nucleic acid" is used to characterize a DNA or RNA sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or even 100% identical to a known reference sequence.

**[000142]** The term "orthologous nucleic acid" as used herein includes a nucleic acid sequence from one species that is functionally equivalent to a known reference sequence in another species.

**[000143]** The term "large targeting vector" or "LTVEC" as used herein includes large targeting vectors for eukaryotic cells that are derived from fragments of cloned genomic nucleic acid larger than those typically used by other approaches intended to perform homologous targeting in eukaryotic cells. The size of the LTVEC is too large to enable screening of targeting events by conventional assays, *e.g.*, southern blotting and long-range (*e.g.*, 1kb-5kb) PCR. Examples of the LTVEC, include, but are not limited to, vectors derived from a bacterial artificial chromosome (BAC) and a yeast artificial chromosome (YAC).

**[000144]** The term "modification of allele" or "MOA" includes the modification of the exact DNA sequence of one allele of a gene(s) or chromosomal locus (loci) in a genome. Examples of "modification of allele (MOA)" include, but are not limited to, deletions, substitutions, or insertions of as little as a single nucleotide or deletions of many kilobases spanning a gene(s) or chromosomal locus (loci) of interest, as well as any and all possible modifications between these two extremes.

**[000145]** The term "nuclease" as used herein includes an agent that induces a break in a nucleic acid sequence, *e.g.*, a single or a double strand break in a double-stranded DNA sequence. Nucleases include those which bind a preselected or specific sequence and cut at or near the preselected or specific sequence, *e.g.*, engineered zinc finger nucleases and engineered TAL effector nucleases. Nucleases are not limited to ZFNs and TAL effector nuclease, but can be any nuclease suitable for use with an LTVEC to achieve improved targeting efficiency. Non-limiting examples include other zinc finger-based nucleases and engineered meganucleases that cut at preselected or specific sequences.

**[000146]** TAL effector nucleases suitable for use with the invention include any TAL nucleases known in the art. Examples of suitable TAL nucleases, and methods for preparing suitable TAL nucleases, are disclosed, e.g., in US Patent Application No. 2011/0239315 A1, 2011/0269234 A1, 2011/0145940 A1, 2003/0232410 A1, 2005/0208489 A1, 2005/0026157 A1, 2005/0064474 A1, 2006/0188987 A1, and 2006/0063231 A1 (each hereby incorporated by reference). In various embodiments, TAL effector nucleases are engineered that cut in or near a target nucleic acid sequence in, *e.g.*, a genome of interest, wherein the target nucleic acid sequence is at or near a sequence to be modified by an LTVEC. TAL effector nucleases are proteins that comprise an endonuclease domain and one or more TAL effector DNA binding domains, wherein the one or more TAL effector DNA binding domains comprise a sequence that recognizes a preselected or specific nucleic acid sequence. The TAL nucleases suitable for use with the invention include those that are specifically designed to bind at or near target nucleic acid sequences to be modified by LTVECs as described herein.

**[000147]** The phrase "operably linked" includes a relationship wherein the components operably linked function in their intended manner. In one instance, a nucleic acid sequence encoding a protein may be operably linked to regulatory sequences (*e.g.*, promoter, enhancer, silencer sequence, *etc.*) so as to retain proper transcriptional regulation. In one instance, a nucleic acid sequence of an immunoglobulin variable region (or V(D)J segments) may be operably linked to a nucleic acid sequence of an immunoglobulin constant region so as to allow proper recombination between the sequences into an immunoglobulin heavy or

light chain sequence.

**[000148]** The term "promoter" and "promoter regulatory element", and the like, as used herein include a nucleotide sequence element within a nucleic acid fragment or gene that controls the expression of that gene.

**[000149]** The term "recombination site" as used herein includes a nucleotide sequence that is recognized by a site-specific recombinase and that can serve as a substrate for a recombination event.

**[000150]** The term "site-specific recombinase" as used herein includes a group of enzymes that can facilitate recombination between "recombination sites" where the two recombination sites are physically separated within a single nucleic acid molecule or on separate nucleic acid molecules. Examples of "site-specific recombinase" include, but are not limited to, Cre, Flp, and Dre recombinases.

**[000151] Modification of Genomic Loci Using an LTVEC and a Nuclease Agent**

**[000152]** While progress has been made in targeting various genomic loci of non-human animals, there still remain many genomic loci or cell types that cannot be targeted with conventional targeting strategies. The reasons for the failure may vary, but, as used herein, include loci or cells that are either not targeted successfully at all, or are targeted improperly or at a significantly low efficiency by conventional targeting methods. Conventional targeting methods include targeting using homologous recombination employing conventional targeting vectors. Loci that are difficult to target include loci that cannot be targeted even with LTVECs alone, *i.e.*, in the absence of assistance in the form of a recombinogenic single or double strand-break, or that are targeted with LTVECs improperly or at a low efficiency in the absence of the recombinogenic single or double-strand break.

**[000153]** Compositions and methods are provided for targeting nucleic acid sequences employing a nuclease agent capable of forming a recombinogenic single or double-strand break, in conjunction with a large targeting vector, or LTVEC, wherein the targeted nucleic acid sequence (or a sequence near the targeted nucleic acid sequence) is modified by the LTVEC. The compositions and methods are useful for modifying genomic nucleic acid sequences that are difficult or impossible to modify using conventional targeting strategies, even when using LTVECs alone.

**[000154]** In various aspects, compositions and methods are provided for employing an LTVEC to make a modification to a target nucleic acid, *e.g.*, a locus in a genome, wherein

the target nucleic acid comprises a target sequence that is to be modified by a sequence of the LTVEC (by homologous recombination of the target nucleic acid with the LTVEC), wherein a single or a double-strand break is made in the target nucleic acid at or near the target sequence.

**[000155]** The presence of a single or a double strand break at or near the target nucleic acid, in various embodiments, increases the efficiency and/or frequency of recombination between an LTVEC and a target nucleic acid. In one embodiment the recombination is homologous recombination. In another embodiment the recombination is an insertion by non-homologous end joining. In various embodiments, in the presence of the single or double strand bread, targeting efficiency of an LTVEC sequence at the target genomic locus is at least about 2-fold higher, at least about 3-fold higher, at least about 4-fold higher than in the absence of the single or double-strand break (using, *e.g.*, the same LTVEC and the same target nucleic acid comprising the same target sequence but in the absence of an added nuclease that makes the single or double strand break).

**[000156]** LTVECs suitable for use with the invention, and methods for making them, are described, *e.g.*, in US Pat. No. 6,586,251, 6,596,541, 7,105,348, and WO 2002/036789 (PCT/US01/45375).

**[000157]** Although embodiments directed to introducing an LTVEC into a mouse pluripotent cell, *e.g.*, a mouse ES cell, are extensively discussed, other methods that introduce an LTVEC into a variety of mammalian cell types are also provided herein. Such mammalian cells include any mammalian cells that can be genetically modified according to the method as disclosed herein, including, *e.g.*, a mouse cell, a rat cell, a rabbit cell, a pig cell, a bovine cell, a deer cell, a sheep cell, a goat cell, a chicken cell, a cat cell, a dog cell, a ferret cell, a primate (*e.g.*, marmoset, rhesus monkey) cell, and the like. In some embodiments, for those mammals for which suitable genetically modifiable pluripotent cells are not readily available, other methods are employed in order to reprogram somatic cells into pluripotent cells, *e.g.*, via introduction into somatic cells of a combination of pluripotency-inducing factors, including, but not limited to, Oct3/4, Sox2, KLF4, Myc, Nanog, LIN28, and Glis1.

**[000158]** In one embodiment, the upstream and downstream homology arms are from the same genome as the targeted genome. In one embodiment, the homology arms are from a related genome, *e.g.*, the targeted genome is a mouse genome of a first strain, and the targeting arms are from a mouse genome of a second strain, wherein the first strain and the second strain are different. In one embodiment, the targeting arms are derived from a BAC

library, a cosmid library, or a P1 phage library. In one embodiment, the homology arms are derived from a synthetic DNA. In one embodiment, the homology arms are derived from a gene that is not targetable using conventional methods. In a specific embodiment, the homology arms are from a gene that cannot be targeted using conventional targeting technology, or can be targeted only incorrectly or only with significantly low efficiency, in the absence of a single or double-strand break induced by a nuclease agent.

**[000159]** In various embodiments, in order to facilitate identification of the targeted modification, a high-throughput quantitative assay, namely, modification of allele (MOA) assay, is employed. The MOA assay described herein allows a large-scale screening of a modified allele(s) in a parental chromosome following a genetic modification. The MOA assay can be carried out via various analytical techniques, including, but not limited to, a quantitative PCR, *e.g.*, a real-time PCR (qPCR). For example, the real-time PCR comprises a first primer set that recognizes the target locus and a second primer set that recognizes a non-targeted reference locus. In addition, the primer set comprises a fluorescent probe that recognizes the amplified sequence. The quantitative assay can also be carried out via a variety of analytical techniques, including, but not limited to, fluorescence-mediated in situ hybridization (FISH), comparative genomic hybridization, isothermic DNA amplification, quantitative hybridization to an immobilized probe(s), Invader Probes®, MMP assays®, TaqMan® Molecular Beacon, and Eclipse™ probe technology. (See, for example, US2005/0144655, incorporated by reference herein in its entirety).

**[000160]** In some embodiments, various genetic modifications of the target genomic loci described herein can be carried out by a series of homologous recombination reactions (BHR) in bacterial cells using an LTVEC derived from Bacterial Artificial Chromosome (BAC) DNA using VELOCIGENE® genetic engineering technology (see, *e.g.*, US Pat. No. 6,586,251 and Valenzuela, D. M. *et al.* (2003), High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, *Nature Biotechnology* 21(6): 652-659, which is incorporated herein by reference in their entireties).

**[000161]** In some embodiments, targeted mouse ES cells comprising various genetic modifications as described herein are used as donor ES cells and introduced into a pre-morula stage mouse embryo, *e.g.*, an 8-cell stage mouse embryo, via the VELOCIMOUSE® method (see, *e.g.*, US 7,576,259, US 7,659,442, US 7,294,754, and US 2008-0078000 A1, all of which are incorporated by reference herein in their entireties). The mouse embryo comprising the genetically modified ES cells is incubated until the blastocyst stage and then implanted into a surrogate mother to produce an F0 mouse. Mice bearing the genetically modified genomic locus can be identified via modification of allele (MOA) assay as described

herein. The resulting F0 generation mouse derived from the genetically modified ES cells is crossed to a wild-type mouse to obtain F1 generation offspring. Following genotyping with specific primers and/or probes, F1 pups that are heterozygous for the genetically modified genomic locus are crossed to each other to produce mice that are homozygous for the genetically modified genomic locus.

**[000162]** It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural references unless the context clearly dictates otherwise. All technical and scientific terms used herein have the same meaning.

**[000163]** The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the described invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates, which may need to be independently confirmed.

**[000164]** The described invention may be embodied in other specific forms without departing from the spirit or essential attributes thereof and, accordingly, reference should be made to the appended claims, rather than to the foregoing specification, as indicating the scope of the invention.

## EXAMPLES

**[000165]** The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.*, amounts, temperature, *etc.*) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

### Example 1. Enhanced LTVEC targeting by a Zinc Finger Nuclease (ZFN)

**[000166]** In order to test if inducing a double-strand break in a gene by a zinc finger nuclease (ZFN) could enhance the targeting of the same gene by an LTVEC (a large BAC-based targeting vector), three electroporations were performed into F1H4 ES cells with an LTVEC targeting the *Il2rg* gene and plasmids encoding each half of a ZFN pair in the following combinations: (1) 1.5 µg of *Il2rg* LTVEC alone; (2) 20 µg of *Il2rg* ZFN-1 + 20 µg of

*Il2rg* LTVEC ZFN-2 + 1.5 µg of *Il2rg* LTVEC; and (3) 20 µg of *Il2rg* ZFN-1 + 20 µg of *Il2rg* ZFN-2 without LTVEC.

**[000167]** ZFN-1 and ZFN-2 used herein were designed to contain (1) zinc finger DNA binding domains that recognize two contiguous target DNA sequences in each strand of the target sequence separated by a 6bp cleavage site; and (2) FokI nucleases that dimerize and make a double-strand break at the target site. More specifically, the zinc finger domain of ZFN-1 was designed to recognize 5'-AGCTCCAAGGTCCTC-3' (SEQ ID NO: 1) in the sense strand of exon 1; and the zinc finger domain of ZFN-2 was designed to recognize 5'-GTCTTCATTCGCACT-3' (SEQ ID NO: 2) in the anti-sense strand of exon 1 in the *Il2rg* gene.

**[000168]** The LTVEC (VelociGene MAID 5057L1) was designed to delete the approximately 3,000 base pair (3kb) of the *Il2rg* gene, including all of the coding sequence for the IL-2 receptor gamma chain and to replace the endogenous sequence with a selection cassette that expresses hygromycin phosphotransferase, which imparts resistance to hygromycin B. LTVEC 5057L1 is derived from parental BAC clone 290o15 isolated from a commercially available BAC library (BAC ES Release 2; Incyte Genomics) and contains two large homology arms with about 90kb (homology arm 1) and about 8 kb (homology arm 2).

**[000169]** For electroporations (1) and (2), hygromycin B-resistant colonies were isolated. For electroporation (3), colonies were allowed to form without drug selection. Colonies were picked from each electroporation and cultured in 96-well plates, followed by screening for targeting events by loss-of-allele (LOA) assays designed to detect the 3kb deletion created by correct targeting of the *Il2rg* gene by the LTVEC. LOA assays were also used to detect cleavage events induced by the ZFN in exon 1.

**[000170] Table 1. LOA assays for the targeted 3kb deletion**

| | | |
|---|---|---|
| **5057U2 assay (intron 2)** | | |
| Forward Primer | 5′-GGAGGGTAGCACGGGAAGAAG-3′ | SEQ ID NO: 3 |
| Reverse Primer | 5′-GCTGGCTACCCACTTGATTGG-3′ | SEQ ID NO: 4 |
| TaqMan® probe | 5′-TCAAGCAGTCTCTCCCAGCTAACCTCCCT-3′ | SEQ ID NO: 5 |
| **5057D2 assay (intron 7)** | | |
| Forward Primer | 5′-CAGGATGTGGCTGACCAAATG-3′ | SEQ ID NO: 6 |
| Reverse Primer | 5′-GGCTCCTAATGCCCTGTAGTTTC-3′ | SEQ ID NO: 7 |
| TaqMan® probe | 5′-CCGTCTCTCTGCCTAGCCCACCCT-3′ | SEQ ID NO: 8 |

**[000171] Table 2. LOA assay for the ZFN cleavage**

| | | |
|---|---|---|
| Forward Primer | 5′-CAGCTGCTCCTGCTGAGG-3′ | SEQ ID NO: 9 |
| Reverse Primer | 5′-CCTACCAGCTTTGATGTCTTCATTC-3′ | SEQ ID NO: 10 |
| TaqMan® probe | 5′-AGCTCCAAGGTCCTCATGTCCAGT-3′ | SEQ ID NO: 11 |

**[000172] Table 3. Comparison of Targeting Efficiency of LTVECs**

| EP | LTVEC (μg) | ZFN-1 (μg) | ZFN-2 (μg) | Number of clones assayed | Clones correctly targeted by LTVEC | **Targeting Efficiency (%)** | Clones with ZFN Cleavage | ZFN cleavage efficiency (%) |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.5 | 0 | 0 | 26 | 5 | **19.2** | ND[1] | ND[1] |
| 2 | 1.5 | 20 | 20 | 102 | 76 | **74.5** | 7[2] | 6.9 |
| 3 | 0 | 20 | 20 | 192 | 0 | **0** | 6 | 3.1 |

[1] not determined
[2] among the clones that were not correctly targeted by the LTVEC

**[000173]** As shown in Table 3, when plasmids encoding a ZFN designed to cleave a site in a target gene were combined with an LTVEC that targets the same gene (EP #2), a significant enhancement (approximately about 4-fold increase in the experiment described) of targeting efficiency was achieved compared to the LTVEC alone (EP #1).

**[000174]** While the described invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adopt a particular situation, material, composition of matter, process, process step or steps, to the objective spirit and scope of the described invention. All such modifications are intended to be within the scope of the claims appended hereto.

* * * * *

## CLAIMS

What is claimed is:

1. A method for modifying a target genomic locus in a mouse embryonic stem (ES) cell, comprising:

(a) introducing into the mouse ES cell:

(i) a zinc finger nuclease (ZFN) that makes a single or double-strand break at or near the target genomic locus; and

(ii) a large targeting vector (LTVEC) comprising an insert nucleic acid flanked by an upstream homology arm and a downstream homology arm, wherein the insert nucleic acid ranges from about 5 kb to about 30 kb in length, and the sum total of the upstream and downstream homology arms is at least 10 kb; and

(b) selecting a targeted mouse ES cell comprising the insert nucleic acid in the target genomic locus, wherein integration of the insert nucleic acid into the target genomic locus results in deletion of an endogenous nucleic acid sequence at the target genomic locus and replacement with the insert nucleic acid, wherein combined use of the LTVEC with the ZFN results in an increased targeting efficiency compared to use of the LTVEC alone.

2. The method of claim 1, wherein the targeting efficiency of the LTVEC is increased at least 2-fold compared to use of the LTVEC alone.

3. The method of claim 1 or 2, wherein an expression construct comprising a nucleic acid sequence encoding the ZFN is introduced into the mouse ES cell, and wherein the nucleic acid sequence is operably linked to a promoter active in the mouse ES cell.

4. The method of claim 1 or 2, wherein an mRNA encoding the ZFN is introduced into the mouse ES cell.

5. The method of any one of claims 1-4, wherein a target sequence of the ZFN is located in an intron, exon, a promoter, a promoter regulatory region, or an enhancer region in the target genomic locus.

6. The method of any one of claims 1-5, wherein the insert nucleic acid

comprises a selection cassette, a reporter gene, a nucleic acid flanked by site-specific recombination target sequences, or a conditional allele.

7. The method of any one of claims 1-6, wherein the insert nucleic acid comprises a nucleic acid sequence from a different species.

8. The method of claim 7, wherein the insert nucleic acid comprises a human nucleic acid sequence.

9. The method of claim 8, wherein the insert nucleic acid comprises a genomic nucleic acid sequence that encodes a human immunoglobulin heavy chain variable region amino acid sequence.

10. The method of claim 8, wherein the insert nucleic comprises a genomic nucleic acid sequence that encodes a human immunoglobulin light chain variable region amino acid sequence.

11. The method of claim 10, wherein the genomic nucleic acid sequence comprises an unrearranged human λ and/or κ light chain variable region nucleic acid sequence.

12. The method of claim 10, wherein the genomic nucleic acid sequence comprises a rearranged human λ and/or κ light chain variable region nucleic acid sequence.

13. The method of claim 8, wherein the insert nucleic acid comprises a human disease allele.

14. The method of claim 9, wherein the insert nucleic acid comprises a rearranged human immunoglobulin heavy chain variable region nucleic acid sequence operably linked to an immunoglobulin heavy chain constant region nucleic acid sequence.

15. The method of claim 9, wherein the insert nucleic acid comprises an unrearranged human immunoglobulin heavy chain variable region nucleic acid sequence operably linked to an immunoglobulin heavy chain constant region nucleic acid sequence.

16. The method of claim 14 or 15, wherein the immunoglobulin heavy chain

constant region sequence is a mouse immunoglobulin heavy chain constant region sequence, a human immunoglobulin heavy chain constant region sequence, or a combination thereof.

17. The method of any one of claims 14-16, wherein the immunoglobulin heavy chain constant region nucleic acid sequence is selected from a $C_H1$, a hinge, a $C_H2$, a $C_H3$, and a combination thereof.

18. The method of claim 11, wherein the insert nucleic acid comprises an unrearranged human λ or κ light chain variable region nucleic acid sequence operably linked to a mouse or human immunoglobulin light chain constant region nucleic acid sequence selected from a λ light chain constant region nucleic acid sequence and a κ light chain constant region nucleic acid sequence.

19. The method of claim 12, wherein the insert nucleic acid comprises a rearranged human λ or κ light chain variable region nucleic acid sequence operably linked to a mouse or human immunoglobulin light chain constant region nucleic acid sequence selected from a λ light chain constant region nucleic acid sequence and a κ light chain constant region nucleic acid sequence.

20. The method of any one of claims 1-19, wherein the insert nucleic acid comprises a nucleic acid sequence homologous or orthologous to the replaced nucleic acid sequence at the target genomic locus of the mouse ES cell.

21. The method of any one of claims 1-20, wherein integration of the insert nucleic acid into the target genomic locus results in a knockout, a knock-in, a point mutation, a domain swap, an exon swap, an intron swap, a regulatory sequence swap, a gene swap, or a combination thereof.

22. The method of any one of claims 1-21, wherein selecting step (b) is carried out via a modification of allele (MOA) assay.

23. The method of any one of claims 1-22, wherein the LTVEC is from about 50 kb to about 300 kb.

24. The method of any one of claims 1-23, wherein the sum total of the

upstream and the downstream homology arms is from 10 kb to about 200 kb.

25. The method of any one of claims 1-24, wherein each of the upstream and downstream homology arms is between about 5 kb and about 200 kb in length.

26. The method of any one of claims 1-25, wherein each of the upstream and downstream homology arms is between about 5 kb and about 200 kb in length, and wherein the LTVEC ranges from about 50 kb to about 300 kb in length.

27. The method of any one of claims 1-26, wherein integration of the insert nucleic acid into the target genomic locus results in a deletion of an endogenous gene at the target genomic locus.

28. The method of any one of claims 1-27, further comprising:

(a) introducing the modified mouse ES cell into a pre-morula stage embryo;

(b) incubating the embryo until the blastocyst stage and implanting the embryo into a surrogate mother to produce an F0 mouse; and

(c) identifying a mouse bearing the genetically modified genomic locus.