METHODS AND COMPOSITIONS FOR THE TARGETED MODIFICATION OF A GENOME

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 61/914,768, filed December 11, 2013, U.S. Provisional Patent Application No. 62/017,416, filed June 26, 2014, U.S. Provisional Patent Application No. 62/029,261, filed July 25, 2014, U.S. Provisional Patent Application No. 62/052,906, filed September 19, 2014, U.S. Provisional Patent Application No. 62/059,527, filed October 03, 2014, and U.S. Provisional Patent Application No. 62/064,384, filed October 15, 2014, each of which is incorporated by reference in its entirety for all purposes.

REFERENCE TO A SEQUENCE LISTING SUBMITTED AS A TEXT FILE VIA EFS WEB

[0002] The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file named 453460SEQLIST.TXT, created on October 15, 2014, and having a size of 27.5 kilobytes, and is filed concurrently with the specification. The sequence listing contained in this ASCII formatted document is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

[0003] While rats have been regarded as an important animal model system that can recapitulate the pathology of various human diseases, including, but not limited to, cardiovascular (*e.g.*, hypertension), metabolic (*e.g.*, obesity, diabetes), neurological (*e.g.*, pain pathologies), and a variety of cancers, the use of rats in modeling human diseases has been limited as compared to mice, due in part to unavailability of germline-transmittable pluripotent rat cells, which can sustain their pluripotency following a series of genetic modifications *in vitro*, *e.g.*, one or more serial electroporations, and due in part

to lack of efficient targeting technologies that allow introduction or deletion of large genomic DNA sequences, or replacement of large endogenous genomic DNA sequences with exogenous nucleic acid sequences in pluripotent rat cells.

[0004] There is a need in the art for compositions and methods that allow precise targeted changes in the genome of an organism, which can open or expand current areas of target discovery and validate therapeutic agents more quickly and easily.

SUMMARY

[0005] Methods are provided for modifying a genomic locus of interest in a eukaryotic cell via targeted genetic modification. Such a method comprises (a) introducing into the eukaryotic cell: (i) a large targeting vector (LTVEC) comprising a first nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the LTVEC is at least 10 kb; (ii) a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding a Cas protein, (iii) a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding a guide RNA (gRNA) comprising a nucleotide sequence that hybridizes to a target sequence and a trans-activating CRISPR RNA (tracrRNA), wherein the first and the second promoters are active in the eukaryotic cell; and (b) identifying a modified eukaryotic cell comprising a targeted genetic modification at the genomic locus of interest.

[0006] In one embodiment, the targeted genetic modification is a biallelic genetic modification.

[0007] In one embodiment, the LTVEC is at least 15 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, or at least 90 kb. In another embodiment, the LTVEC is at least 100 kb, at least 150 kb, or at least 200 kb.

[0008] In one embodiment, the eukaryotic cell is a mammalian cell. In one embodiment, the mammalian cell is a fibroblast.

[0009] In one embodiment, the eukaryotic cell is a pluripotent cell. In one embodiment, the pluripotent cell is a human pluripotent cell. In one embodiment the human pluripotent cell is a human embryonic stem (ES) cell or a human adult stem cell.

In another embodiment, the human pluripotent cell is a developmentally restricted human progenitor cell. In another embodiment, the human pluripotent cell is a human induced pluripotent stem (iPS) cell.

[0010] In one embodiment, the Cas protein is Cas9.

[0011] In one embodiment, the target sequence is flanked by a Protospacer Adjacent Motif (PAM) sequence. In one embodiment, the target sequence is immediately flanked on the 3' end by a Protospacer Adjacent Motif (PAM) sequence.

[0012] In some embodiments, the sum total of the 5' and the 3' homology arms is from about 10 kb to about 150 kb. In some embodiments, the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 120 kb, or from about 120 kb to 150 kb.

[0013] The methods further provide that the targeted genetic modification comprises: (a) a replacement of an endogenous nucleic acid sequence with a homologous or an orthologous nucleic acid sequence; (b) a deletion of an endogenous nucleic acid sequence; (c) a deletion of an endogenous nucleic acid sequence, wherein the deletion ranges from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb; (d) insertion of an exogenous nucleic acid sequence; (e) insertion of an exogenous nucleic acid sequence ranging from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb; (f) insertion of an exogenous nucleic acid sequence comprising a homologous or an orthologous nucleic acid sequence; (g)

insertion of a chimeric nucleic acid sequence comprising a human and a non-human nucleic acid sequence; (h) insertion of a conditional allele flanked with site-specific recombinase target sequences; (i) insertion of a selectable marker or a reporter gene operably linked to a third promoter active in the pluripotent cell; or (j) a combination thereof.

[0014] In one embodiment, the genomic locus of interest comprises (i) a 5' target sequence that is homologous to the 5' homology arm; and (ii) a 3' target sequence that is homologous to the 3' homology arm.

[0015] In some embodiments, the 5' target sequence and the 3' target sequence is separated by at least 5 kb but less than 3 Mb. In some embodiments, the 5' target sequence and the 3' target sequence is separated by at least 5 kb but less than 10 kb, at least 10 kb but less than 20 kb, at least 20 kb but less than 40 kb, at least 40 kb but less than 60 kb, at least 60 kb but less than 80 kb, at least about 80 kb but less than 100 kb, at least 100 kb but less than 150 kb, or at least 150 kb but less than 200 kb, at least about 200 kb but less than about 300 kb, at least about 300 kb but less than about 400 kb, at least about 400 kb but less than about 500 kb, at least about 500 kb but less than about 1 Mb, at least about 1 Mb but less than about 1.5 Mb, at least about 1.5 Mb but less than about 2 Mb, at least about 2 Mb but less than about 2.5 Mb, or at least about 2.5 Mb but less than about 3 Mb.

[0016] In one embodiment, the genomic locus of interest comprises the Interleukin-2 receptor gamma locus, the *ApoE* locus, the *Rag1* locus, the *Rag2* locus, or both of the *Rag1* and the *Rag2* loci.

[0017] In one embodiment, the first and the second expression constructs are on a single nucleic acid molecule.

[0018] Further provided is a method for modifying a genome, comprising exposing the genome to a Cas protein and a CRISPR RNA in the presence of a large targeting vector (LTVEC) comprising a nucleic acid sequence of at least 10 kb, wherein following exposure to the Cas protein, the CRISPR RNA, and the LTVEC, the genome is modified to contain at least 10 kb nucleic acid sequence.

[0019] In some such methods, the LTVEC comprises a nucleic acid sequence of at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at

least 80 kb, or at least 90 kb. In some such methods, the LTVEC comprises a nucleic acid sequence of at least 100 kb, at least 150 kb, or at least 200 kb.

[0020] Further provided is a method for modifying a genome, comprising contacting the genome with a Cas protein, a CRISPR RNA that hybridizes to a target sequence, and a tracrRNA in the presence of a large targeting vector (LTVEC), wherein the LTVEC is at least 10 kb and comprises a first nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein following contacting with the Cas protein, CRISPR RNA, and tracrRNA in the presence of the LTVEC, the genome is modified at a genomic locus of interest to contain the first nucleic acid. The target sequence can be at or near the genomic locus of interest.

[0021] In some such methods, the genome is in a eukaryotic cell, and the Cas protein, the CRISPR RNA, the tracrRNA, and the LTVEC are introduced into the eukaryotic cell. Some such methods further comprise identifying a modified eukaryotic cell comprising a targeted genetic modification at the genomic locus of interest.

[0022] In some such methods, the CRISPR RNA and the tracrRNA are introduced together in the form of a single guide RNA (gRNA). In other methods, the CRISPR RNA and the tracrRNA are introduced separately.

[0023] In some such methods (a) the Cas protein is introduced into the eukaryotic cell in the form of a protein, a messenger RNA (mRNA) encoding the Cas protein, or a DNA encoding the Cas protein; (b) the CRISPR RNA is introduced into the eukaryotic cell in the form of an RNA or a DNA encoding the CRISPR RNA; and (c) the tracrRNA is introduced into the eukaryotic cell in the form of an RNA or a DNA encoding the tracrRNA.

[0024] In some methods (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; (b) the DNA encoding the CRISPR RNA is in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding the CRISPR RNA; and (c) the DNA encoding the tracrRNA is in the form of a third expression construct comprising a third promoter operably linked to a fourth nucleic acid encoding the tracrRNA, wherein the first, second, and third promoters

are active in the eukaryotic cell. Optionally, the first, second, and/or third expression constructs are on a single nucleic acid molecule.

[0025] In some methods (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; and (b) the DNA encoding the CRISPR RNA and the DNA encoding the tracrRNA are in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding a gRNA comprising the CRISPR RNA and the tracrRNA; wherein the first and second promoters are active in the eukaryotic cell. Optionally, the first and the second expression constructs are on a single nucleic acid molecule.

[0026] In some methods, the Cas protein, the CRISPR RNA, and the tracrRNA are introduced into the eukaryotic cell as a protein-RNA complex.

[0027] In some methods, the targeted genetic modification comprises simultaneous deletion of an endogenous nucleic acid sequence at the genomic locus of interest and insertion of the first nucleic acid at the genomic locus of interest. In some methods, the deleted endogenous nucleic acid sequence is about 30 kb to about 110 kb, and the inserted first nucleic acid is about 40 kb to about 140 kb. In some methods, the deleted endogenous nucleic acid sequence is about 38 kb to about 110 kb, and the inserted first nucleic acid is about 43 kb to about 134 kb.

[0028] In some methods, the targeted genetic modification is a biallelic genetic modification. Optionally, the biallelic genetic modification comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid at the genomic locus of interest in two homologous chromosomes.

[0029] In some methods, the modified eukaryotic cell is compound heterozygous at the genomic locus of interest. In some methods, the modified eukaryotic cell is hemizygous at the genomic locus of interest. Optionally, the targeted genetic modification at the genomic locus of interest in one chromosome comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid. Optionally, the targeted genetic modification comprises: (1) deletion of an endogenous nucleic acid sequence at the genomic locus of interest in two homologous chromosomes; and (2) insertion of the first nucleic acid into the genomic locus of interest in a first chromosome

and disruption of the genomic locus of interest in a second chromosome. The first chromosome can be one of the two homologous chromosomes, and the second chromosome can be the other homologous chromosome.

[0030] In some methods, the LTVEC is at least 15 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, or at least 90 kb. Optionally, the LTVEC is at least 100 kb, at least 150 kb, or at least 200 kb.

[0031] In some methods, the first nucleic acid is at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb, at least 200 kb, at least 250 kb, or at least 300 kb. In some methods, the first nucleic acid is about 40 kb to about 140 kb. In some methods, the first nucleic acid is about 43 kb to about 134 kb.

[0032] In some methods, the eukaryotic cell is a mammalian cell, a fibroblast, a pluripotent cell, a non-human pluripotent cell, a rodent pluripotent cell, a mouse or rat embryonic stem (ES) cell, a human pluripotent cell, a human embryonic stem (ES) cell, a human adult stem cell, a developmentally restricted human progenitor cell, or a human induced pluripotent stem (iPS) cell.

[0033] In some methods, the Cas protein is Cas9. In some methods, the target sequence is immediately flanked by a Protospacer Adjacent Motif (PAM) sequence.

[0034] In some methods, the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 150 kb. Optionally, the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 120 kb, or from about 120 kb to 150 kb.

[0035] In some methods, the targeted genetic modification comprises: (a) a replacement of an endogenous nucleic acid sequence with a homologous or an orthologous nucleic acid sequence; (b) a deletion of an endogenous nucleic acid sequence; (c) a deletion of an endogenous nucleic acid sequence, wherein the deletion ranges from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150

kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb; (d) insertion of an exogenous nucleic acid sequence; (e) insertion of an exogenous nucleic acid sequence ranging from about 5kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb; (f) insertion of an exogenous nucleic acid sequence comprising a homologous or an orthologous nucleic acid sequence; (g) insertion of a chimeric nucleic acid sequence comprising a human and a non-human nucleic acid sequence; (h) insertion of a conditional allele flanked with site-specific recombinase target sequences; (i) insertion of a selectable marker or a reporter gene operably linked to a third promoter active in the pluripotent cell; or (j) a combination thereof.

[0036] In some methods, the genomic locus of interest comprises (i) a 5' target sequence that is homologous to the 5' homology arm; and (ii) a 3' target sequence that is homologous to the 3' homology arm. Optionally, the 5' target sequence and the 3' target sequence are separated by at least 5 kb but less than 3 Mb. Optionally, the 5' target sequence and the 3' target sequence are separated by at least 5 kb but less than 10 kb, at least 10 kb but less than 20 kb, at least 20 kb but less than 40 kb, at least 40 kb but less than 60 kb, at least 60 kb but less than 80 kb, at least about 80 kb but less than 100 kb, at least 100 kb but less than 150 kb, or at least 150 kb but less than 200 kb, at least about 200 kb but less than about 300 kb, at least about 300 kb but less than about 400 kb, at least about 400 kb but less than about 500 kb, at least about 500 kb but less than about 1Mb, at least about 1 Mb but less than about 1.5 Mb, at least about 1.5 Mb but less than about 2 Mb, at least about 2 Mb but less than about 2.5 Mb, or at least about 2.5 Mb but less than about 3 Mb. Optionally, the 5' target sequence and the 3' target sequence are separated by at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 110 kb, at least 120 kb,

at least 130 kb, at least 140 kb, at least 150 kb, at least 160 kb, at least 170 kb, at least 180 kb, at least 190 kb, or at least 200 kb. In some methods, the 5' and 3' target sequences are separated by about 30 kb to about 110 kb. In some methods, the 5' and 3' target sequences are separated by about 38 kb to about 110 kb.

[0037] In some methods, the genomic locus of interest comprises the Interleukin-2 receptor gamma locus, the *ApoE* locus, the *Rag1* locus, the *Rag2* locus, or both of the *Rag1* and the *Rag2* loci. In other methods, the genomic locus of interest comprises the *Adamts5* locus, the *Trpa1* locus, the *Folh1* locus, or the *Erbb4* locus. In yet other methods, the genomic locus of interest comprises the *Lrp5* locus. In yet other methods, the genomic locus of interest comprises the *C5* (*Hc*) locus, the *Ror1* locus, or the *Dpp4* locus.

[0038] Further provided is a method for producing an F0 generation non-human animal that comprises a targeted genetic modification at a genomic locus of interest, the method comprising: (a) contacting the genome in a non-human ES cell with a Cas protein, a CRISPR RNA, and a tracrRNA in the presence of a large targeting vector (LTVEC) to form a modified non-human ES cell, wherein the LTVEC is at least 10 kb and comprises a first nucleic acid flanked with a 5' homology arm and a 3' homology arm; (b) identifying the modified non-human ES cell comprising the targeted genetic modification at the genomic locus of interest; (c) introducing the modified non-human ES cell into a non-human host embryo; and (d) gestating the non-human host embryo in a surrogate mother, wherein the surrogate mother produces the F0 generation non-human animal comprising the targeted genetic modification at the genomic locus of interest.

[0039] In some such methods, the CRISPR RNA and the tracrRNA are introduced together in the form of a single guide RNA (gRNA). In other such methods, the CRISPR RNA and the tracrRNA are introduced separately.

[0040] In some such methods, (a) the Cas protein is introduced into the non-human ES cell in the form of a protein, a messenger RNA (mRNA) encoding the Cas protein, or a DNA encoding the Cas protein; (b) the CRISPR RNA is introduced into the non-human ES cell in the form of an RNA or a DNA encoding the CRISPR RNA; and (c) the tracrRNA is introduced into the non-human ES cell in the form of an RNA or a DNA encoding the tracrRNA.

[0041] In some such methods, (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; (b) the DNA encoding the CRISPR RNA is in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding the CRISPR RNA; and (c) the DNA encoding the tracrRNA is in the form of a third expression construct comprising a third promoter operably linked to a fourth nucleic acid encoding the tracrRNA, wherein the first, second, and third promoters are active in the non-human ES cell. Optionally, the first, second, and third expression constructs are on a single nucleic acid molecule.

[0042] In some such methods, (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; and (b) the DNA encoding the CRISPR RNA and the DNA encoding the tracrRNA are in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding a gRNA comprising the CRISPR RNA and the tracrRNA; wherein the first and second promoters are active in the non-human ES cell. Optionally, the first and the second expression constructs are on a single nucleic acid molecule.

[0043] In some such methods, the Cas protein, the CRISPR RNA, and the tracrRNA are introduced into the non-human ES cell as a protein-RNA complex.

[0044] In some such methods, the targeted genetic modification comprises simultaneous deletion of an endogenous nucleic acid sequence at the genomic locus of interest and insertion of the first nucleic acid at the genomic locus of interest.

[0045] In some such methods, the targeted genetic modification is a biallelic genetic modification. Optionally, the biallelic genetic modification comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid at the genomic locus of interest in two homologous chromosomes.

[0046] In some such methods, the modified non-human ES cell is compound heterozygous at the genomic locus of interest. In some such methods, the modified non-human ES cell is hemizygous at the genomic locus of interest. Optionally, the targeted genetic modification at the genomic locus of interest in one chromosome comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid.

Optionally, the targeted genetic modification comprises: (1) deletion of an endogenous nucleic acid sequence at the genomic locus of interest in two homologous chromosomes; and (2) insertion of the first nucleic acid into the genomic locus of interest in a first chromosome and disruption of the genomic locus of interest in a second chromosome. The first chromosome can be one of the two homologous chromosomes, and the second chromosome can be the other homologous chromosome.

[0047] In some such methods, the Cas protein is Cas9.

[0048] Further provided are methods for modifying a genome at a genomic locus of interest in a eukaryotic cell, a mouse cell, or a human cell, comprising contacting the genome with a Cas protein, a CRISPR RNA that hybridizes to a target sequence at the genomic locus of interest, and a tracrRNA in the presence of a large targeting vector (LTVEC), wherein the LTVEC is at least 10 kb and comprises a first nucleic acid flanked with a 5' homology arm that is homologous to a 5' target sequence at the genomic locus of interest and a 3' homology arm that is homologous to a 3' target sequence at the genomic locus of interest, wherein the first nucleic acid is at least 30 kb and/or the 5' target sequence and the 3' target sequence are separated by at least 30 kb, wherein following contacting with the Cas protein, the CRISPR RNA, and the tracrRNA in the presence of the LTVEC, the genome is modified to comprise a targeted genetic modification comprising insertion of the first nucleic acid at the genomic locus of interest.

[0049] Any of the above methods can further comprise introducing the Cas protein, the CRISPR RNA, the tracrRNA, and the LTVEC into the eukaryotic cell, the mouse cell, or the human cell. Any of the above methods can further comprise identifying the modified eukaryotic cell, the modified mouse cell, or the modified human cell comprising the targeted genetic modification at the genomic locus of interest.

[0050] In some of the above methods, the CRISPR RNA and the tracrRNA are introduced together in the form of a single transcript. In some of the above methods, the CRISPR RNA and the tracrRNA are introduced separately.

[0051] In some of the above methods, (a) the Cas protein is introduced into the eukaryotic cell, the mouse cell, or the human cell in the form of a protein, a messenger RNA (mRNA) encoding the Cas protein, or a DNA encoding the Cas protein; (b) the

CRISPR RNA is introduced into the eukaryotic cell, the mouse cell, or the human cell in the form of an RNA or a DNA encoding the CRISPR RNA; and (c) the tracrRNA is introduced into the eukaryotic cell, the mouse cell, or the human cell in the form of an RNA or a DNA encoding the tracrRNA. In some of the above methods, the Cas protein, the CRISPR RNA, and the tracrRNA are introduced into the eukaryotic cell, the mouse cell, or the human cell as a protein-RNA complex.

[0052] In some of the above methods, (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; (b) the DNA encoding the CRISPR RNA is in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding the CRISPR RNA; and (c) the DNA encoding the tracrRNA is in the form of a third expression construct comprising a third promoter operably linked to a fourth nucleic acid encoding the tracrRNA; wherein the first, second, and third promoters are active in the eukaryotic cell, the mouse cell, or the human cell. In some of the above methods, the first, second, and/or third expression constructs are on a single nucleic acid molecule.

[0053] In some of the above methods, (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; and (b) the DNA encoding the CRISPR RNA and the DNA encoding the tracrRNA are in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding a gRNA comprising the CRISPR RNA and the tracrRNA in a single transcript; wherein the first and second promoters are active in the eukaryotic cell, the mouse cell, or the human cell. In some of the above methods, the first and the second expression constructs are on a single nucleic acid molecule.

[0054] In some of the above methods, the LTVEC is at least 15 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, or at least 90 kb. In some of the above methods, the LTVEC is at least 100 kb, at least 150 kb, or at least 200 kb.

[0055] In some of the above methods, the first nucleic acid is at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at

least 90 kb, at least 100 kb, at least 150 kb, at least 200 kb, at least 250 kb, or at least 300 kb. In some of the above methods, the first nucleic acid is about 40 kb to about 140 kb.

[0056] In some of the above methods, the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 150 kb. In some of the above methods, the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 120 kb, or from about 120 kb to 150 kb.

[0057] In some of the above methods, the 5' target sequence and the 3' target sequence are separated by at least 5 kb but less than 3 Mb. In some of the above methods, the 5' target sequence and the 3' target sequence are separated by at least 5 kb but less than 10 kb, at least 10 kb but less than 20 kb, at least 20 kb but less than 40 kb, at least 40 kb but less than 60 kb, at least 60 kb but less than 80 kb, at least about 80 kb but less than 100 kb, at least 100 kb but less than 150 kb, or at least 150 kb but less than 200 kb, at least about 200 kb but less than about 300 kb, at least about 300 kb but less than about 400 kb, at least about 400 kb but less than about 500 kb, at least about 500 kb but less than about 1Mb, at least about 1 Mb but less than about 1.5 Mb, at least about 1.5 Mb but less than about 2 Mb, at least about 2 Mb but less than about 2.5 Mb, or at least about 2.5 Mb but less than about 3 Mb. In some of the above methods, the 5' target sequence and the 3' target sequence are separated by at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 110 kb, at least 120 kb, at least 130 kb, at least 140 kb, at least 150 kb, at least 160 kb, at least 170 kb, at least 180 kb, at least 190 kb, or at least 200 kb. In some of the above methods, the 5' target sequence and the 3' target sequence are separated by from about 30 kb to about 110 kb.

[0058] In some of the above methods, the eukaryotic cell is not a rat cell. In some of the above methods, the eukaryotic cell is a pluripotent cell, a non-pluripotent cell, a mammalian cell, a human cell, a non-human mammalian cell, a rodent cell, a mouse cell, a hamster cell, a non-human pluripotent cell, a human pluripotent cell, a rodent pluripotent cell, or a fibroblast. In some of the above methods, the eukaryotic cell is a

primary cell or an immortalized cell. In some of the above methods, the rodent pluripotent cell is a mouse or rat embryonic stem (ES) cell.

[0059] In some of the above methods, the mouse cell, or the human cell is a primary cell or an immortalized cell. In some of the above methods, the mouse cell, or the human cell is a pluripotent cell. In some of the above methods, the mouse pluripotent cell is a mouse embryonic stem (ES) cell. In some of the above methods, the human pluripotent cell is a human embryonic stem (ES) cell, a human adult stem cell, a developmentally restricted human progenitor cell, or a human induced pluripotent stem (iPS) cell. In some of the above methods, the human iPS cells is being maintained in a medium comprising a base medium and supplements, wherein the medium comprises: (a) a leukemia inhibitory factor (LIF) polypeptide; (b) a glycogen synthase kinase (GSK3) inhibitor; and (c) a MEK inhibitor; wherein the medium has an osmolality of about 175 mOsm/kg to about 280 mOsm/kg.

[0060] In some of the above methods, the Cas protein is Cas9. In some of the above methods, the target sequence is immediately flanked by a Protospacer Adjacent Motif (PAM) sequence.

[0061] In some of the above methods, the targeted genetic modification comprises simultaneous deletion of an endogenous nucleic acid sequence at the genomic locus of interest and insertion of the first nucleic acid at the genomic locus of interest in a single step. In some of the above methods, the deleted endogenous nucleic acid sequence is from about 30 kb to about 110 kb, and the inserted first nucleic acid is from about 40 kb to about 140 kb.

[0062] In some of the above methods, the targeted genetic modification is a biallelic genetic modification. In some of the above methods, the biallelic genetic modification comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid at the genomic locus of interest in two homologous chromosomes. In some of the above methods, the modified eukaryotic cell, the modified mouse cell, or the modified human cell is compound heterozygous at the genomic locus of interest. In some of the above methods, the modified eukaryotic cell, the modified mouse cell, or the modified human cell is hemizygous at the genomic locus of interest. In some of the above methods, the targeted genetic modification at the genomic locus of interest in one

chromosome comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid. In some of the above methods, the targeted genetic modification comprises: (1) deletion of an endogenous nucleic acid sequence at the genomic locus of interest in first and second homologous chromosomes; and (2) insertion of the first nucleic acid into the genomic locus of interest in the first homologous chromosome and disruption of the genomic locus of interest in the second homologous chromosome.

[0063] In some of the above methods, the targeted genetic modification comprises: (a) a replacement of an endogenous nucleic acid sequence with a homologous or an orthologous nucleic acid sequence; (b) a deletion of an endogenous nucleic acid sequence; (c) a deletion of an endogenous nucleic acid sequence, wherein the deletion ranges from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb; (d) insertion of an exogenous nucleic acid sequence; (e) insertion of an exogenous nucleic acid sequence ranging from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb; (f) insertion of an exogenous nucleic acid sequence comprising a homologous or an orthologous nucleic acid sequence; (g) insertion of a chimeric nucleic acid sequence comprising a human and a non-human nucleic acid sequence; (h) insertion of a conditional allele flanked with site-specific recombinase target sequences; (i) insertion of a selectable marker or a reporter gene operably linked to a promoter active in the pluripotent cell; or (j) a combination thereof.

[0064] In some of the above methods, the genomic locus of interest comprises the Interleukin-2 receptor gamma locus, the *ApoE* locus, the *Rag1* locus, the *Rag2* locus, both of the *Rag1* and the *Rag2* loci, the *Adamts5* locus, the *Trpa1* locus, the *Folh1* locus,

the *Erbb4* locus, the *Lrp5* locus, the *C5* (*Hc*) locus, the *Ror1* locus, or the *Dpp4* locus. In some of the above methods, the genomic locus of interest comprises extrachromosomal DNA.

[0065] Also provided are methods for producing an F0 generation non-human animal or mouse that comprises a targeted genetic modification at a genomic locus of interest, comprising: (a) modifying a non-human or mouse ES cell using any of the above methods; (b) identifying the modified non-human or mouse ES cell comprising the targeted genetic modification at the genomic locus of interest; (c) introducing the modified non-human or mouse ES cell into a non-human or mouse host embryo; and (d) gestating the non-human or mouse host embryo in a surrogate mother, wherein the surrogate mother produces the F0 generation non-human animal or mouse comprising the targeted genetic modification at the genomic locus of interest.

BRIEF DESCRIPTION OF THE DRAWINGS

[0066] Figure 1 depicts rat ESCs, which grow as compact spherical colonies that routinely detach and float in the dish.

[0067] Figure 2A through D depict various pluripotency markers expressed by rat ESCs: A depicts Oct-4 (green); B depicts Sox-2 (red); C depicts DAPI (blue); D depicts an overlay of pluripotency markers expressed by rESCs.

[0068] Figure 3 depicts that the rat ESCs express light levels of alkaline phosphatase (a pluripotency marker).

[0069] Figure 4 depicts the karyotype for line DA.2B, which is 42X,Y. Karyotyping was done because rat ESCs often become tetraploid; lines were thus pre-screened by counting metaphase chromosome spreads, and lines with mostly normal counts were then formally karyotyped.

[0070] Figure 5A-B provides photographs showing the analysis of the chromosome number of the ACI.G1 rat ES cell line.

[0071] Figure 6A-B provides photographs showing the analysis of the chromosome number of the DA.2B rat ES cell line.

[0072] Figure 7A-B provides photographs showing the analysis of the chromosome number of the DA.2C rat ES cell line.

[0073] Figure 8 depicts a closer view of a rat ESC of Figure 1.

[0074] Figure 9 depicts production of chimeras by blastocyst injection and transmission of the rat ESC genome through the germline. Chimeras were produced by blastocyst injection using parental ACI.G1 rat ESCs. High percentage chimeras usually have albino snouts.

[0075] Figure 10 depicts F1 agouti pups with albino littermates, sired by ACI/SD chimera labeled with an asterisk (*) in Figure 9.

[0076] Figure 11 provides a schematic of the rat *ApoE* locus and denotes with grey bars the cutting site for zinc finger nucleases (ZFN1 and ZFN2). The genomic regions corresponding to the 5' and 3' homology arms (5 kb and 5.4 kb, respectively) are denoted by the dark grey boxes. Exon 1 of the *ApoE* gene is non-coding and is shown as an open box closest to the 5' homology arm. The three introns of the *ApoE* gene are denoted as lines. Exons 2 and 3 comprise coding regions and are shown as stippled grey boxes. Exon 4 contains both coding and non-coding sequences as denoted by the stippled grey shading and the open box.

[0077] Figure 12 depicts targeting of the rat *Rosa26* locus, which lies between the *Setd5* and *Thumpd3* genes as in mouse, with the same spacing. Panel A shows the structure of the mouse *Rosa26* locus. Mouse *Rosa26* transcripts consist of 2 or 3 exons. Panel B depicts the structure of the rat *Rosa26* locus; the rat locus contains a second exon 1 (Ex1b) in addition to the homologous exon to mouse exon1 (Ex1a); no third exon has been identified in rat. Panel C depicts a targeted rat *Rosa26* allele; homology arms of 5 kb each were cloned by PCR using genomic DNA from DA rESC; the targeted allele contains a Splicing Acceptor (SA)-lacZ-hUB-neo cassette replacing a 117 bp deletion in the rat *Rosa26* intron.

[0078] Figure 13A depicts a control brain of a 14-week-old wild type rat, which was stained with X-gal. The control brain showed a low level of background staining for LacZ (dorsal view).

[0079] Figure 13B depicts LacZ expression in the brain of an rRosa26 heterozygous rat (14-week old). The lacZ reporter was expressed ubiquitously throughout the brain of the rRosa26 heterozygote.

[0080] Figure 13C depicts a control heart and thymus (inset) of a 14-week-old wild type rat, which were treated with X-gal. The control heart and thymus showed a low level of background staining for LacZ.

[0081] Figure 13D depicts LacZ expression in the heart and thymus (inset) of a 14-week-old rRosa26 heterozygous rat. The lacZ reporter was expressed ubiquitously throughout the heart and thymus of the rROSA26 heterozygote.

[0082] Figure 13E depicts a control lung of a 14-week-old wild type rat, which was treated with X-gal. The control lung showed a low level of background staining for LacZ.

[0083] Figure 13F depicts LacZ expression in the lung of a 14-week-old rRosa26 heterozygote rat. The lacZ reporter was expressed ubiquitously throughout the lung of the rRosa26 heterozygote.

[0084] Figure 13G and H depict LacZ expression in E12.5 rat embryos. In contrast to the wild-type control embryo (H), which shows a low level of background LacZ staining, the rRosa26 heterozygous embryo exhibited ubiquitous expression of the LacZ reporter throughout the embryo.

[0085] Figure 13I and J depict LacZ expression in E14.5 rat embryos. In contrast to the wild-type control embryo (J), which shows a low level of background LacZ staining, the rRosa26 heterozygous rat embryo exhibited ubiquitous expression of the LacZ reporter throughout the embryo.

[0086] Figure 14 illustrates a homologous or non-homologous recombination event that occurs inside a rat ES cell following an electroporation of a targeting vector comprising a selection cassette (lacZ-neo cassette).

[0087] Figure 15 illustrates the mechanism by which genome-editing endonucleases (e.g., ZFNs and TALENs) introduce a double strand break (DSB) in a target genomic sequence and activate non-homologous end-joining (NHEJ) in an ES cell.

[0088] Figure 16 illustrates a gene targeting technique that utilizes ZFN/TALENs to improve the efficiency of homologous recombination of a targeting vector. DSB represents double strand break.

[0089] Figure 17 shows ApoE-ZFN-AB5 chimeras produced by chimera production and germline transmission of the modified rat *ApoE* locus. The targeted modification was assisted by zinc finger nucleases.

[0090] Figure 18 provides a schematic of the IL2r-γ targeting event in combination with zinc finger nucleases that target ZFN U and ZFN D. The region of the rat IL2r-γ locus targeted by ZFN U and ZFN D is shown (SEQ ID NO: 93). ZFN cut sites are noted in the figure.

[0091] Figure 19 provides a schematic of the IL2r-γ targeting event in combination with zinc finger nucleases that target ZFN U and ZFN D or in combination with gRNAs (gRNA1, gRNA2, gRNA3, gRNA4). The regions of the rat IL2r-γ locus targeted by ZFN U and ZFN D or gRNAs1-4 are shown, and ZFN cut sites are noted.

[0092] Figure 20 provides a schematic of the rat *ApoE* locus and a targeting plasmid. The upper schematic shows the genomic structure of the rat *ApoE* locus and the genomic regions corresponding to the 5' and 3' homology arms (5 kb and 5.4 kb respectively; dark grey boxes). Exon 1 of the *ApoE* gene is non-coding and is shown as an open box closest to the 5' homology arm. The three introns of the *ApoE* gene are denoted as lines. Exons 2 and 3 comprise coding regions and are shown as stippled grey boxes. Exon 4 contains both coding and non-coding sequences as denoted by the stippled grey shading and the open box. The lower panel shows the targeting plasmid. The 5' and 3' homology arms (5 kb and 5.4 kb, respectively) are denoted by the dark grey boxes. The targeting vector comprises a reporter gene (lacZ) and a self-deleting cassette flanked by loxP sites (open arrows). The self-deleting cassette comprises a mouse *Prm1* promoter operably linked to the Crei gene and a drug selection cassette comprising a human ubiquitin promoter operably linked to a neomycin resistance gene.

[0093] Figure 21A provides a schematic for targeting the *ApoE* locus in rat ES cells using zinc-finger nucleases and a targeting vector comprising a reporter gene (LacZ) and a self-deleting cassette comprising a mouse *Prm1* promoter operably linked to the Crei gene and a drug selection cassette comprising a human ubiquitin promoter operably linked to a neomycin resistance gene. Figure 21B depicts a homozygous targeted *ApoE* locus.

[0094] Figure 22 provides a schematic of the rat *ApoE* locus and a large targeting vector (LTVEC). The upper panel shows the genomic organization of the rat *ApoE* locus and the genomic regions corresponding to the 5' and 3' homology arms (45 kb and 23 kb, respectively; the dark grey boxes). Exon 1 of *ApoE* is non-coding and is shown as an open box closest to the 5' homology arm. The three introns of the *ApoE* gene are denoted as lines and exons 2 and 3 comprise coding regions and are shown as stippled grey boxes. Exon 4 contains both coding and non-coding sequences as denoted by the stippled grey shading and the open box. The lower panel shows the LTVEC for modifying the rat *ApoE* locus. The 5' and 3' homology arms (45 kb and 23 kb, respectively) are denoted by the dark grey boxes. The LTVEC comprises a reporter gene (lacZ) and a self-deleting cassette flanked by loxP sites (open arrows), which comprises a mouse *Prm1* promoter operably linked to the Crei gene and a drug selection cassette comprising a human ubiquitin promoter operably linked to a neomycin resistance gene.

[0095] Figure 23 provides a schematic of the rat *ApoE* locus and denotes with grey bars the cutting sites for zinc finger nucleases (ZFN1 and ZFN2) used together with the large targeting vector (LTVEC) to enhance homologous recombination between the targeting vector and the target cognate chromosomal region.

[0096] Figure 24 depicts the rat IL2r-γ locus that has been disrupted by a 3.2 kb deletion and the insertion of a reporter gene (eGFP) and a self-deleting cassette comprising a drug selection cassette (hUb-neo) and the Crei gene operably linked to a mouse *Prm1* promoter.

[0097] Figure 25 provides another depiction of the rat IL2r-γ locus that has been disrupted by a 3.2 kb deletion and the insertion of a reporter gene (eGFP) and a self-deleting cassette comprising the Crei gene operably linked to a mouse *Prm1* promoter and a drug selection cassette (hUb-Neo).

[0098] Figure 26 provides a schematic of the rat *Rag2* locus and a large targeting vector (LTVEC) for modifying the rat *Rag2* locus. The upper panel shows the genomic organization of the rat *Rag2* locus and the cognate genomic regions corresponding to the 5' and 3' homology arms (48 kb and 84 kb, respectively; dark grey boxes). *Rag2* comprises single exon denoted by the stippled grey shading. The lower panel is the LTVEC. The 5' and 3' homology arms (48 kb and 84 kb, respectively) are denoted by

the dark grey boxes. The LTVEC comprises a reporter gene (lacZ) and a self-deleting cassette flanked by loxP sites (open arrows) that contains a rat *Prm1* promoter operably linked to the Crei gene and a drug selection cassette containing a human ubiquitin promoter operably linked to a neomycin resistance gene.

[0099] Figure 27 provides the genomic structure of the rat *Rag1*/*Rag2* locus and the genomic regions deleted by either *Rag2* targeting (*Rag2* deletion) or *Rag2*/*Rag1* double targeting (*Rag2*/*Rag1* deletion).

[00100] Figure 28 provides a schematic of the rat *Rag2* and *Rag1* loci and a large targeting vector (LTVEC) used for modifying the loci. The upper panel shows the genomic organization of the *Rag1* and *Rag2* loci and the cognate genomic regions corresponding to the 5' and 3' homology arms (48 kb and 15 kb, respectively; dark grey boxes). *Rag2* and *Rag1* each comprise a single exon denoted by the stippled grey shading. The lower panel is the LTVEC. The 5' and 3' homology arms (48 kb and 15 kb, respectively) are denoted by the dark grey boxes. The LTVEC comprises a reporter gene (lacZ) and a self-deleting cassette flanked by loxP sites (open arrows), which comprises a rat *Prm1* promoter operably linked to the Crei gene and a drug selection cassette comprising a human ubiquitin promoter operably linked to a neomycin resistance gene.

[00101] Figure 29 shows flow cytometry analysis for GFP expression and for T-cell marker CD3 (panels A and D), B-cell marker B220 (panels B and E), and NK cell marker CD161a (panels C and F) in peripheral blood mononuclear cells (PBMCs) from an II2rg-/y chimeric rat (panels A-C) and a WT DA rat (panels D-F). Double-positive cells are showin in quadrant R8. Figure 29 shows that II2rg-/y PBMC do not express mature lymphocyte markers.

[00102] Figure 30 shows that GFP-positive lymphocytes were detected in peripheral blood in 2 of the 3 II2rg-/y chimeras.

[00103] Figure 31 provides a schematic of the rat *Il2rg* locus and a targeting plasmid for the full humanization of the rat *Il2rg* locus. The upper panel shows the genomic organization of the rat *Il2rg* locus and the cognate genomic regions corresponding to the 5' and 3' homology arms (4.3 kb and 4.0 kb, respectively; grey boxes). The lower panel is the targeting plasmid. The 5' and 3' homology arms (4.3 kb

and 4.0 kb, respectively) are denoted by the grey boxes. The targeting plasmid comprises the human IL-2rg genomic region and a deletion cassette flanked by loxP sites (open arrows) that contains a drug selection cassette containing a human ubiquitin promoter operably linked to a neomycin resistance gene.

[00104] Figure 32 provides a schematic of the rat *Il2rg* locus and a targeting plasmid for the ecto-domain humanization of the rat *Il2rg* locus. The upper panel shows the genomic organization of the rat *Il2rg* locus and the cognate genomic regions corresponding to the 5' and 3' homology arms (4.3 kb and 4.0 kb, respectively; grey boxes). The lower panel is the targeting plasmid. The 5' and 3' homology arms (4.3 kb and 4.0 kb, respectively) are denoted by the grey boxes. The targeting plasmid comprises the human ecto-domain of the IL-2Rg genomic region and a self-deleting cassette flanked by loxP sites (open arrows) that contains a rat *Prm1* promoter operably linked to the Crei gene and a drug selection cassette containing a human ubiquitin promoter operably linked to a neomycin resistance gene.

[00105] Figure 33 provides a sequence alignment of the human IL-2rg protein (SEQ ID NO: 20; NP_000197.1); the rat IL-2rg protein (SEQ ID NO: 21; NP_543165.1); and the chimeric IL-2rg protein (SEQ ID NO: 22) comprising the human ecto-domain of IL-2rg fused to the remainder of the rat IL-2rg protein. The junction between the human and rat IL-2rg is noted by the vertical line.

[00106] Figure 34 provides a schematic of CRISPR/Cas9-assisted humanization of the mouse *Lrp5* gene; the LTVEC is shown the top panel and the mouse *Lrp5* locus is shown in the bottom panel. The region humanized is the ectodomain. The arrows indicate target sites for each gRNA (gA, gB, gB2, gC, gD, gE2, gE, gF) and ZFN (a-d).

[00107] Figure 35 depicts the percent targeting efficiency of LTVECs targeting genes of increasing size for deletion (Figure 35A) and LTVECs with human gene insertions of increasing size (Figure 35B). The LTVECs were used alone (gray squares or triangles) or in combination with ZFNs (black squares or triangles).

[00108] Figure 36 provides a schematic of CRISPR/Cas9-assisted humanization of the entire coding region of the mouse *Trpa1* gene; the LTVEC is shown the top panel and the mouse *Trpa1* locus is shown in the bottom panel. The arrows indicate target sites for each gRNA (gA, gA2, gB, gC, gD, gE2, gE, gF).

[00109] Figure 37 provides a schematic of CRISPR/Cas9-assisted humanization of the ectodomain (exon 2 to stop codon) of the mouse *Folh1* gene; the LTVEC is shown the top panel and the mouse *Folh1* locus is shown in the bottom panel. The arrows indicate target sites for each gRNA (gA, gA2, gB, gC, gD, gE, gE2, gF).

[00110] Figure 38 provides a schematic of CRISPR/Cas9-assisted humanization of the region from exon 2 to the stop codon of the mouse *C5* (*Hc*) gene; the LTVEC is shown the top panel and the mouse *C5* (*Hc*) locus is shown in the bottom panel. The arrows indicate target sites for each gRNA (gA, gB, gB2, gC, gD, gE2, gE, gF).

[00111] Figure 39 provides a schematic of CRISPR/Cas9-assisted humanization of the entire coding region of the mouse *Adamts5* gene; the LTVEC is shown the top panel and the mouse *Adamts5* locus is shown in the bottom panel. The arrows indicate target sites for each gRNA (gA, gA2, gB, gC, gD, gE2, gE, gF).

[00112] Figure 40 provides a schematic of CRISPR/Cas9-assisted humanization of exons 4-15 of the mouse *Erbb4* gene; the LTVEC is shown the top panel and the mouse *Erbb4* locus is shown in the bottom panel. The arrows indicate target sites for each gRNA (gA, gB, gB2, gC, gD, gE2, gE, gF).

[00113] Figure 41 provides a schematic of CRISPR/Cas9-assisted humanization of exons 2-7 of the mouse *Ror1* gene; the LTVEC is shown the top panel and the mouse *Ror1* locus is shown in the bottom panel. The arrows indicate target sites for each gRNA (gA, gB, gC, gD, gE, gF).

[00114] Figure 42 provides a schematic of CRISPR/Cas9-assisted humanization of the region from exon 2 to the stop codon of the mouse *Dpp4* gene; the LTVEC is shown the top panel and the mouse *Dpp4* locus is shown in the bottom panel. The arrows indicate target sites for each gRNA (gA, gB, gB2, gC, gD, gE2, gE, gF).

[00115] Figure 43 shows 12-week-old female rat brains stained with X-gal. Figure 43A-C show a brain from a wild type rat, and Figure 43D-F show a brain from an *ApoE* $^{+/-}$ rat. Figure 43A and D show dorsal views, Figure 43B and E show ventral views, and Figure 43C and F show close-up views.

[00116] Figure 44 shows 12-week-old female rat hearts (A and C) and corresponding close-ups of blood vessels (B and D) stained with X-gal. Figure 44A and B show a heart and blood vessels, respectively, from a wild type rat, and Figure 44C and

D show a heart and blood vessels, respectively, from an *ApoE* $^{+/-}$ rat. Staining was present in the atria of the heart and in some vessels (e.g., vena cava).

[00117] Figure 45 shows 12-week-old female rat livers stained with X-gal. Figure 45A and B show a liver from a wild type rat, and Figure 45C and D show a liver from an *ApoE* $^{+/-}$ rat. Figure 45B and D are close-ups of the livers.

[00118] Figure 46 shows detection of cholesterol, LDL, HDL, and triglyceride levels (Figure 46A-D, respectively) in homozygous *ApoE*-targeted rats, heterozygous *ApoE*-targeted rats, and wild type rats at 6 weeks, 9 weeks, 12 weeks, and 15 weeks.

[00119] Figure 47 shows a schematic of the rat *ApoE* locus (upper panel) and a large targeting vector (LTVEC) that targets the rat *ApoE* locus (lower panel). The upper panel shows the genomic organization of the rat *ApoE* locus and the genomic regions corresponding to the 5' and 3' homology arms (45 kb and 23 kb, respectively; the dark grey boxes). Exon 1 of *ApoE* is non-coding and is shown as an open box closest to the 5' homology arm. The three introns of the *ApoE* gene are denoted as lines and exons 2 and 3 comprise coding regions and are shown as stippled grey boxes. Exon 4 contains both coding and non-coding sequences as denoted by the stippled grey shading and the open box. Target sites for *ApoE* gRNA2 (SEQ ID NO: 87) and gRNA3 (SEQ ID NO: 88) are indicated. The lower panel shows the LTVEC for modifying the rat *ApoE* locus. The 5' and 3' homology arms (45 kb and 23 kb, respectively) are denoted by the dark grey boxes. The LTVEC comprises a reporter gene (lacZ) and a self-deleting cassette flanked by loxP sites (open arrows), which comprises a mouse *Prm1* promoter operably linked to the Crei gene and a drug selection cassette comprising a human ubiquitin promoter operably linked to a neomycin resistance gene.

[00120] Figure 48 shows a schematic of the rat *Rag2* locus (upper panel) and a large targeting vector (LTVEC) that targets the rat *Rag2* locus (lower panel). The upper panel shows the genomic organization of the rat *Rag2* locus and the cognate genomic regions corresponding to the 5' and 3' homology arms (48 kb and 84 kb, respectively; dark grey boxes). *Rag2* comprises a single exon denoted by the stippled grey shading. Target sites for *Rag2* gRNA1 (SEQ ID NO: 89) and gRNA4 (SEQ ID NO: 90) are indicated. The lower panel is the LTVEC. The 5' and 3' homology arms (48 kb and 84 kb, respectively) are denoted by the dark grey boxes. The LTVEC comprises a reporter

gene (lacZ) and a self-deleting cassette flanked by loxP sites (open arrows) that contains a rat *Prm1* promoter operably linked to the Crei gene and a drug selection cassette containing a human ubiquitin promoter operably linked to a hygromycin resistance gene.

[00121] Figure 49 shows a schematic of the rat *Il2rg* locus (upper panel) and a targeting plasmid for ectodomain humanization of the rat *Il2rg* locus (lower panel). The upper panel shows the genomic organization of the rat *Il2rg* locus and the cognate genomic regions corresponding to the 5' and 3' homology arms (4.3 kb and 4.0 kb, respectively; grey boxes). Target sites for *Il2rg* gRNA2 (SEQ ID NO: 91) and gRNA4 (SEQ ID NO: 92) are indicated. The lower panel is the targeting plasmid. The 5' and 3' homology arms (4.3 kb and 4.0 kb, respectively) are denoted by the grey boxes. The targeting plasmid comprises the human ecto-domain of the IL-2Rg genomic region and a self-deleting cassette flanked by loxP sites (open arrows) that contains a rat *Prm1* promoter operably linked to the Crei gene and a drug selection cassette containing a human ubiquitin promoter operably linked to a neomycin resistance gene.

[00122] Figure 50 shows a schematic of the rat *Rag2* and *Rag1* loci and a large targeting vector (LTVEC) used for modifying the loci in *Il2rg*-targeted rat ES cells (clone *Il2rg*-CG12). The upper panel shows the genomic organization of the *Rag1* and *Rag2* loci and the cognate genomic regions corresponding to the 5' and 3' homology arms (48 kb and 15 kb, respectively; grey boxes). *Rag2* and *Rag1* each comprise a single exon denoted by the unshaded arrows. The lower panel is the LTVEC. The 5' and 3' homology arms (48 kb and 15 kb, respectively) are denoted by the grey boxes. The LTVEC comprises a reporter gene (eGFP) and a puromycin resistance gene separated by an internal ribosome entry site (IRES) and operably linked to an actin promoter. The LTVEC further comprises a self-deleting cassette flanked by loxP sites (open arrows), which comprises a rat *Prm1* promoter operably linked to the Crei gene and a drug selection cassette comprising a human ubiquitin promoter operably linked to a neomycin resistance gene.

[00123] Figure 51 depicts a schematic for replacement of a portion of the human *ADAM6* locus with a nucleic acid comprising the mouse *Adam6a* and mouse *Adam6b* loci using an LTVEC and a guide RNA in human iPS cells. The target site for the guide RNA is indicated by the arrow.

[00124] Figure 52A depicts the morphology displayed by human iPS cells cultured for 8 days in 2i medium. Figure 52B depicts the morphology displayed by human iPS cells cultured for 12 days in 2i medium.

[00125] Figures 53A-53D depict the morphology of human iPS cells cultured in mTeSR™ -hLIF medium or low osmolality VG2i medium for 6 days. Figures 53A and 53B depict the morphology of human iPS cells cultured in mTeSR™ -hLIF medium (Figure 3A) or VG2i medium (Figure 53B) for 6 days. Figures 53C and 53D depict the morphology of human iPS cells cultured on newborn human foreskin fibroblast (NuFF) feeder cells in mTeSR™ -hLIF medium (Figure 53C) or VG2i medium (Figure 53D) for 6 days.

[00126] Figure 54A depicts reprogrammed human iPS cells cultured in VG2i medium that have been stained for alkaline phosphatase. Figures 54B and 54C depict reprogrammed human iPS cells cultured in VG2i medium that have been immunostained for the expression of NANOG.

[00127] Figures 55A-55C illustrate enzymatic dissociation and subculture of reprogrammed human iPS cells cultured in VG2i medium. Figure 55A depicts reprogrammed human iPS cells cultured in VG2i medium prior to enzymatic dissociation with trypsin in the absence of a ROCK inhibitor. Figure 55B depicts human iPS cells cultured in VG2i medium for 1 day after subculture. Figure 55C depicts human iPS cells cultured in VG2i medium for 4 days after subculture.

DETAILED DESCRIPTION OF THE INVENTION

[00128] Compositions and methods are provided for modifying a rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse, or hamster genomic locus of interest via bacterial homologous recombination (BHR) in a prokaryotic cell. Compositions and methods are also provided for genetically modifying a genomic locus of interest, for example, rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, or mouse genomic locus of interest using a large targeting vector (LTVEC) in combination with endonucleases. Compositions and methods are also provided for producing a genetically modified non-human animal, for example, a rat, mouse, rodent, or non-rat rodent,

comprising one or more targeted genetic modifications. Also provided are isolated human and non-human totipotent or pluripotent stem cells, in particular rat embryonic stem cells, that are capable of sustaining pluripotency following one or more serial genetic modifications *in vitro,* and that are capable of transmitting the targeted genetic modifications to subsequent generations through germline.

Glossary

[00129] The term "embryonic stem cell" or "ES cell" as used herein includes an embryo-derived totipotent or pluripotent cell that 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 types. The term "non-pluripotent cell" includes cells that are not pluripotent cells.

[00130] 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 sequence having amino acid sequence that is 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.

[00131] 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.

[00132] 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 DNA larger than those typically used by other approaches intended to perform homologous gene targeting in eukaryotic cells. Examples of LTVEC, include, but are not limited to, bacterial homologous chromosome (BAC) and yeast artificial chromosome (YAC).

[00133] The term "modification of allele" (MOA) as used herein 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)" as described herein includes, but is 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.

[00134] 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.

[00135] "Serial" genetic modifications include two or more modifications conducted independently to a cell (*e.g.*, a eukaryotic cell, a non-rat eukaryotic cell, a mammalian cell, a human cell, a non-human mammalian cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, a human iPS cell, a human cell, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell, a fibroblast, or a Chinese hamster ovary (CHO) cell). The first modification may be achieved by electroporation, or any other method known in the art. Then a second modification is made to the same cell genome employing a suitable second nucleic acid construct. The second modification may be achieved by a second electroporation, or any other method known in the art. In various embodiments, following the first and the second genetic modifications of the same cell, a third, a fourth, a fifth, a sixth, and so on, serial genetic modifications (one following another) may be achieved using, *e.g.,* serial electroporation or any other suitable method (serially) known in the art.

[00136] 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.

[00137] The term "germline" in reference to a nucleic acid sequence includes a nucleic acid sequence that can be passed to progeny.

[00138] 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 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."

[00139] 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 framework (FR) regions, 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.

[00140] The phrase "operably linked" comprises 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.

1. Target Locus Comprising a Nucleic Acid

[00141] Various methods and compositions are provided, which allow for the integration of at least one insert nucleic acid at a target locus. As used herein, a "genomic locus of interest" comprises any segment or region of DNA within the genome that one desires to integrate an insert nucleic acid. The terms "genomic locus of interest" and "target genomic locus of interest" can be used interchangeable. The genomic locus of interest can be native to the cell, or alternatively can comprise a heterologous or exogenous segment of DNA that was integrated into the genome of the cell. Such heterologous or exogenous segments of DNA can include transgenes, expression cassettes, polynucleotide encoding selection makers, or heterologous or exogenous regions of genomic DNA. The term "locus" is a defined herein as a segment of DNA within the genomic DNA. Genetic modifications as described herein can include one or more deletions from a locus of interest, additions to a locus of interest, replacement of a locus of interest, and/or any combination thereof. The locus of interest can comprise coding regions or non-coding regulatory regions.

[00142] The genomic locus of interest can further comprise any component of a targeted integration system including, for example, a recognition site, a selection marker, a previously integrated insert nucleic acid, polynucleotides encoding nuclease agents, promoters, etc. Alternatively, the genomic locus of interest can be located within an extrachromosomal DNA within the cell, such as a yeast artificial chromosome (YAC), a bacterial artificial chromosome (BAC), a human artificial chromosome, or any other engineered genomic region contained in an appropriate host cell. In various embodiments, the targeted locus can comprise native, heterologous, or exogenous nucleic acid sequence from a prokaryote, a eukaryote, a non-rat eukaryote, yeast, bacteria, a non-human mammal, a non-human cell, a rodent, a non-rat rodent, a human, a rat, a mouse, a

hamster, a rabbit, a pig, a bovine, a deer, a sheep, a goat, a chicken, a cat, a dog, a ferret, a primate (*e.g.*, marmoset, rhesus monkey), domesticated mammal or an agricultural mammal or any other organism of interest or a combination thereof. In some embodiments, the genomic locus of interest comprises a nucleic acid sequence from a human, a mouse, or a combination thereof.

[00143] In specific embodiments, the target locus is from, for example, a eukaryotic cell, a non-rat eukaryotic cell, a mammalian cell, human cell, a non-human mammalian cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, a human iPS cell, a human cell, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell, a fibroblast, or a CHO cell.

[00144] In specific embodiments, the genomic locus of interest comprises a target locus of a "rat nucleic acid." Such a region comprises a nucleic acid from a rat that is integrated within the genome of a cell. Non-limiting examples of the target locus include a genomic locus that encodes a protein expressed in a B cell, a genomic locus that expresses a polypeptide in an immature B cell, a genomic locus that expresses a polypeptide in a mature B cell, an immunoglobulin (Ig) loci, or a T cell receptor loci, including, for example, a T cell receptor alpha locus. Additional examples of target genomic locus include 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, an IL2 receptor locus, including, for example, an IL2 Receptor gamma (*Il2rg*) locus, an *ApoE* locus, a *Rag1* locus, a *Rag2* locus, a *Rag1*/*Rag2* locus, and an *Erbb4* locus. Any such target locus can be from a rat or can be from a eukaryotic cell, a non-rat eukaryotic cell, a mammalian cell, a human cell, or a non-human mammalian cell.

[00145] In one embodiment, the target locus encodes a mammalian immunoglobulin heavy chain variable region amino acid sequence. In one embodiment, the target locus encodes a rat immunoglobulin heavy chain variable region amino acid sequence. In one embodiment, the target locus comprises a genomic DNA sequence comprising an unrearranged rat, mouse, or 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 rat, mouse, or human immunoglobulin heavy chain constant region nucleic acid sequence selected from a CH1, a hinge, a CH2, a CH3, and a combination thereof. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a CH1-hinge-CH2-CH3. In one embodiment, the target locus comprises a rearranged rat, mouse, or 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 rat, mouse, or human immunoglobulin heavy chain constant region nucleic acid sequence selected from a CH1, a hinge, a CH2, a CH3, and a combination thereof. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a CH1-hinge-CH2-CH3.

[00146] In one embodiment, the target locus comprises a genomic DNA sequence that encodes a mammalian immunoglobulin light chain variable region amino acid sequence. In one embodiment, the genomic DNA sequence comprises an unrearranged mammalian λ and/or κ light chain variable region nucleic acid sequence.

[00147] In one embodiment, the genomic DNA sequence comprises a rearranged mammalian λ and/or κ light chain variable region nucleic acid sequence. In one embodiment, the unrearranged λ or κ light chain variable region nucleic acid sequence is operably linked to a mammalian 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. In one embodiment, the mammalian immunoglobulin light chain constant region nucleic acid sequence is a rat immunoglobulin light chain constant region nucleic acid sequence. In one embodiment, the mammalian immunoglobulin light chain constant region nucleic acid sequence is a mouse immunoglobulin light chain constant region nucleic acid sequence. In one embodiment, the mammalian immunoglobulin light chain constant region nucleic acid sequence is a human immunoglobulin light chain constant region nucleic acid sequence.

[00148] As used herein, an *ApoE* locus, an interleukin-2 receptor gamma (*Il2rg*) locus, a *Rag2* locus, a *Rag1* locus and/or a *Rag2*/*Rag1* locus comprise the respective regions of the genome (i.e., a mammalian genome, a human genome or a non-human mammalian genome) in which each of these genes or gene combinations are located.

Modifying any one of the *ApoE* locus, interleukin-2 receptor gamma (*Il2rg*) locus, *Rag2* locus, *Rag1* locus and/or *Rag2*/*Rag1* locus (i.e., a mammalian, a human, or a non-human mammalian *ApoE* locus, the interleukin-2 receptor gamma locus, the *Rag2* locus, the *Rag1* locus and/or the combined *Rag2*/*Rag1* locus) can comprise any desired alteration to the given locus. Non-limiting examples of modification to the given locus (i.e., a mammalian, a human, or a non-human mammalian locus) are discussed in further detail herein.

[00149] For example, in specific embodiments, one or more of the *ApoE* locus, interleukin-2 receptor gamma (*Il2rg*) locus, *Rag2* locus, *Rag1* locus and/or *Rag2*/*Rag1* locus (i.e., a mammalian, a human, or a non-human mammalian *ApoE* locus, a mammalian, a human, or a non-human mammalian interleukin-2 receptor gamma locus, a mammalian, a human, or a non-human mammalian *Rag2* locus, and/or the *Rag2*/*Rag1* locus) is modified such that the activity and/or level of the encoded ApoE protein or the interleukin-2 receptor gamma protein or the Rag1 protein or the Rag2 protein or a combination of the Rag1 and Rag2 proteins are decreased. In other embodiments, the activity of the ApoE protein, the interleukin-2 receptor gamma protein, the Rag1 protein, or the Rag2 protein, or a combination of the Rag1 and Rag2 proteins is absent.

[00150] By "decreased" is intended any decrease in the level or activity of the gene/protein encoded at the locus of interest. For example, a decrease in activity can comprise either (1) a statistically significant decrease in the overall level or activity of a given protein (i.e., ApoE, interleukin-2 receptor gamma, Rag2, Rag2 or a combination of Rag1 and Rag2) including, for example, a decreased level or activity of 0.5%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 120% or greater when compared to an appropriate control. Methods to assay for a decrease in the concentration and/or the activity of anyone of ApoE, interleukin-2 receptor gamma, Rag1 and Rag2 are known in the art.

[00151] In other embodiments, one or more of the a mammalian, a human, or a non-human mammalian *ApoE* locus, the a mammalian, a human, or a non-human mammalian interleukin-2 receptor gamma locus, a mammalian, a human, or a non-human mammalian *Rag2* locus, a mammalian, a human, or a non-human mammalian *Rag1* locus and/or a mammalian, a human, or a non-human mammalian *Rag2*/*Rag1*

locus comprise a modification such that the activity and/or level of the encoded ApoE polypeptide, the interleukin-2 receptor gamma polypeptide, the Rag2 polypeptide, the Rag1 polypeptide, or both the Rag1 and Rag2 polypeptide is increased. By "increased" is intended any increase in the level or activity of the gene/polypeptide encoded at the locus of interest. For example, an increase in activity can comprise either (1) a statistically significant increase in the overall level or activity of a given protein (i.e., ApoE, interleukin-2 receptor gamma, Rag1, Rag2 or Rag1 and Rag2) including, for example, an increased level or activity of 0.5%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 120% or greater when compared to an appropriate control. Methods to assay for an increase in the concentration and/or the activity of anyone of the ApoE, Rag1, Rag2 and interleukin-2 receptor gamma proteins are known in the art.

[00152] The genetic modification to the a mammalian, a human, or a non-human mammalian *ApoE* locus, a mammalian, a human, or a non-human mammalian interleukin-2 receptor gamma locus, a mammalian, a human, or a non-human mammalian *Rag2* locus, a mammalian, a human, or a non-human mammalian *Rag1* locus and/or a mammalian, a human, or a non-human mammalian *Rag2*/*Rag1* locus can comprise a deletion of an endogenous nucleic acid sequence at the genomic locus, an insertion of an exogenous nucleic acid at the genomic locus, or a combination thereof. The deletion and/or insertion can occur anywhere within the given locus as discussed elsewhere herein.

[00153] Further embodiments provided herein comprise the modification of one or more of the mammalian, human, or non-human mammalian *ApoE* locus, interleukin-2 receptor gamma locus, *Rag2* locus, *Rag1* locus and/or *Rag2*/*Rag1* locus through the replacement of a portion of the *ApoE* locus, interleukin-2 receptor gamma (*Il2rg*) locus, *Rag2* locus, *Rag1* locus and/or *Rag2*/*Rag1* locus with the corresponding homologous or orthologous portion of an *ApoE* locus, an interleukin-2 receptor gamma locus, a *Rag2* locus, a *Rag1* locus and/or a *Rag2*/*Rag1* locus from another organism.

[00154] In still other embodiments, the modification of one or more of the mammalian, human, or non-human mammalian *ApoE* locus, the interleukin-2 receptor gamma locus, *Rag2* locus, *Rag1* locus, and/or *Rag2*/*Rag1* locus is carried out through the replacement of a portion of the *ApoE* locus, interleukin-2 receptor gamma (*Il2rg*) locus,

Rag2 locus, *Rag1* locus and/or *Rag2/Rag1* locus with an insert polynucleotide sharing across its full length least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% to a portion of an *ApoE* locus, an interleukin-2 receptor gamma locus, a *Rag2* locus, a *Rag1* locus and/or a *Rag2/Rag1* locus it is replacing.

[00155] The given insert polynucleotide and/or the corresponding region of the locus being deleted can be a coding region, an intron, an exon, an untranslated region, a regulatory region, a promoter, or an enhancer or any combination thereof or any portion thereof. Moreover, the given insert polynucleotide and/or the region of the locus, for example, being deleted can be of any desired length, including for example, between 10-100 nucleotides in length, 100-500 nucleotides in length, 500-1 kb nucleotide in length, 1 Kb to 1.5 kb nucleotide in length, 1.5 kb to 2 kb nucleotides in length, 2 kb to 2.5 kb nucleotides in length, 2.5 kb to 3 kb nucleotides in length, 3 kb to 5 kb nucleotides in length, 5 kb to 8 kb nucleotides in length, 8 kb to 10 kb nucleotides in length or more. In other instances, the size of the insertion or replacement is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, from about 350 kb to about 400 kb, from about 400 kb to about 800 kb, from about 800 kb to 1 Mb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, from about 2.5 Mb to about 2.8 Mb, from about 2.8 Mb to about 3 Mb. In other embodiments, the given insert polynucleotide and/or the region of the locus being deleted is at least 100, 200, 300, 400, 500, 600, 700, 800, or 900 nucleotides or at least 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb, 12 kb, 13 kb, 14 kb, 15 kb, 16 kb or greater. In other embodiments, the given insert polynucleotide and/or the region of the locus being deleted is at least 10 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb, at least 200 kb, at least 250 kb, or at least 300 kb or greater.

[00156] The given insert polynucleotide can be from any organism, including, for example, a rodent, a non-rat rodent, a rat, a mouse, a hamster, a mammal, a non-human mammal, a eukaryote, a non-rat eukaryote, a human, an agricultural animal or a domestic animal.

[00157] As discussed in further detail herein, various methods are provided to generate targeted modifications of any locus of interest, including for example, targeted modifications in the *ApoE* locus, interleukin-2 receptor gamma (*Il2rg*) locus, *Rag2* locus, *Rag1* locus and/or *Rag2*/*Rag1* locus. Further provided are genetically modified non-human animals, genetically modified non-human mammals, genetically modified non-rat eukaryotes, genetically modified non-pluripotent cells, or genetically modified pluripotent cells (e.g., a pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, or a human iPS cell), which comprise a deletion, an insertion, a replacement and/or any combination thereof at the interleukin-2 receptor gamma locus, at the *ApoE* locus, at the *Rag2* locus, at the *Rag1* locus, and/or at the *Rag2*/*Rag1* locus. Such genetic modifications (including those that result in an absence, a decrease, an increase or a modulation in activity of the target locus) and are also capable of being transmitted through the germline. In specific embodiments, the genetic modifications result in a knockout of the desired target locus. Such non-human animals, for example, find use in in a variety of experimental systems as discussed elsewhere herein.

[00158] For example, *ApoE* (Apolipoprotein E) knockouts offer an animal model to study endothelial function, including, but not limited to, plaque formation, transcriptional changes (Whole Transcriptome Shotgun Sequencing (RNA-Seq), and *ex vivo* function. ApoE is an important transport molecule and can transport lipids, such as cholesterol, through the bloodstream. ApoE can also function in the nervous system, for example, to clear β-amyloid from the brain. Modifications in ApoE have been implicated in various conditions, including, for example, atherosclerosis, hyperlipidemia, and Alzheimer's disease. *ApoE* knockout animals display impaired clearing of lipoproteins from the blood and develop atherosclerosis. Thus, *ApoE* knockout animals provide a model to study conditions and/or processes such as, for example, endothelia function, plaque formation, transcriptional changes (RNA-Seq), hyperlipidemia, atherosclerosis

and Alzheimer's disease. Assays to measure ApoE activity are known in the art. For example, a decrease in ApoE activity can be measured by assaying for a decrease in the ApoE levels in a blood sample obtained from a subject by immunoassays, such as by ELISA or by Immunoblotting techniques. Moreover, the large size of rats facilitates all these assays and improves the quality of the data.

[00159] RAG1 (Recombination-Activating Gene 1) and RAG2 (Recombination-Activating Gene 2) are enzymes that are part of a multi-subunit complex having VDJ recombination activity and play an important role in the rearrangement and recombination of immunoglobulin and T-cell receptor genes in lymphocytes. RAG1 and RAG2 induce a double stranded DNA cleavage to facilitate recombination and join of segments of the T cell receptor and B cell receptor (i.e., immunoglobulin) genes. Knockout of RAG1 and/or RAG2 causes a loss of B cells and T cells in the animal resulting in severe immunodeficiency. RAG1 and/or RAG2 knockout animals find use, for example, in studies of xenografts (i.e., human cell xenografts in rats), cancer, vaccine development, autoimmune disease, infectious disease and graft versus host disease (GVHD). Various assays to measure RAG1 and/or RAG2 activity are known in the art and include, for example, measuring recombination efficiency or assaying for the presence or absence of B cells and/or T cells in a subject.

[00160] The IL-2 receptor (IL-2R) is expressed on the surface of certain immune cells and binds to the cytokine interleukin-2 (IL-2). The IL-2R is an integral membrane protein comprising at least three separate subunit chains, including, an alpha chain (IL-2Ra, CD25), a beta chain (IL-2Rb, CD122) and a gamma chain (IL2-Rg, CD132). The IL-2 receptor gamma (also referred to as IL2r-γ or IL2Rg) chain is a common gamma chain that is shared by various cytokine receptors, including, for example, the receptors for IL-2, IL-4, IL-7, IL-9, IL-15 and IL-21. IL-2Rg comprises an ectodomain on the extracellular surface of the cell, which contributes to the binding of the ligand, a transmembrane domain, and an intracellular domain, which can interact with various molecules to induce intracellular signal transduction pathways. The *Il2rg* gene is found on the X-chromosome in mammals and certain mutations in the gamma chain gene in humans can cause human X-linked severe combined immunodeficiency (XSCID) characterized by a profound T-cell defect. In addition, the gamma chain ecto-domain can

be shed off of the transmembrane receptor and released as a soluble gamma chain receptor. The soluble gamma chain receptor can be detected in the blood of a subject and can function to regulate cytokine signaling.

[00161] In some embodiments, the non-human IL-2Rg chain is replaced with the human IL2-Rg chain such that the genetically modified animal expresses a fully human IL-2Rg chain. In other instances, it may be useful to replace only the ectodomain of a non-human IL-2Rg chain with the ectodomain of the human IL-2Rg chain. In such cases, the resulting humanized IL-2Rg chain expressed in a non-human comprises a human ectodomain, with the remainder of the molecule being from the native organism.

[00162] The full-length humanization of IL-2Rg is useful because non-human mammals having this modified locus will produce human IL-2Rg. This will allow for the detection of human IL-2Rg in non-human mammals with antibodies specific to human IL-2Rg. The ecto-humanization (i.e., replacing the ecto-domain of IL-2Rg a non-human mammal with the human ecto-domain of IL-2Rg) will result in an IL-2Rg polypeptide that will bind the human ligands for IL2-Rg, but because the cytoplasmic domain is still from the non-human mammal, the ecto-humanized form of IL-2Rg will also interact with the non-human mammal signaling machinery.

2. Modifying a Target Locus

A. Targeting Vectors and Insert Nucleic Acids

i. Insert Nucleic Acid

[00163] As used herein, the "insert nucleic acid" comprises a segment of DNA that one desires to integrate at the target locus. In one embodiment, the insert nucleic acid comprises one or more polynucleotides of interest. In other embodiments, the insert nucleic acid can comprise one or more expression cassettes. A given expression cassette can comprise a polynucleotide of interest, a polynucleotide encoding a selection marker and/or a reporter gene along with the various regulatory components that influence expression. Non-limiting examples of polynucleotides of interest, selection markers, and reporter genes that can be included within the insert nucleic acid are discussed in detail elsewhere herein.

[00164] In specific embodiments, the insert nucleic acid can comprise a nucleic acid from rat, which can include a segment of genomic DNA, a cDNA, a regulatory region, or any portion or combination thereof. In other embodiments, the insert nucleic acid can comprise a nucleic acid from a eukaryote, a non-rat eukaryote, a mammal, a human, a non-human mammal, a rodent, a non-rat rodent, a human, a rat, a mouse, a hamster, a rabbit, a pig, a bovine, a deer, a sheep, a goat, a chicken, a cat, a dog, a ferret, a primate (*e.g.*, marmoset, rhesus monkey), a domesticated mammal, or an agricultural mammal or any other organism of interest. As outlined in further detail herein, the insert nucleic acid employed in the various methods and compositions can result in the "humanization" of the a target locus of interest.

[00165] In one embodiment, the insert nucleic acid comprises a knock-in allele of at least one exon of an endogenous gene. In one embodiment, the insert nucleic acid comprises a knock-in allele of the entire endogenous gene (i.e., "gene-swap knock-in").

[00166] In one embodiment, the insert nucleic acid comprises a regulatory element, including for example, a promoter, an enhancer, or a transcriptional repressor-binding element.

[00167] In further embodiments, 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 specific embodiments, 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.

[00168] The insert nucleic acid ranges from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to

about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb.

[00169] In one embodiment, the insert nucleic acid comprises a deletion of, for example, a eukaryotic cell, a non-rat eukaryotic cell, a mammalian cell, a human cell or a non-human mammalian cell genomic DNA sequence ranging from about 1 kb to about 200 kb, from about 2 kb to about 20 kb, or from about 0.5 kb to about 3 Mb. In one embodiment, the extent of the deletion of the genomic DNA sequence is greater than a total length of the 5' homology arm and the 3' homology arm. In one embodiment, the extent of the deletion of the genomic DNA sequence ranges from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 20 kb to about 30 kb, from about 30 kb to about 40 kb, from about 40 kb to about 50 kb, from about 50 kb to about 60 kb, from about 60 kb to about 70 kb, from about 70 kb to about 80 kb, from about 80 kb to about 90 kb, from about 90 kb to about 100 kb, from about 100 kb to about 110 kb, from about 110 kb to about 120 kb, from about 120 kb to about 130 kb, from about 130 kb to about 140 kb, from about 140 kb to about 150 kb, from about 150 kb to about 160 kb, from about 160 kb to about 170 kb, from about 170 kb to about 180 kb, from about 180 kb to about 190 kb, from about 190 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, from about 350 kb to about 400 kb, from about 400 kb to about 800 kb, from about 800 kb to 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb, to about 2.5 Mb, from about 2.5 Mb to about 2.8 Mb, from about 2.8 Mb to about 3 Mb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb.

[00170] In one embodiment, the insert nucleic acid comprises an insertion or a replacement of a eukaryotic, a non-rat eukaryotic, a mammalian, a human or a non-human mammalian nucleic acid sequence with a homologous or orthologous human

nucleic acid sequence. In one embodiment, the insert nucleic acid comprises an insertion or replacement of a DNA sequence with a homologous or orthologous human nucleic acid sequence at an endogenous locus that comprises the corresponding DNA sequence.

[00171] In one embodiment, the genetic modification is an addition of a nucleic acid sequence. In one embodiment, the added nucleotide sequence ranges from 5 kb to 200 kb.

[00172] 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.

In one embodiment, the insert nucleic acid comprises an insertion or a replacement of a eukaryotic, a non-rat eukaryotic, a mammalian, a human, or a non-human mammalian, nucleic acid sequence with a homologous or orthologous human nucleic acid sequence. In one embodiment, the insert nucleic acid comprises an insertion or replacement of a rat DNA sequence with a homologous or orthologous human nucleic acid sequence at an endogenous rat locus that comprises the corresponding rat DNA sequence.

[00173] In one embodiment, the genetic modification 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 element. In one embodiment, the genetic modification comprises a deletion of a promoter. In one embodiment, the genetic modification comprises an addition of a promoter or a regulatory element. In one embodiment, the genetic modification comprises a replacement of a promoter or a regulatory element.

[00174] In one embodiment, the nucleic acid sequence of the targeting vector can comprise a polynucleotide that when integrated into the genome will produce a genetic modification of a region of the mammalian, human, or a non-human mammalian *ApoE* locus, wherein the genetic modification at the *ApoE* locus results in a decrease in ApoE activity, increase in ApoE activity, or a modulation of ApoE activity. In one embodiment, an *ApoE* knockout ("null allele) is generated.

[00175] In one embodiment, the nucleic acid sequence of the targeting vector can comprise a polynucleotide that when integrated into the genome will produce a genetic modification of a region of the mammalian, human cell, or non-human mammalian interleukin-2 receptor locus, wherein the genetic modification at the interleukin-2 receptor locus results in a decrease in interleukin-2 receptor activity. In one embodiment, an interleukin-2 receptor knockout ("null allele") is generated.

[00176] In further embodiments, the insert nucleic acid results in the replacement of a portion of the mammalian, human cell, or non-human mammalian *ApoE* locus, the interleukin-2 receptor gamma locus and/or *Rag2* locus, and/or *Rag1* locus and/or *Rag2*/*Rag1* locus with the corresponding homologous or orthologous portion of an *ApoE* locus, an interleukin-2 receptor gamma locus, a *Rag2* locus, a *Rag1* locus and/or a *Rag2*/*Rag1* locus from another organism.

[00177] Still other embodiments, the insert nucleic acid comprises a polynucleotide sharing across its full length least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% to a portion of an *ApoE* locus, an interleukin-2 receptor gamma locus, a *Rag2* locus, a *Rag1* locus and/or a *Rag2*/*Rag1* locus it is replacing.

[00178] The given insert polynucleotide and the corresponding region of the mammalian, human cell, or non-human mammalian locus being replaced can be a coding region, an intron, an exon, an untranslated region, a regulatory region, a promoter, or an enhancer or any combination thereof. Moreover, the given insert polynucleotide and/or the region of the mammalian, human cell, or non-human mammalian locus being deleted can be of any desired length, including for example, between 10-100 nucleotides in length, 100-500 nucleotides in length, 500-1 kb nucleotide in length, 1 Kb to 1.5 kb nucleotide in length, 1.5 kb to 2 kb nucleotides in length, 2 kb to 2.5 kb nucleotides in length, 2.5 kb to 3 kb nucleotides in length, 3 kb to 5 kb nucleotides in length, 5 kb to 8 kb nucleotides in length, 8 kb to 10 kb nucleotides in length or more. In other instances, the size of the insertion or replacement is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, from about 350

kb to about 400 kb, from about 400 kb to about 800 kb, from about 800 kb to 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb, to about 2.5 Mb, from about 2.5 Mb to about 2.8 Mb, from about 2.8 Mb to about 3 Mb. In other embodiments, the given insert polynucleotide and/or the region of the mammalian, human cell, or non-human mammalian locus being deleted is at least 100, 200, 300, 400, 500, 600, 700, 800, or 900 nucleotides or at least 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb, 12 kb, 13 kb, 14 kb, 15 kb, 16 kb or greater.

[00179] In one embodiment, the promoter is constitutively active promoter.

[00180] In one embodiment, the promoter is an inducible promoter. In one embodiment, the inducible promoter is a chemically-regulated promoter. In one embodiment, the chemically-regulated promoter is an alcohol-regulated promoter. In one embodiment, the alcohol-regulated promoter is an alcohol dehydrogenase (alcA) gene promoter. In one embodiment, the chemically-regulated promoter is a tetracycline-regulated promoter. In one embodiment, the tetracycline-regulated promoter is a tetracycline-responsive promoter. In one embodiment, the tetracycline-regulated promoter is a tetracycline operator sequence (tetO). In one embodiment, the tetracycline-regulated promoter is a tet-On promoter. In one embodiment, the tetracycline-regulated promoter a tet-Off promoter. In one embodiment, the chemically-regulated promoter is a steroid regulated promoter. In one embodiment, the steroid regulated promoter is a promoter of a rat glucocorticoid receptor. In one embodiment, the steroid regulated promoter is a promoter of an estrogen receptor. In one embodiment, the steroid-regulated promoter is a promoter of an ecdysone receptor. In one embodiment, the chemically-regulated promoter is a metal-regulated promoter. In one embodiment, the metal-regulated promoter is a metalloprotein promoter. In one embodiment, the inducible promoter is a physically-regulated promoter. In one embodiment, the physically-regulated promoter is a temperature-regulated promoter. In one embodiment, the temperature-regulated promoter is a heat shock promoter. In one embodiment, the physically-regulated promoter is a light-regulated promoter. In one embodiment, the light-regulated promoter is a light-inducible promoter. In one embodiment, the light-regulated promoter is a light-repressible promoter.

[00181] In one embodiment, the promoter is a tissue-specific promoter. In one embodiment, the promoter is a neuron-specific promoter. In one embodiment, the promoter is a glia-specific promoter. In one embodiment, the promoter is a muscle cell-specific promoter. In one embodiment, the promoter is a heart cell-specific promoter. In one embodiment, the promoter is a kidney cell-specific promoter. In one embodiment, the promoter is a bone cell-specific promoter. In one embodiment, the promoter is an endothelial cell-specific promoter. In one embodiment, the promoter is an immune cell-specific promoter. In one embodiment, the immune cell promoter is a B cell promoter. In one embodiment, the immune cell promoter is a T cell promoter.

[00182] In one embodiment, the promoter is a developmentally-regulated promoter. In one embodiment, the developmentally-regulated promoter is active only during an embryonic stage of development. In one embodiment, the developmentally-regulated promoter is active only in an adult cell.

[00183] In specific embodiments, the promoter may be selected based on the cell type. Thus the various promoters find use in a eukaryotic cell, a non-rat eukaryotic cell, a mammalian cell, a non-human mammalian cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, a human iPS cell, a human cell, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell, a fibroblast or a CHO cell.

[00184] In some embodiments, the insert nucleic acid comprises a nucleic acid flanked with site-specific recombination target sequences. It is recognized the while the entire insert nucleic acid can be flanked by such site-specific recombination target sequences, any region or individual polynucleotide of interest within the insert nucleic acid can also be flanked by such sites. The site-specific recombinase can be introduced into the cell by any means, including by introducing the recombinase polypeptide into the cell or by introducing a polynucleotide encoding the site-specific recombinase into the host cell. The polynucleotide encoding the site-specific recombinase can be located within the insert nucleic acid or within a separate polynucleotide. The site-specific recombinase can be operably linked to a promoter active in the cell including, for example, an inducible promoter, a promoter that is endogenous to the cell, a promoter

that is heterologous to the cell, a cell-specific promoter, a tissue-specific promoter, or a developmental stage-specific promoter. Site-specific recombination target sequences, which can flank the insert nucleic acid or any polynucleotide of interest in the insert nucleic acid can include, but are not limited to, loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox, and a combination thereof.

[00185] In some embodiments, the site-specific recombination sites flank a polynucleotide encoding a selection marker and/or a reporter gene contained within the insert nucleic acid. In such instances following integration of the insert nucleic acid at the targeted locus the sequences between the site-specific recombination sites can be removed.

[00186] In one embodiment, the insert nucleic acid comprises a polynucleotide encoding a selection marker. The selection marker can be contained in a selection cassette. Such selection markers include, but are not limited, to neomycin phosphotransferase (neo^r), hygromycin B phosphotransferase (hyg^r), puromycin-N-acetyltransferase ($puro^r$), blasticidin S deaminase (bsr^r), xanthine/guanine phosphoribosyl transferase (gpt), or herpes simplex virus thymidine kinase (HSV-k), or a combination thereof. In one embodiment, the polynucleotide encoding the selection marker is operably linked to a promoter active in the cell, rat cell, pluripotent rat cell, the ES rat cell, a eukaryotic cell, a non-rat eukaryotic cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, a human iPS cell, a mammalian cell, a non-human mammalian cell, a human cell, a rodent cell, a non-rat rodent cell, a mouse cell, a hamster cell, a fibroblast, or a CHO cell. When serially tiling polynucleotides of interest into a targeted locus, the selection marker can comprise a recognition site for a nuclease agent, as outlined above. In one embodiment, the polynucleotide encoding the selection marker is flanked with a site-specific recombination target sequences.

[00187] The insert nucleic acid can further comprise a reporter gene operably linked to a promoter, wherein the reporter gene encodes a reporter protein selected from the group consisting of or comprising 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/or a combination thereof. Such reporter genes can be operably linked to a promoter active in the cell. Such promoters can be an inducible promoter, a promoter that is endogenous to the reporter gene or the cell, a promoter that is heterologous to the reporter gene or to the cell, a cell-specific promoter, a tissue-specific promoter, or a developmental stage-specific promoter.

[00188] In one embodiment, nucleic acid insert can comprise a mammalian 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 mammalian 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 nucleic acid comprises a genomic locus that encodes a protein expressed in a spleen cell.

[00189] In one embodiment, the mammalian 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 mammalian 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 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 DNA sequence, a rat genomic DNA sequence, eukaryotic genomic DNA sequence, a non-rat eukaryotic genomic DNA sequence, a mammalian genomic DNA sequence, a human genomic DNA sequence, or non-human DNA sequence mammalian, or a combination thereof. In one embodiment, the genomic locus comprises, in any order, rat and human genomic DNA sequences. In one embodiment, the genomic locus comprises, in any order, mouse and human genomic DNA sequences. In one embodiment, the genomic locus comprises, in any order, mouse

and rat genomic DNA sequences. In one embodiment, the genomic locus comprises, in any order, rat, mouse, and human genomic DNA sequences.

[00190] In one embodiment, the genomic locus comprises a mouse genomic DNA sequence, a rat genomic DNA sequence, a hamster genomic DNA sequence, a human genomic DNA sequence, eukaryotic genomic DNA sequence, a non-rat eukaryotic genomic DNA sequence, a mammalian genomic DNA sequence, or non-human DNA sequence mammalian, or a combination thereof. In one embodiment, the genomic locus comprises, in any order, rat and human genomic DNA sequences. In one embodiment, the genomic locus comprises, in any order, mouse and human genomic DNA sequences. In one embodiment, the genomic locus comprises, in any order, mouse and rat genomic DNA sequences. In one embodiment, the genomic locus comprises, in any order, rat, mouse, and human genomic DNA sequences.

[00191] In one embodiment, the genetic modification comprises at least one human disease allele of a human gene. In one embodiment, the human disease is a neurological disease. In one embodiment, the human disease is a cardiovascular disease. In one embodiment, the human disease is a kidney disease. In one embodiment, the human disease is a muscle disease. In one embodiment, the human disease is a blood disease. In one embodiment, the human disease is a cancer. In one embodiment, the human disease is an immune system disease.

[00192] 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.

[00193] In one embodiment, the genetic modification produces a mutant form of a protein with an altered binding characteristic, altered localization, altered expression, and/or altered expression pattern.

[00194] In one embodiment, the insert nucleic acid comprises a selection cassette. In one embodiment, the selection cassette comprises a nucleic acid sequence encoding a selective marker, wherein the nucleic acid sequence is operably linked to a promoter active in rat ES cells. In one embodiment, the selective marker is selected from or comprises a hygromycin resistance gene or a neomycin resistance gene.

[00195] 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.

[00196] In one embodiment, the insert nucleic acid comprises a regulatory element. In one embodiment, the regulatory element is a promoter. In one embodiment, the regulatory element is an enhancer. In one embodiment, the regulatory element is a transcriptional repressor-binding element.

[00197] In one embodiment, the genetic modification 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 element. In one embodiment, the genetic modification comprises a deletion of a regulatory element. In one embodiment, the genetic modification comprises an addition of a promoter or a regulatory element. In one embodiment, the genetic modification comprises a replacement of a promoter or a regulatory element.

ii. Expression Cassettes

[00198] Provided herein are polynucleotides or nucleic acid molecules comprising the various components employed in a targeted genomic integration system provided herein (*i.e.*, any one of or any combination of nuclease agents, recognition sites, insert nucleic acids, polynucleotides of interest, targeting vectors, selection markers, and other components).

[00199] The terms "polynucleotide," "polynucleotide sequence," "nucleic acid sequence," and "nucleic acid fragment" are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. Polynucleotides can comprise deoxyribonucleotides and

ribonucleotides include both naturally occurring molecules and synthetic analogues, and any combination these. The polynucleotides provided herein also encompass all forms of sequences including, but not limited to, single-stranded forms, double-stranded forms, hairpins, stem-and-loop structures, and the like.

[00200] Further provided are recombinant polynucleotides comprising the various components of the targeted genomic integration system. The terms "recombinant polynucleotide" and "recombinant DNA construct" are used interchangeably herein. A recombinant construct comprises an artificial or heterologous combination of nucleic acid sequences, *e.g.*, regulatory and coding sequences that are not found together in nature. In other embodiments, a recombinant construct may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. Such a construct may be used by itself or may be used in conjunction with a vector. If a vector is used, then the choice of vector is dependent upon the method that is used to transform the host cells as is well known to those skilled in the art. For example, a plasmid vector can be used. Genetic elements required to successfully transform, select, and propagate host cells comprising any of the isolated nucleic acid fragments provided herein are also provided. Screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, immunoblotting analysis of protein expression, or phenotypic analysis, among others.

[00201] In specific embodiments, one or more of the components of the targeted genomic integration system described herein can be provided in an expression cassette for expression in a prokaryotic cell, a eukaryotic cell, a non-rat eukaryotic cell, a bacterial, a yeast cell, or a mammalian cell or other organism or cell type of interest. The cassette can include 5' and 3' regulatory sequences operably linked to a polynucleotide provided herein. "Operably linked" comprises a relationship wherein the components operably linked function in their intended manner. For example, an operable linkage between a polynucleotide of interest and a regulatory sequence (i.e., a promoter) is a functional link that allows for expression of the polynucleotide of interest. Operably linked elements may be contiguous or non-contiguous. When used to refer to the joining of two protein coding regions, operably linked means that the coding regions are in the

same reading frame. In another 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.

[00202] The cassette may additionally contain at least one additional polynucleotide of interest to be co-introduced into the organism. Alternatively, the additional polynucleotide of interest can be provided on multiple expression cassettes. Such an expression cassette is provided with a plurality of restriction sites and/or recombination sites for insertion of a recombinant polynucleotide to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selection marker genes.

[00203] The expression cassette can include in the 5'-3' direction of transcription, a transcriptional and translational initiation region (i.e., a promoter), a recombinant polynucleotide provided herein, and a transcriptional and translational termination region (i.e., termination region) functional in mammalian cell or a host cell of interest. The regulatory regions (i.e., promoters, transcriptional regulatory regions, and translational termination regions) and/or a polynucleotide provided herein may be native/analogous to the host cell or to each other. Alternatively, the regulatory regions and/or a polynucleotide provided herein may be heterologous to the host cell or to each other. For example, a promoter operably linked to a heterologous polynucleotide is from a species different from the species from which the polynucleotide was derived, or, if from the same/analogous species, one or both are substantially modified from their original form and/or genomic locus, or the promoter is not the native promoter for the operably linked polynucleotide. Alternatively, the regulatory regions and/or a recombinant polynucleotide provided herein may be entirely synthetic.

[00204] The termination region may be native with the transcriptional initiation region, may be native with the operably linked recombinant polynucleotide, may be native with the host cell, or may be derived from another source (i.e., foreign or

heterologous) to the promoter, the recombinant polynucleotide, the host cell, or any combination thereof.

[00205] In preparing the expression cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation. Toward this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, in vitro mutagenesis, primer repair, restriction, annealing, resubstitutions, e.g., transitions and transversions, may be involved.

[00206] A number of promoters can be used in the expression cassettes provided herein. The promoters can be selected based on the desired outcome. It is recognized that different applications can be enhanced by the use of different promoters in the expression cassettes to modulate the timing, location and/or level of expression of the polynucleotide of interest. Such expression constructs may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible, constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

[00207] The expression cassette containing the polynucleotides provided herein can also comprise a selection marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues.

[00208] Where appropriate, the sequences employed in the methods and compositions (i.e., the polynucleotide of interest, the nuclease agent, etc.) may be optimized for increased expression in the cell. That is, the genes can be synthesized using codons preferred in a given cell of interest including, for example, mammalian-preferred codons, human-preferred codons, rodent-preferred codons, non-rat-rodent-preferred codons, mouse-preferred codons, rat-preferred codons, hamster-preferred codons, etc. for improved expression.

[00209] The various methods and compositions provided herein can employ selection markers. Various selection markers can be used in the methods and compositions disclosed herein. Such selection markers can, for example, impart

resistance to an antibiotic such as G418, hygromycin, blasticidin, neomycin, or puromycin. Such selection markers include neomycin phosphotransferase (neo^r), hygromycin B phosphotransferase (hyg^r), puromycin-N-acetyltransferase ($puro^r$), and blasticidin S deaminase (bsr^r). In still other embodiments, the selection marker is operably linked to an inducible promoter and the expression of the selection marker is toxic to the cell. Non-limiting examples of such selection markers include xanthine/guanine phosphoribosyl transferase (gpt), hypoxanthine-guanine phosphoribosyltransferase (HGPRT) or herpes simplex virus thymidine kinase (HSV-TK). The polynucleotide encoding the selection markers are operably linked to a promoter active in the cell.

iii. Targeting Vectors

[00210] Targeting vectors are employed to introduce the insert nucleic acid into the target locus of the rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster nucleic acid. The targeting vector comprises the insert nucleic acid and further comprises a 5' and a 3' homology arm, which flank the insert nucleic acid. The homology arms, which flank the insert nucleic acid, correspond to regions within the target locus of the rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster nucleic acid. For ease of reference, the corresponding cognate genomic regions within the targeted genomic locus are referred to herein as "target sites". For example, a targeting vector can comprise a first insert nucleic acid flanked by a first and a second homology arm complementary to a first and a second target site. As such, the targeting vector thereby aids in the integration of the insert nucleic acid into the target locus of the rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster nucleic acid through a homologous recombination event that occurs between the homology arms and the complementary target sites within the genome of the cell.

[00211] In one embodiment, the target locus of the rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster nucleic acid comprises a first nucleic acid sequence that is complementary to

the 5' homology arm and a second nucleic acid sequence that is complementary to the 3' homology arm. In one embodiment, the first and the second nucleic acid sequences are separated by at least 5 kb. In another embodiment, the first and the second nucleic acid sequences are separated by at least 5 kb but less than 200 kb. In one embodiment, the first and the second nucleic acid sequences are separated by at least 10 kb. In one embodiment, the first and the second nucleic acid sequences are separated by at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 110 kb, at least 120 kb, at least 130 kb, at least 140 kb, at least 150 kb, at least 160 kb, at least 170 kb, at least 180 kb, at least 190 kb, or at least 200 kb. In still further embodiments, the first and the second nucleic acid sequence is separated by at least 5 kb but less than 10 kb, at least 5 kb but less than 3 Mb, at least 10 kb but less than 20 kb, at least 20 kb but less than 40 kb, at least 40 kb but less than 60 kb, at least 60 kb but less than 80 kb, at least about 80 kb but less than 100 kb, at least 100 kb but less than 150 kb, or at least 150 kb but less than 200 kb, at least about 200 kb but less than about 300 kb, at least about 300 kb but less than about 400 kb, at least about 400 kb but less than about 500 kb, at least about 500 kb but less than about 1 Mb, at least about 1.5 Mb but less than about 2 Mb, at least about 1 Mb but less than about 1.5 Mb, at least about 2 Mb but less than 2.5 Mb, at least about 2.5 Mb but less than 3 Mb, or at least about 2 Mb but less than about 3 Mb.

[00212] A homology arm of the targeting vector can be of any length that is sufficient to promote a homologous recombination event with a corresponding target site, including for example, at least 5-10 kb, 5-15 kb, 10-20 kb, 20-30 kb, 30-40 kb, 40-50 kb, 50-60 kb, 60-70 kb, 70-80 kb, 80-90 kb, 90-100 kb, 100-110 kb, 110-120 kb, 120-130 kb, 130-140 kb, 140-150 kb, 150-160 kb, 160-170 kb, 170-180 kb, 180-190 kb, 190-200 kb in length or greater. As outlined in further detail below, large targeting vectors can employ targeting arms of greater length. In a specific embodiment, the sum total of the 5' homology arm and the 3' homology arm is at least 10 kb or the sum total of the 5' homology arm and the 3' homology arm is at least about 16 kb to about 100 kb or about 30 kb to about 100 kb. In other embodiments, the size of the sum total of the total of the 5' and 3' homology arms of the LTVEC is about 10 kb to about 150 kb, about 10 kb to

about 100 kb, about 10 kb to about 75 kb, about 20 kb to about 150 kb, about 20 kb to about 100 kb, about 20 kb to about 75 kb, about 30 kb to about 150 kb, about 30 kb to about 100 kb, about 30 kb to about 75 kb, about 40 kb to about 150 kb, about 40 kb to about 100 kb, about 40 kb to about 75 kb, about 50 kb to about 150 kb, about 50 kb to about 100 kb, or about 50 kb to about 75 kb, about 10 kb to about 30 kb, about 20 kb to about 40 kb, about 40 kb to about 60 kb, about 60 kb to about 80 kb, about 80 kb to about 100 kb, about 100 kb to about 120 kb, or from about 120 kb to about 150 kb. In one embodiment, the size of the deletion is the same or similar to the size of the sum total of the 5' and 3' homology arms of the LTVEC.

[00213] In one embodiment, the genomic locus of interest comprises (i) a 5' target sequence that is homologous to the 5' homology arm; and (ii) a 3' target sequence that is homologous to the 3' homology arm. In one embodiment, the 5' target sequence and the 3' target sequence are separated by at least 5 kb but less than 3 Mb. In still further embodiments, the 5' target sequence and the 3' target sequence are separated by at least 5 kb but less than 10 kb, at least 5 kb but less than 3 Mb, at least 10 kb but less than 20 kb, at least 20 kb but less than 40 kb, at least 40 kb but less than 60 kb, at least 60 kb but less than 80 kb, at least about 80 kb but less than 100 kb, at least 100 kb but less than 150 kb, or at least 150 kb but less than 200 kb, at least about 200 kb but less than about 300 kb, at least about 300 kb but less than about 400 kb, at least about 400 kb but less than about 500 kb, at least about 500 kb but less than about 1 Mb, at least about 1.5 Mb but less than about 2 Mb, at least about 1 Mb but less than about 1.5 Mb, at least about 2 Mb but less than 2.5 Mb, at least about 2.5 Mb but less than about 3 Mb, or at least about 2 Mb but less than about 3 Mb.

[00214] When nuclease agents are employed, the cognate genomic regions corresponding to the 5' and 3' homology arms of a targeting vector are "located in sufficient proximity" to nuclease target sites so as to promote the occurrence of a homologous recombination event between the cognate genomic regions and the homology arms upon a nick or double-strand break at the recognition site. For example, the nuclease target sites can be located anywhere between the cognate genomic regions corresponding to the 5' and 3' homology arms. In specific embodiments, the recognition site is immediately adjacent to at least one or both of the cognate genomic regions.

[00215] As used herein, a homology arm and a target site (i.e., cognate genomic region) "complement" or are "complementary" to one another when the two regions share a sufficient level of sequence identity to one another to act as substrates for a homologous recombination reaction. By "homology" is meant DNA sequences that are either identical or share sequence identity to a corresponding or "complementary" sequence. The sequence identity between a given target site and the corresponding homology arm found on the targeting vector can be any degree of sequence identity that allows for homologous recombination to occur. For example, the amount of sequence identity shared by the homology arm of the targeting vector (or a fragment thereof) and the target site (or a fragment thereof) can be at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, such that the sequences undergo homologous recombination. Moreover, a complementary region of homology between the homology arm and the complementary target site can be of any length that is sufficient to promote homologous recombination at the cleaved recognition site. For example, a given homology arm and/or complementary target site can comprise complementary regions of homology that are at least 5-10 kb, 5-15 kb, 10-20 kb, 20-30 kb, 30-40 kb, 40-50 kb, 50-60 kb, 60-70 kb, 70-80 kb, 80-90 kb, 90-100 kb, 100-110 kb, 110-120 kb, 120-130 kb, 130-140 kb, 140-150 kb, 150-160 kb, 160-170 kb, 170-180 kb, 180-190 kb, 190-200 kb, 200 kb to 300 kb in length or greater (such as described in the LTVEC vectors described elsewhere herein) such that the homology arm has sufficient homology to undergo homologous recombination with the corresponding target sites within the genome of the cell. For ease of reference the homology arms are referred to herein as a 5' and a 3' homology arm. This terminology relates to the relative position of the homology arms to the insert nucleic acid within the targeting vector.

[00216] The homology arms of the targeting vector are therefore designed to be complementary to a target site with the targeted locus. Thus, the homology arms can be complementary to a locus that is native to the cell, or alternatively they can be complementary to a region of a heterologous or exogenous segment of DNA that was integrated into the genome of the cell, including, but not limited to, transgenes, expression cassettes, or heterologous or exogenous regions of genomic DNA.

Alternatively, the homology arms of the targeting vector can be complementary to a region of a human artificial chromosome or any other engineered genomic region contained in an appropriate host cell. Still further, the homology arms of the targeting vector can be complementary to or be derived from a region of a BAC library, a cosmid library, or a P1 phage library. Thus, in specific embodiments, the homology arms of the targeting vector are complementary to a rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster genomic locus that is native, heterologous or exogenous to a given cell. In further embodiments, the homology arms are complementary to a rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster genomic locus that is not targetable using a conventional method or can be targeted only incorrectly or only with significantly low efficiency, in the absence of a nick or double-strand break induced by a nuclease agent. In one embodiment, the homology arms are derived from a synthetic DNA.

[00217] In still other embodiments, the 5' and 3' homology arms are complementary to 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 rat genome of a first strain, and the targeting arms are from a rat genome of a second strain, wherein the first strain and the second strain are different. In other embodiments, the homology arms are from the genome of the same animal or are from the genome of the same strain, e.g., the targeted genome is a rat genome of a first strain, and the targeting arms are from a rat genome from the same rat or from the same strain.

[00218] The targeting vector (such as a large targeting vector) can also comprise a selection cassette or a reporter gene as discussed elsewhere herein. The selection cassette can comprise a nucleic acid sequence encoding a selection marker, wherein the nucleic acid sequence is operably linked to a promoter. The promoter can be active in a prokaryotic cell of interest and/or active in a eukaryotic cell of interest. Such promoters can be an inducible promoter, a promoter that is endogenous to the reporter gene or the cell, a promoter that is heterologous to the reporter gene or to the cell, a cell-specific promoter, a tissue-specific promoter or a developmental stage-specific promoter. In one embodiment, the selection marker is selected from or comprises 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/or a combination thereof. The selection marker of the targeting vector can be flanked by the 5' and 3' homology arms or found either 5' or 3' to the homology arms.

[00219] In one embodiment, the targeting vector (such as a large targeting vector) comprises a reporter gene operably linked to a promoter, wherein the reporter gene encodes a reporter protein selected from the group consisting of or comprises 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/or a combination thereof. Such reporter genes can be operably linked to a promoter active in the cell. Such promoters can be an inducible promoter, a promoter that is endogenous to the report gene or the cell, a promoter that is heterologous to the reporter gene or to the cell, a cell-specific promoter, a tissue-specific promoter or a developmental stage-specific promoter.

[00220] In one embodiment, combined use of the targeting vector (including, for example, a large 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, at least three-fold, or at least 4-fold when compared to when the targeting vector is used alone.

[00221] When employing a targeting vector, the vector design can be such as to allow for the insertion of a given sequence that is from about 5 kb to about 200 kb as described herein. In one embodiment, the insertion is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 30 kb, from about 30 kb to about 40 kb, from about 40 kb to about 50 kb, from about 50 kb to about 60 kb, from about 60 kb to about 70 kb, from about 80 kb to about 90 kb, from about 90 kb to about 100 kb, from about 100 kb to about 110 kb, from about 110 kb to about 120 kb, from about 120 kb to about 130 kb, from about 130 kb to about 140 kb, from about 140 kb to about 150 kb, from about 150 kb to about 160 kb, from about 160 kb to about 170 kb,

from about 170 kb to about 180 kb, from about 180 kb to about 190 kb, or from about 190 kb to about 200 kb, from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb.

[00222] When employing a targeting vector, the vector design can be such as to allow for the replacement of a given sequence that is from about 5 kb to about 200 kb or from about 5 kb to about 3.0 Mb as described herein. In one embodiment, the replacement is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 30 kb, from about 30 kb to about 40 kb, from about 40 kb to about 50 kb, from about 50 kb to about 60 kb, from about 60 kb to about 70 kb, from about 80 kb to about 90 kb, from about 90 kb to about 100 kb, from about 100 kb to about 110 kb, from about 110 kb to about 120 kb, from about 120 kb to about 130 kb, from about 130 kb to about 140 kb, from about 140 kb to about 150 kb, from about 150 kb to about 160 kb, from about 160 kb to about 170 kb, from about 170 kb to about 180 kb, from about 180 kb to about 190 kb, from about 190 kb to about 200 kb, from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb.

[00223] In one embodiment, the targeting vector comprises a site-specific recombinase gene. In one embodiment, the site-specific recombinase gene encodes a Cre recombinase. In one embodiment, the Cre recombinase gene is Crei, wherein two exons encoding the Cre recombinase are separated by an intron to prevent its expression in a prokaryotic cell.

[00224] In one embodiment, the Cre recombinase gene further comprises a nuclear localization signal to facilitate localization of Cre (or any recombinase or nuclease agent) to the nucleus (e.g., the gene is an NL-Cre gene). In a specific embodiment, the Cre

recombinase gene further comprises a nuclear localization signal and an intron (e.g., NL-Crei).

[00225] In various embodiments, a suitable promoter for expression of the nuclease agent (including the Cre or Crei recombinase discussed above) is selected from or comprises a *Prm1*, *Blimp1*, *Gata6*, *Gata4*, *Igf2*, *Lhx2*, *Lhx5*, and/or *Pax3*. In a specific embodiment, the promoter is the *Gata6* or *Gata4* promoter. The various promoters can be from any organism, including for example, a rodent such as a mouse or a rat, a non-rat rodent, a eukaryote, a non-rat eukaryote, a non-human mammal, a mammal, a human or a hamster. In another specific embodiment, the promoter is a *Prm1* promoter. In another specific embodiment, the promoter is a rat *Prm1* promoter. In another specific embodiment, the promoter is a mouse *Prm1* promoter. In another specific embodiment, the promoter is a *Blimp1* promoter or a fragment thereof, e.g., a 1 kb or 2 kb fragment of a *Blimp1* promoter. See, for example, U.S. Patent 8,697,851 and U.S. Application Publication 2013-0312129, both of which are herein incorporated by reference in their entirety.

iv. Large Targeting Vectors

[00226] The term "large targeting vector" or "LTVEC" as used herein comprises large targeting vectors that comprise homology arms that correspond to and are derived from nucleic acid sequences larger than those typically used by other approaches intended to perform homologous targeting in cells and/or comprising insert nucleic acids comprising nucleic acid sequences larger than those typically used by other approaches intended to perform homologous recombination targeting in cells. For example, the LTVEC make possible the modification of large loci that cannot be accommodated by traditional plasmid-based targeting vectors because of their size limitations. In specific embodiments, the homology arms and/or the insert nucleic acid of the LTVEC comprises genomic sequence of a eukaryotic cell or a non-rat eukaryotic cell. 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.*, 1 kb-5 kb) PCR. Examples of the LTVEC, include, but are not limited to, vectors derived from a bacterial artificial chromosome (BAC), a human artificial chromosome or a yeast artificial chromosome (YAC). Non-

limiting examples of LTVECs 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), and US 2013/0137101, each of which is herein incorporated by reference.

[00227] The LTVEC can be of any length, including, but not limited to, from about 20 kb to about 400 kb, from about 20 kb to about 30 kb, from about 30 kb to 40 kb, from about 40 kb to about 50 kb, from about 50 kb to about 75 kb, from about 75 kb to about 100 kb, from about 100 kb to 125 kb, from about 125 kb to about 150 kb, from about 150 kb to about 175 kb, about 175 kb to about 200 kb, from about 200 kb to about 225 kb, from about 225 kb to about 250 kb, from about 250 kb to about 275 kb or from about 275 kb to about 300 kb, from about 200 kb to about 300 kb, from about 300 kb to about 350 kb, from about 350 kb to about 400 kb, from about 350 kb to about 550 kb. In one embodiment, the LTVEC is about 100 kb.

[00228] In some embodiments, the LTVEC is at least 10 kb, at least 15 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb or at least 200 kb.

[00229] In some embodiments, the LTVEC comprises a nucleic acid sequence of at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb or at least 200 kb.

[00230] In one embodiment, the LTVEC comprises an insert nucleic acid ranging from about 5 kb to about 200 kb, from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 30 kb, from about 0.5 kb to about 30 kb, from about 0.5 kb to about 40 kb, from about 30 kb to about 150 kb, from about 0.5 kb to about 150 kb, from about 30 kb to about 40 kb, from about 40 kb to about 50 kb, from about 60 kb to about 70 kb, from about 80 kb to about 90 kb, from about 90 kb to about 100 kb, from about 100 kb to about 110 kb, from about 120 kb to about 130 kb, from about 130 kb to about 140 kb, from about 140 kb to about 150 kb, from about 150 kb to about 160 kb, from about 160 kb to about 170 kb, from about 170 kb to about 180 kb, from about 180 kb to about 190 kb, or from about 190 kb to about 200 kb, from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb,

from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb.

[00231] In one embodiment, the LTVEC comprises a nucleic acid sequence of at least 100 kb, at least 150 kb, or at least 200 kb.

[00232] When employing a LTVEC, the vector design can be such as to allow for the replacement of a given sequence that is from about 5 kb to about 200 kb or from about 5 kb to about 3 Mb as described herein. In one embodiment, the replacement is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 30 kb, from about 30 kb to about 40 kb, from about 40 kb to about 50 kb, from about 50 kb to about 60 kb, from about 60 kb to about 70 kb, from about 80 kb to about 90 kb, from about 90 kb to about 100 kb, from about 100 kb to about 110 kb, from about 110 kb to about 120 kb, from about 120 kb to about 130 kb, from about 130 kb to about 140 kb, from about 140 kb to about 150 kb, from about 150 kb to about 160 kb, from about 160 kb to about 170 kb, from about 170 kb to about 180 kb, from about 180 kb to about 190 kb, from about 190 kb to about 200 kb, from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb.

[00233] In one embodiment, the homology arms of the LTVEC are derived from a BAC library, a cosmid library, or a P1 phage library. In other embodiments, the homology arms are derived from the targeted genomic locus of the cell and in some instances the target genomic locus, which the LTVEC is designed to target is not targetable using a conventional method. In still other embodiments, the homology arms are derived from a synthetic DNA.

[00234] In one embodiment, a sum total of the 5' homology arm and the 3' homology arm in the LTVEC is at least 10 kb. In other embodiments, the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 30 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about

80 kb, from about 80 kb to about 100 kb, from 100 kb to about 120 kb, from about 120 kb to about 140 kb, from about 140 kb to about 160 kb, from about 160 kb to about 180 kb, from about 180 kb to about 200 kb. In one embodiment the sum total of the 5' and the 3' homology arms of the LTVEC is from about 30 kb to about 100 kb. In other embodiments, the size of the sum total of the total of the 5' and 3' homology arms of the LTVEC is about 10 kb to about 150 kb, about 10 kb to about 100 kb, about 10 kb to about 75 kb, about 20 kb to about 150 kb, about 20 kb to about 100 kb, about 20 kb to about 75 kb, about 30 kb to about 150 kb, about 30 kb to about 100 kb, about 30 kb to about 75 kb, about 40 kb to about 150 kb, about 40 kb to about 100 kb, about 40 kb to about 75 kb, about 50 kb to about 150 kb, about 50 kb to about 100 kb, or about 50 kb to about 75 kb, about 10 kb to about 30 kb, about 20 kb to about 40 kb, about 40 kb to about 60 kb, about 60 kb to about 80 kb, about 80 kb to about 100 kb, about 100 kb to about 120 kb, or from about 120 kb to about 150 kb. In one embodiment, the size of the deletion is the same or similar to the size of the sum total of the 5' and 3' homology arms of the LTVEC.

[00235] In other embodiments, the 5' homology arm ranges from about 5 kb to about 100 kb. In one embodiment, the 3' homology arm ranges from about 5 kb to about 100 kb. In other embodiments, the sum total of the 5' and 3' homology arms are from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 30 kb, from about 30 kb to about 40 kb, from about 40 kb to about 50 kb, from about 50 kb to about 60 kb, from about 60 kb to about 70 kb, from about 70 kb to about 80 kb, from about 80 kb to about 90 kb, from about 90 kb to about 100 kb, from about 100 kb to about 110 kb, from about 110 kb to about 120 kb, from about 120 kb to about 130 kb, from about 130 kb to about 140 kb, from about 140 kb to about 150 kb, from about 150 kb to about 160 kb, from about 160 kb to about 170 kb, from about 170 kb to about 180 kb, from about 180 kb to about 190 kb, from about 190 kb to about 200 kb, or from about 30 kb to about 100 kb, about 10 kb to about 30 kb, about 20 kb to about 40 kb, about 40 kb to about 60 kb, about 60 kb to about 80 kb, about 80 kb to about 100 kb, about 100 kb to about 120 kb, or from about 120 kb to about 150 kb.

[00236] In one embodiment, the LTVEC comprises an insert nucleic acid that is homologous or orthologous to a rat nucleic acid sequence flanked by the LTVEC

homology arms. In one embodiment, the insert nucleic acid sequence is from a species other than a rat. In one embodiment, the insert nucleic acid sequence is from a eukaryote. In one embodiment, the insert nucleic acid that is homologous or orthologous to the rat nucleic acid sequence is a mammalian nucleic acid. In one embodiment, the insert nucleic acid that is homologous or orthologous to the rat nucleic acid sequence is a non-human mammalian nucleic acid. In one embodiment, the mammalian nucleic acid is a mouse nucleic acid. In one embodiment, the mammalian nucleic acid is a human nucleic acid. In one embodiment, the mammalian nucleic acid is a hamster nucleic acid. In one embodiment, the insert nucleic acid is a genomic DNA. In one embodiment, the insert is from 5 kb to 200 kb as described above.

[00237] In one embodiment, the LTVEC comprises a selection cassette or a reporter gene. Various forms of the selection cassette and reporter gene that can be employed are discussed elsewhere herein.

As described elsewhere herein, the LTVEC can also be used in the methods provided herein in combination with a nuclease agent that promotes a homologous recombination between the targeting vector and the target locus of a rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster nucleic acid in a pluripotent or non-pluripotent rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster cell.

[00238] In one embodiment, the large targeting vector (LTVEC) comprises a site-specific recombinase gene. In one embodiment, the site-specific recombinase gene encodes a Cre recombinase. In one embodiment, the Cre recombinase gene is Crei, wherein two exons encoding the Cre recombinase are separated by an intron to prevent its expression in a prokaryotic cell. In one embodiment, the Cre recombinase gene further comprises a nuclear localization signal to facilitate localization of Cre (or any recombinase or nuclease agent) to the nucleus (e.g., the gene is an NL-Cre gene). In a specific embodiment, the Cre recombinase gene further comprises a nuclear localization signal and an intron (e.g., NL-Crei)

[00239] In various embodiments, a suitable promoter for expression of the nuclease agent (including the Cre or Crei recombinase discussed above) is selected from

or comprises a *Prm1*, *Blimp1*, *Gata6*, *Gata4*, *Igf2*, *Lhx2*, *Lhx5*, and/or *Pax3*. In a specific embodiment, the promoter is the *Gata6* or *Gata4* promoter. The various promoters can be from any organism, including for example, a rodent such as a mouse or a rat, a non-rat rodent, a eukaryote, a non-rat eukaryote, a non-human mammal, a mammal, a human or a hamster. In another specific embodiment, the promoter is a *Prm1* promoter. In another specific embodiment, the promoter is a rat *Prm1* promoter. In another specific embodiment, the promoter is a mouse *Prm1* promoter. In another specific embodiment, the promoter is a *Blimp1* promoter or a fragment thereof, e.g., a 1 kb or 2 kb fragment of a *Blimp1* promoter. See, for example, U.S. Patent 8,697,851 and U.S. Application Publication 2013-0312129, both of which are herein incorporated by reference in their entirety.

[00240] In one embodiment, the LTVEC comprises an insert nucleic acid that can produce a deletion, addition, replacement or a combination thereof of a region of the rat, a eukaryotic, a non-rat eukaryotic, a mammalian, non-human mammalian, a human, a rodent, a non-rat rodent, a mouse or a hamster *ApoE* locus, the *Il2rg* locus, the *Rag2* locus, the *Rag1* locus and/or the *Rag2*/*Rag1* locus as discussed in detail elsewhere herein. In specific embodiments, the genetic modification at the *ApoE* locus results in a decrease, an increase or a modulation in ApoE activity, IL-2Rg activity, Rag2 activity, Rag1 activity and/or Rag2 and Rag1 activity. In one embodiment, an *ApoE* knockout, and *Il2rg* knockout, a *Rag2* knockout, a *Rag1* knockout, a *Rag2*/*Rag1* knockout is generated. As discussed below, nuclease agents can be employed with any of the LTVEC targeting systems to target any genomic locus of interest.

[00241] In another embodiment, the genome is exposed to a Cas protein and a CRISPR RNA in the presence of a large targeting vector (LTVEC) comprising a nucleic acid sequence of at least 10 kb. In such cases, following exposure to the Cas protein, the CRISPR RNA, and the LTVEC, the genome is modified to contain at least 10 kb of nucleic acid sequence. In specific embodiments, the LTVEC comprises a nucleic acid sequence of at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb or at least 200 kb.

v. Nuclease Agents and Recognition Sites for Nuclease Agents

[00242] As outlined in detail above, nuclease agents may be utilized in the methods and compositions disclosed herein to aid in the modification of the target locus both in a prokaryotic cell or within a pluripotent or non-pluripotent rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent ,mouse or hamster cell. Such a nuclease agent may promote homologous recombination between the targeting vector and the target locus. In one embodiment, the nuclease agent comprises an endonuclease agent.

[00243] As used herein, the term "recognition site for a nuclease agent" comprises a DNA sequence at which a nick or double-strand break is induced by a nuclease agent. The recognition site for a nuclease agent can be endogenous (or native) to the cell or the recognition site can be exogenous to the cell. In specific embodiments, the recognition site is exogenous to the cell and thereby is not naturally occurring in the genome of the cell. In still further embodiments, the recognition site is exogenous to the cell and to the polynucleotides of interest that one desired to be positioned at the target genomic locus. In further embodiments, the exogenous or endogenous recognition site is present only once in the genome of the host cell. In specific embodiments, an endogenous or native site that occurs only once within the genome is identified. Such a site can then be used to design nuclease agents that will produce a nick or double-strand break at the endogenous recognition site.

[00244] The length of the recognition site can vary, and includes, for example, recognition sites that are at least 4, 6, 8, 10, 12, 14, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70 or more nucleotides in length. In one embodiment, each monomer of the nuclease agent recognizes a recognition site of at least 9 nucleotides. In other embodiments, the recognition site is from about 9 to about 12 nucleotides in length, from about 12 to about 15 nucleotides in length, from about 15 to about 18 nucleotides in length, or from about 18 to about 21 nucleotides in length, and any combination of such subranges (e.g., 9-18 nucleotides). The recognition site could be palindromic, that is, the sequence on one

strand reads the same in the opposite direction on the complementary strand. It is recognized that a given nuclease agent can bind the recognition site and cleave that binding site or alternatively, the nuclease agent can bind to a sequence that is the different from the recognition site. Moreover, the term recognition site comprises both the nuclease agent binding site and the nick/cleavage site irrespective whether the nick/cleavage site is within or outside the nuclease agent binding site. In another variation, the cleavage by the nuclease agent can occur at nucleotide positions immediately opposite each other to produce a blunt end cut or, in other cases, the incisions can be staggered to produce single-stranded overhangs, also called "sticky ends", which can be either 5' overhangs, or 3' overhangs.

[00245] Any nuclease agent that induces a nick or double-strand break into a desired recognition site can be used in the methods and compositions disclosed herein. A naturally-occurring or native nuclease agent can be employed so long as the nuclease agent induces a nick or double-strand break in a desired recognition site. Alternatively, a modified or engineered nuclease agent can be employed. An "engineered nuclease agent" comprises a nuclease that is engineered (modified or derived) from its native form to specifically recognize and induce a nick or double-strand break in the desired recognition site. Thus, an engineered nuclease agent can be derived from a native, naturally-occurring nuclease agent or it can be artificially created or synthesized. The modification of the nuclease agent can be as little as one amino acid in a protein cleavage agent or one nucleotide in a nucleic acid cleavage agent. In some embodiments, the engineered nuclease induces a nick or double-strand break in a recognition site, wherein the recognition site was not a sequence that would have been recognized by a native (non-engineered or non-modified) nuclease agent. Producing a nick or double-strand break in a recognition site or other DNA can be referred to herein as "cutting" or "cleaving" the recognition site or other DNA.

[00246] Active variants and fragments of the exemplified recognition sites are also provided. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the given recognition site, wherein the active variants retain biological activity and hence are capable of being recognized and cleaved by a nuclease agent in a sequence-specific

manner. Assays to measure the double-strand break of a recognition site by a nuclease agent are known in the art and generally measure the ability of a nuclease to cut the recognition site.

[00247] The recognition site of the nuclease agent can be positioned anywhere in or near the target locus. The recognition site can be located within a coding region of a gene, or within regulatory regions, which influence expression of the gene. Thus, a recognition site of the nuclease agent can be located in an intron, an exon, a promoter, an enhancer, a regulatory region, or any non-protein coding region.

[00248] In one embodiment, the nuclease agent is a Transcription Activator-Like Effector Nuclease (TALEN). TAL effector nucleases are a class of sequence-specific nucleases that can be used to make double-strand breaks at specific target sequences in the genome of a prokaryotic or eukaryotic organism. TAL effector nucleases are created by fusing a native or engineered transcription activator-like (TAL) effector, or functional part thereof, to the catalytic domain of an endonuclease, such as, for example, *FokI*. The unique, modular TAL effector DNA binding domain allows for the design of proteins with potentially any given DNA recognition specificity. Thus, the DNA binding domains of the TAL effector nucleases can be engineered to recognize specific DNA target sites and thus, used to make double-strand breaks at desired target sequences. *See*, WO 2010/079430; Morbitzer *et al.* (2010) *PNAS* 10.1073/pnas.1013133107; Scholze & Boch (2010) *Virulence* 1:428-432; Christian *et al. Genetics* (2010) 186:757-761; Li *et al.* (2010) *Nuc. Acids Res.* (2010) doi:10.1093/nar/gkq704; and Miller *et al.* (2011) *Nature Biotechnology* 29:143–148; all of which are herein incorporated by reference.

[00249] 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 genomic locus of interest, wherein the target nucleic acid sequence is at or near a sequence to be modified by a targeting vector. The TAL nucleases suitable for use with the various methods and compositions provided herein include those that are specifically designed to bind at or

near target nucleic acid sequences to be modified by targeting vectors as described herein.

[00250] In one embodiment, each monomer of the TALEN comprises 12-25 TAL repeats, wherein each TAL repeat binds a 1 bp 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 6 bp to about 40 bp cleavage site, and wherein the FokI nucleases dimerize and make a double strand break at a target sequence.

[00251] 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 a 5 bp or 6 bp cleavage site, and wherein the FokI nucleases dimerize and make a double strand break.

[00252] The nuclease agent employed in the various methods and compositions disclosed herein can further comprise 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 3 bp subsite. In other embodiments, 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 6 bp to about 40 bp cleavage site or about a 5 bp to about 6

bp cleavage site, and wherein the FokI nucleases dimerize and make a double strand break. See, for example, US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308A2; US20130123484; US20100291048; and, WO/2011/017293A2, each of which is herein incorporated by reference.

[00253] In one embodiment of the methods provided herein, the nuclease agent comprises (a) a chimeric protein comprising a zinc finger-based DNA binding domain fused to a FokI endonuclease; or (b) a chimeric protein comprising a Transcription Activator-Like Effector Nuclease (TALEN) fused to a FokI endonuclease.

[00254] In still another embodiment, the nuclease agent is a meganuclease. Meganucleases have been classified into four families based on conserved sequence motifs, the families are the LAGLIDADG (SEQ ID NO: 16), GIY-YIG, H-N-H, and His-Cys box families. These motifs participate in the coordination of metal ions and hydrolysis of phosphodiester bonds. HEases are notable for their long recognition sites, and for tolerating some sequence polymorphisms in their DNA substrates. Meganuclease domains, structure and function are known, see for example, Guhan and Muniyappa (2003) *Crit Rev Biochem Mol Biol* 38:199-248; Lucas *et al.*, (2001) *Nucleic Acids Res* 29:960-9; Jurica and Stoddard, (1999) *Cell Mol Life Sci* 55:1304-26; Stoddard, (2006) *Q Rev Biophys* 38:49-95; and Moure *et al.*, (2002) *Nat Struct Biol* 9:764. In some examples a naturally occurring variant, and/or engineered derivative meganuclease is used. Methods for modifying the kinetics, cofactor interactions, expression, optimal conditions, and/or recognition site specificity, and screening for activity are known, see for example, Epinat *et al.*, (2003) *Nucleic Acids Res* 31:2952-62; Chevalier *et al.*, (2002) *Mol Cell* 10:895-905; Gimble *et al.*, (2003) *Mol Biol* 334:993-1008; Seligman *et al.*, (2002) *Nucleic Acids Res* 30:3870-9; Sussman *et al.*, (2004) *J Mol Biol* 342:31-41; Rosen *et al.*, (2006) *Nucleic Acids Res* 34:4791-800; Chames *et al.*, (2005) *Nucleic Acids Res* 33:e178; Smith *et al.*, (2006) *Nucleic Acids Res* 34:e149; Gruen *et al.*, (2002) *Nucleic Acids Res* 30:e29; Chen and Zhao, (2005) *Nucleic Acids Res* 33:e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; and WO2004031346.

[00255] Any meganuclease can be used herein, including, but not limited to, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-

CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, PI-PspI, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoeI, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NcIIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, PI-TliII, or any active variants or fragments thereof.

[00256] 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 (SEQ ID NO: 16) family of homing nuclease. In one embodiment, the LAGLIDADG (SEQ ID NO: 16) family of homing nuclease is selected from I-SceI, I-CreI, and I-Dmol.

[00257] Nuclease agents can further comprise restriction endonucleases, which include Type I, Type II, Type III, and Type IV endonucleases. Type I and Type III restriction endonucleases recognize specific recognition sites, but typically cleave at a variable position from the nuclease binding site, which can be hundreds of base pairs away from the cleavage site (recognition site). In Type II systems the restriction activity is independent of any methylase activity, and cleavage typically occurs at specific sites within or near to the binding site. Most Type II enzymes cut palindromic sequences, however Type IIa enzymes recognize non-palindromic recognition sites and cleave outside of the recognition site, Type IIb enzymes cut sequences twice with both sites outside of the recognition site, and Type IIs enzymes recognize an asymmetric recognition site and cleave on one side and at a defined distance of about 1-20 nucleotides from the recognition site. Type IV restriction enzymes target methylated DNA. Restriction enzymes are further described and classified, for example in the REBASE database (webpage at rebase.neb.com; Roberts *et al.*, (2003) *Nucleic Acids Res* 31:418-20), Roberts *et al.*, (2003) *Nucleic Acids Res* 31:1805-12, and Belfort *et al.*,

(2002) in *Mobile DNA II*, pp. 761-783, Eds. Craigie *et al.*, (ASM Press, Washington, DC).

[00258] The nuclease agent employed in the various methods and compositions can also comprise a CRISPR/Cas system. Such systems can employ, for example, a Cas9 nuclease, which in some instances, is codon-optimized for the desired cell type in which it is to be expressed. Such systems can also employ a guide RNA (gRNA) that comprises two separate molecules. An exemplary two-molecule gRNA comprises a crRNA-like ("CRISPR RNA" or "targeter-RNA" or "crRNA" or "crRNA repeat") molecule and a corresponding tracrRNA-like ("trans-acting CRISPR RNA" or "activator-RNA" or "tracrRNA" or "scaffold") molecule. A crRNA comprises both the DNA-targeting segment (single stranded) of the gRNA and a stretch of nucleotides that forms one half of a double stranded RNA (dsRNA) duplex of the protein-binding segment of the gRNA. A corresponding tracrRNA (activator-RNA) comprises a stretch of nucleotides that forms the other half of the dsRNA duplex of the protein-binding segment of the gRNA. Thus, a stretch of nucleotides of a crRNA are complementary to and hybridize with a stretch of nucleotides of a tracrRNA to form the dsRNA duplex of the protein-binding domain of the gRNA. As such, each crRNA can be said to have a corresponding tracrRNA. The crRNA additionally provides the single stranded DNA-targeting segment. Accordingly, a gRNA comprises a sequence that hybridizes to a target sequence, and a tracrRNA. Thus, a crRNA and a tracrRNA (as a corresponding pair) hybridize to form a gRNA. If used for modification within a cell, the exact sequence and/or length of a given crRNA or tracrRNA molecule can be designed to be specific to the species in which the RNA molecules will be used.

[00259] Naturally occurring genes encoding the three elements (Cas9, tracrRNA and crRNA) are typically organized in operon(s). Naturally occurring CRISPR RNAs differ depending on the Cas9 system and organism but often contain a targeting segment of between 21 to 72 nucleotides length, flanked by two direct repeats (DR) of a length of between 21 to 46 nucleotides (see, e.g., WO2014/131833). In the case of *S. pyogenes*, the DRs are 36 nucleotides long and the targeting segment is 30 nucleotides long. The 3' located DR is complementary to and hybridizes with the corresponding tracrRNA, which in turn binds to the Cas9 protein.

[00260] Alternatively, the system further employs a fused crRNA-tracrRNA construct (i.e., a single transcript) that functions with the codon-optimized Cas9. This single RNA is often referred to as a guide RNA or gRNA. Within a gRNA, the crRNA portion is identified as the 'target sequence' for the given recognition site and the tracrRNA is often referred to as the 'scaffold.' Briefly, a short DNA fragment containing the target sequence is inserted into a guide RNA expression plasmid. The gRNA expression plasmid comprises the target sequence (in some embodiments around 20 nucleotides), a form of the tracrRNA sequence (the scaffold) as well as a suitable promoter that is active in the cell and necessary elements for proper processing in eukaryotic cells. Many of the systems rely on custom, complementary oligos that are annealed to form a double stranded DNA and then cloned into the gRNA expression plasmid. The gRNA expression cassette and the Cas9 expression cassette are then introduced into the cell. See, for example, Mali P et al. (2013) *Science* 2013 Feb 15;339(6121):823-6; Jinek M et al. *Science* 2012 Aug 17;337(6096):816-21; Hwang WY et al. *Nat Biotechnol* 2013 Mar;31(3):227-9; Jiang W et al. *Nat Biotechnol* 2013 Mar;31(3):233-9; and Cong L et al. *Science* 2013 Feb 15;339(6121):819-23, each of which is herein incorporated by reference. See also, for example, WO/2013/176772A1, WO/2014/065596A1, WO/2014/089290A1, WO/2014/093622A2, WO/2014/099750A2, and WO/2013142578A1, each of which is herein incorporated by reference.

[00261] In some embodiments, the Cas9 nuclease can be provided in the form of a protein. In some embodiments, the Cas9 protein can be provided in the form of a complex with the gRNA. In other embodiments, the Cas9 nuclease can be provided in the form of a nucleic acid encoding the protein. The nucleic acid encoding the Cas9 nuclease can be RNA (e.g., messenger RNA (mRNA)) or DNA.

[00262] In some embodiments, the gRNA can be provided in the form of RNA. In other embodiments, the gRNA can be provided in the form of DNA encoding the RNA. In some embodiments, the gRNA can be provided in the form of separate crRNA and tracrRNA molecules, or separate DNA molecules encoding the crRNA and tracrRNA, respectively.

[00263] In one embodiment, the method for modifying a genomic locus of interest in a cell further comprises introducing into the cell: (a) a first expression construct

comprising a first promoter operably linked to a first nucleic acid sequence encoding a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) protein; (b) a second expression construct comprising a second promoter operably linked to a genomic target sequence linked to a guide RNA (gRNA), wherein the genomic target sequence is flanked by a Protospacer Adjacent Motif. Optionally, the genomic target sequence is flanked on the 3'end by a Protospacer Adjacent Motif (PAM) sequence. In one embodiment, the cell comprises a eukaryotic cell, a non-rat eukaryotic cell, a mammalian cell, a human cell, a non-human mammalian cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, a human iPS cell, a human cell, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell, a fibroblast, or a CHO cell.

[00264] In one embodiment, the genomic target sequence comprises the nucleotide sequence of GNNNNNNNNNNNNNNNNNNGG (GN_{1-20} GG; SEQ ID NO: 1). In one embodiment, the genomic target sequence comprises SEQ ID NO: 23, wherein N is between 1 and 20 nucleotides in length. In another embodiment, the genomic target sequence comprises between 14 and 20 nucleotides in length of SEQ ID NO: 1.

[00265] In one embodiment, the gRNA comprises a third nucleic acid sequence encoding a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). In specific embodiments, the Cas protein is Cas9.

[00266] In some embodiments, the gRNA comprises (a) the chimeric RNA of the nucleic acid sequence 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' (SEQ ID NO: 2); or (b) the chimeric RNA of the nucleic acid sequence 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCG-3' (SEQ ID NO: 3).

[00267] In another embodiment, the crRNA comprises 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAU-3' (SEQ ID NO: 4); 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAG (SEQ ID NO: 5); or 5'-GAGUCCGAGCAGAAGAAGAAGUUUUA-3' (SEQ ID NO: 6).

[00268] In yet other embodiments, the tracrRNA comprises, 5'-AAGGCUAGUCCG-3' (SEQ ID NO: 7) or 5'-AAGGCUAGUCCGU UAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' (SEQ ID NO: 8).

[00269] In one embodiment, the Cas protein is a type I Cas protein. In one embodiment, the Cas protein is a type II Cas protein. In one embodiment, the type II Cas protein is Cas9. In one embodiment, the first nucleic acid sequence encodes a human codon-optimized Cas protein.

[00270] In certain embodiments, the Cas protein is a "nickase" that can create single strand breaks (i.e., "nicks") at the target site without cutting both strands of double stranded DNA (dsDNA). Cas9, for example, comprises two nuclease domains—a RuvC-like nuclease domain and an HNH-like nuclease domain—which are responsible for cleavage of opposite DNA strands. Mutation in either of these domains can create a nickase. Examples of mutations creating nickases can be found, for example, WO/2013/176772A1 and WO/2013/142578A1, each of which is herein incorporated by reference.

[00271] In certain embodiments, two separate Cas proteins (e.g., nickases) specific for a target site on each strand of dsDNA can create overhanging sequences complementary to overhanging sequences on another nucleic acid, or a separate region on the same nucleic acid. The overhanging ends created by contacting a nucleic acid with two nickases specific for target sites on both strands of dsDNA can be either 5′ or 3′ overhanging ends. For example, a first nickase can create a single strand break on the first strand of dsDNA, while a second nickase can create a single strand break on the second strand of dsDNA such that overhanging sequences are created. The target sites of each nickase creating the single strand break can be selected such that the overhanging end sequences created are complementary to overhanging end sequences on a different nucleic acid molecule. The complementary overhanging ends of the two different nucleic acid molecules can be annealed by the methods disclosed herein. In some embodiments, the target site of the nickase on the first strand is different from the target site of the nickase on the second strand.

[00272] In one embodiment, the first nucleic acid comprises a mutation that disrupts at least one amino acid residue of nuclease active sites in the Cas protein,

wherein the mutant Cas protein generates a break in only one strand of the target DNA region, and wherein the mutation diminishes nonhomologous recombination in the target DNA region.

[00273] In one embodiment, the first nucleic acid that encodes the Cas protein further comprises a nuclear localization signal (NLS). In one embodiment, the nuclear localization signal is a SV40 nuclear localization signal.

[00274] In one embodiment, the second promoter that drives the expression of the genomic target sequence and the guide RNA (gRNA) is an RNA polymerase III promoter. In one embodiment, the RNA polymerase III promoter is a human U6 promoter. In one embodiment, the RNA polymerase III promoter is a rat U6 polymerase III promoter. In one embodiment, the RNA polymerase III promoter is a mouse U6 polymerase III promoter.

[00275] In one embodiment, the nucleic acid sequences encoding crRNA and the tracrRNA are linked via a synthetic loop, wherein, upon expression, the crRNA and the tracrRNA forms a crRNA:tracrRNA duplex.

[00276] The CRISPR/Cas system as described above can be used in combination with large targeting vectors with any of the following cell types: a eukaryotic cell, a non-rat eukaryotic cell, a mammalian cell, a non-human mammalian cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, a human iPS cell, a human cell, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell, a fibroblast or a CHO cell.

[00277] In one embodiment, the first expression construct and the second expression construct are expressed from a same plasmid.

[00278] In one embodiment, the first and the second expression constructs are introduced together with the LTVEC. In one embodiment, the first and the second expression constructs are introduced separately from the LTVEC over a period of time.

[00279] In one embodiment, the method comprises introducing a plurality of the second construct and a plurality of the LTVEC for multiplex editing of distinct target loci as described herein.

[00280] Active variants and fragments of nuclease agents (i.e., an engineered nuclease agent) are also provided. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the native nuclease agent, wherein the active variants retain the ability to cut at a desired recognition site and hence retain nick or double-strand-break-inducing activity. For example, any of the nuclease agents described herein can be modified from a native endonuclease sequence and designed to recognize and induce a nick or double-strand break at a recognition site that was not recognized by the native nuclease agent. Thus in some embodiments, the engineered nuclease has a specificity to induce a nick or double-strand break at a recognition site that is different from the corresponding native nuclease agent recognition site. Assays for nick or double-strand-break-inducing activity are known and generally measure the overall activity and specificity of the endonuclease on DNA substrates containing the recognition site.

[00281] The nuclease agent may be introduced into the cell by any means known in the art. The polypeptide encoding the nuclease agent may be directly introduced into the cell. Alternatively, a polynucleotide encoding the nuclease agent can be introduced into the cell. When a polynucleotide encoding the nuclease agent is introduced into the cell, the nuclease agent can be transiently, conditionally or constitutively expressed within the cell. Thus, the polynucleotide encoding the nuclease agent can be contained in an expression cassette and be operably linked to a conditional promoter, an inducible promoter, a constitutive promoter, or a tissue-specific promoter. Such promoters of interest are discussed in further detail elsewhere herein. Alternatively, the nuclease agent is introduced into the cell as an mRNA encoding or comprising a nuclease agent.

[00282] In one embodiment, the crRNA and the tracrRNA are expressed as separate RNA transcripts.

[00283] In specific embodiments, the polynucleotide encoding the nuclease agent is stably integrated in the genome of the cell and operably linked to a promoter active in the cell. In other embodiments, the polynucleotide encoding the nuclease agent is in the same targeting vector comprising the insert nucleic acid, while in other instances the polynucleotide encoding the nuclease agent is in a vector or a plasmid that is separate from the targeting vector comprising the insert nucleic acid.

[00284] When the nuclease agent is provided to the cell through the introduction of a polynucleotide encoding the nuclease agent, such a polynucleotide encoding a nuclease agent can be modified to substitute codons having a higher frequency of usage in the cell of interest, as compared to the naturally occurring polynucleotide sequence encoding the nuclease agent. For example the polynucleotide encoding the nuclease agent can be modified to substitute codons having a higher frequency of usage in a given prokaryotic or eukaryotic cell of interest, including a bacterial cell, a yeast cell, a human cell, a non-human cell, a non-rat eukaryotic cell, a mammalian cell, a rodent cell, a non-rat rodent cell, a mouse cell, a rat cell, a hamster cell or any other host cell of interest, as compared to the naturally occurring polynucleotide sequence.

[00285] In one embodiment, the endonuclease agent is introduced together with the LTVEC. In one embodiment, the endonuclease agent is introduced separately from the LTVEC over a period of time. In one embodiment, the endonuclease agent is introduced prior to the introduction of the LTVEC. In one embodiment, the endonuclease agent is introduced into the rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster ES cell following introduction of the LTVEC.

[00286] In one embodiment, the endonuclease agent is an expression construct comprising a nucleic acid sequence encoding an endonuclease, 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 pluripotent or non-pluripotent rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster cell. In one embodiment, the endonuclease agent is an mRNA encoding an endonuclease.

B. Methods for Integrating a Polynucleotide of Interest Into a Target Locus

[00287] Methods for modifying a target locus of interest are provided. In one embodiment, a target locus in a pluripotent or non-pluripotent rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster cell is targeted for genetic modification. Such a method comprises: (a)

introducing into the pluripotent or non-pluripotent rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster cell a targeting vector comprising an insert nucleic acid flanked with a 5' rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster homology arm and a 3' rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster homology arm; and (b) identifying a genetically modified pluripotent or non-pluripotent rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster cell comprising the targeted genetic modification at the target locus, wherein the targeted genetic modification is capable of being transmitted through the germline. In specific embodiments, the sum total of the 5' homology arm and the 3' homology arm is at least 10 kb and/or a large targeting vector is employed.

[00288] In other embodiments, the size of the sum total of the total of the 5' and 3' homology arms of the LTVEC is about 10 kb to about 150 kb, about 10 kb to about 100 kb, about 10 kb to about 75 kb, about 20 kb to about 150 kb, about 20 kb to about 100 kb, about 20 kb to about 75 kb, about 30 kb to about 150 kb, about 30 kb to about 100 kb, about 30 kb to about 75 kb, about 40 kb to about 150 kb, about 40 kb to about 100 kb, about 40 kb to about 75 kb, about 50 kb to about 150 kb, about 50 kb to about 100 kb, or about 50 kb to about 75 kb, about 10 kb to about 30 kb, about 20 kb to about 40 kb, about 40 kb to about 60 kb, about 60 kb to about 80 kb, about 80 kb to about 100 kb, about 100 kb to about 120 kb, or from about 120 kb to about 150 kb. In one embodiment, the size of the deletion is the same or similar to the size of the sum total of the 5' and 3' homology arms of the LTVEC.

[00289] The pluripotent cell, for example, a rat cell, can be an embryonic stem cell, for example, a rat embryonic stem cell. In a specific embodiment, (a) the rat ES cell is derived from a DA strain or an ACI strain; or (b) the rat ES cell is characterized by expression of a pluripotency marker comprising Oct-4, Sox-2, alkaline phosphatase, or a combination thereof. In other instances, the rat embryonic stem cell employed comprises a rat ES cell as described in U.S. Patent Application No. 14/185,103, filed on February 20, 2014, herein incorporated by reference in its entirety.

[00290] Any pluripotent or non-pluripotent cell can be used in the methods provided herein. For example, the pluripotent or non-pluripotent cell can be from a eukaryote, a non-rat eukaryote, a non-human mammal, a mammal, a rodent, a non-rat rodent, a rat, a mouse, a human or a hamster.

[00291] As described elsewhere herein, the insert nucleic acid can be any nucleic acid sequence. In non-limiting embodiments, (a) the insert nucleic acid comprises a replacement of an endogenous rat, eukaryotic, non-rat eukaryotic, mammalian, human, rodent, non-rat rodent, mouse or hamster nucleic acid sequence with a homologous or a orthologous mammalian nucleic acid sequence; (b) the insert nucleic acid comprises a deletion of an endogenous rat, eukaryotic, non-rat eukaryotic, mammalian, human, rodent, non-rat rodent, mouse or hamster nucleic acid sequence; (c) the insert nucleic acid comprises a deletion of an endogenous rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammalian, human, rodent, non-rat rodent, mouse or hamster nucleic acid sequence, wherein the deletion ranges from 5 kb to 200 kb or from 5 kb to 3 Mb (as discussed in detail elsewhere herein); (d) the insert nucleic acid comprises an addition of an exogenous nucleic acid sequence (including for example an exogenous nucleic acid sequence ranging from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb); (e) the insert nucleic acid comprises an exogenous nucleic acid sequence comprising a homologous or an orthologous nucleic acid sequence; (f) the homologous or the orthologous nucleic acid sequence of (a) wherein the nucleic acid sequence is a human nucleic acid sequence; (g) the insert nucleic acid comprises the homologous or the orthologous nucleic acid sequence of (a) wherein the nucleic acid sequence is a chimeric nucleic acid sequence comprising a human and a rat nucleic acid sequence; (h) the insert nucleic acid comprises the exogenous nucleic acid sequence of (e), wherein the insert nucleic acid ranges from about 5 kb to about 200 kb; (i) the insert nucleic acid comprises a conditional allele flanked with site-specific recombinase target sequences; (j) the insert nucleic acid comprises a reporter gene operably linked to a promoter; (k) the

insert nucleic acid comprises one or more unrearranged human immunoglobulin heavy chain V_H gene segments, one or more unrearranged human immunoglobulin heavy chain D gene segments, and one or more unrearranged human immunoglobulin heavy chain J_H gene segments, which are operably linked to a rodent heavy chain constant region nucleic acid sequence; (l) the insert nucleic acid comprises a rearranged human immunoglobulin heavy chain variable region nucleic acid sequence operably linked to a rodent heavy chain constant region nucleic acid sequence; (m) the insert nucleic acid comprises one or more unrearranged human immunoglobulin V_κ or V_λ gene segments and one or more unrearranged human immunoglobulin J_κ or J_λ gene segments, which are operably linked to a mammalian immunoglobulin λ or κ light chain light chain constant region nucleic acid sequence; (n) the insert nucleic acid comprises a rearranged human immunoglobulin λ or κ light chain variable region nucleic acid sequence operably linked to a mammalian immunoglobulin λ or κ light chain light chain constant region nucleic acid sequence; (o) the mammalian heavy chain constant region nucleic acid sequence of (k) and/or (l) comprises a rat constant region nucleic acid sequence, a human constant region nucleic acid sequence, or a combination thereof; or (p) the mammalian immunoglobulin λ or κ light chain constant region nucleic acid of (m) and/or (n) comprises a rat constant region nucleic acid sequence, a human constant region nucleic acid sequence, or a combination thereof.

[00292] In one embodiment, the insert nucleic acid comprises one or more functional human V_H gene segments comprising V_H1-2, V_H1-3, V_H1-8, V_H1-18, V_H1-24, V_H1-45, V_H1-46, V_H1-58, V_H1-69, V_H2-5, V_H2-26, V_H2-70, V_H3-7, V_H3-9, V_H3-11, V_H3-13, V_H3-15, V_H3-16, V_H3-20, V_H3-21, V_H3-23, V_H3-30, V_H3-30-3, V_H 3-30-5, V_H3-33, V_H3-35, V_H3-38, V_H3-43, V_H3-48, V_H3-49, V_H3-53, V_H3-64, V_H3-66, V_H3-72, V_H3-73, V_H3-74, V_H4-4, V_H4-28, V_H4-30-1, V_H4-30-2, V_H4-30-4, V_H4-31, V_H4-34, V_H4-39, V_H4-59, V_H4-61, V_H5-51, V_H6-1, V_H7-4-1, V_H7-81, or a combination thereof.

[00293] In one embodiment, the insert nucleic acid comprises one or more functional human D gene segments comprising D1-1, D1-7, D1-14, D1-20, D1-26, D2-2, D2-8, D2-15, D2-21, D3-3, D3-9, D3-10, D3-16, D3-22, D4-4, D4-11, D4-17, D4-23, D5-12, D5-5, D5-18, D5-24, D6-6, D6-13, D6-19, D6-25, D7-27, or a combination thereof.

[00294] In one embodiment, the insert nucleic acid comprises one or more functional J_H gene segments comprising J_H1, J_H2, J_H3, J_H4, J_H5, J_H6, or a combination thereof. In one embodiment, the insert nucleic acid comprises one or more human Vκ gene segments comprising Vκ4-1, Vκ5-2, Vκ 7-3, Vκ 2-4, Vκ1-5, Vκ1-6, Vκ3-7, Vκ1-8, Vκ1-9, Vκ2-10, Vκ3-11, Vκ1-12, Vκ1-13, Vκ2-14, Vκ3-15, Vκ1-16, Vκ1-17, Vκ2-18, Vκ2-19, Vκ3-20, Vκ6-21, Vκ1-22, Vκ1-23, Vκ2-24, Vκ3-25, Vκ2-26, Vκ1-27, Vκ2-28, Vκ2-29, Vκ2-30, Vκ3-31, Vκ1-32, Vκ1-33, Vκ3-34, Vκ1-35, Vκ2-36, Vκ1-37, Vκ2-38, Vκ1-39, Vκ2-40, or a combination thereof.

[00295] In one embodiment, the insert nucleic acid comprises one or more human Vλ gene segments comprising Vλ3-1, Vλ4-3, Vλ2-8, Vλ3-9, Vλ3-10, Vλ2-11, Vλ3-12, Vλ2-14, Vλ3-16, Vλ2-18, Vλ3-19, Vλ3-21, Vλ3-22, Vλ2-23, Vλ3-25, Vλ3-27, or a combination thereof.

[00296] In one embodiment, the insert nucleic acid comprises one or more human Jκ gene segments comprising Jκ1, Jκ2, Jκ3, Jκ4, Jκ5, or a combination thereof.

[00297] In specific embodiments, upon modification of the target locus in a pluripotent or non-pluripotent rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammal, human, rodent, non-rat rodent, mouse or hamster cell, the genetic modification is transmitted through the germline.

[00298] In one embodiment, the insert nucleic acid sequence comprises a polynucleotide that when integrated into the genome will produce a genetic modification of a region of the rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammal, human, rodent, non-rat rodent, mouse or hamster *ApoE* locus, wherein the genetic modification at the *ApoE* locus results in a decrease in ApoE activity, an increase in ApoE activity or a modulation of ApoE activity. In one embodiment, an *ApoE* knockout is generated.

[00299] In one embodiment, the insert nucleic acid sequence comprises a polynucleotide that when integrated into the genome will produce a genetic modification of a region of the rat, eukaryotic, non-rat eukaryotic, mammal, human, non-human mammal, rodent, non-rat rodent, mouse or hamster interleukin-2 receptor gamma locus, wherein the genetic modification at the interleukin-2 receptor gamma locus results in a decrease in interleukin-2 receptor activity, an increase in interleukin-2 receptor gamma

activity, or a modulation of interleukin-2 receptor activity. In one embodiment, an interleukin-2 receptor knockout is generated.

[00300] In still another embodiment, the insert nucleic acid sequence comprises a polynucleotide that when integrated into the genome will produce a genetic modification of a region of the rat, eukaryotic, non-rat eukaryotic, mammal, non-human mammal, human, rodent, non-rat rodent, mouse or hamster *Rag1* locus, the rat, eukaryotic, non-rat eukaryotic, non-human mammal, mammalian, human, rodent, non-rat rodent, mouse or hamster *Rag2* locus and/or the rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammal, human, rodent, non-rat rodent, mouse or hamster *Rag2/Rag1* locus, wherein the genetic modification at the rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammal, human, rodent, non-rat rodent, mouse or hamster *Rag1*, *Rag2* and/or *Rag2/Rag1* locus results in a decrease in in Rag1, Rag2 or Rag1 and Rag2 protein activity, an increase in Rag1, Rag2 or Rag1 and Rag2 protein activity, or a modulation in Rag1, Rag2 or Rag 1 and Rag2 protein activity. In one embodiment, a *Rag1*, *Rag2* or *Rag2/Rag1* knockout is generated.

[00301] In further embodiments, the insert nucleic acid results in the replacement of a portion of the rat, eukaryotic, non-rat eukaryotic, mammalian, non-human mammal, human, rodent, non-rat rodent, mouse or hamster *ApoE* locus, the interleukin-2 receptor gamma locus and/or *Rag2* locus, and/or *Rag1* locus and/or *Rag2/Rag1* locus with the corresponding orthologous portion of an *ApoE* locus, an interleukin-2 receptor gamma locus, a *Rag2* locus, a *Rag1* locus and/or a *Rag2/Rag1* locus from another organism.

[00302] In still other embodiments, the insert nucleic acid comprises a polynucleotide sharing across its full length least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% to a portion of an *ApoE* locus, an interleukin-2 receptor gamma locus, a *Rag2* locus, a *Rag1* locus and/or a *Rag2/Rag1* locus it is replacing.

[00303] The given insert polynucleotide and the corresponding region of the rat, eukaryotic, non-rat eukaryotic, mammal, non-human mammal, human, rodent, non-rat rodent, mouse or hamster locus being replaced can be a coding region, an intron, an exon, an untranslated region, a regulatory region, a promoter, or an enhancer or any combination thereof. Moreover, the given insert polynucleotide and/or the region of the rat, eukaryotic, non-rat eukaryotic, mammalian, human, non-human mammal, rodent,

non-rat rodent, mouse or hamster locus being replaced can be of any desired length, including for example, between 10-100 nucleotides in length, 100-500 nucleotides in length, 500-1 kb nucleotide in length, 1 kb to 1.5 kb nucleotide in length, 1.5 kb to 2 kb nucleotides in length, 2 kb to 2.5 kb nucleotides in length, 2.5 kb to 3 kb nucleotides in length, 3 kb to 5 kb nucleotides in length, 5 kb to 8 kb nucleotides in length, 8 kb to 10 kb nucleotides in length or more. In other instances, the size of the insertion or replacement is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, from about 350 kb to about 400 kb, from about 400 kb to about 800 kb, from about 800 kb to 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb, to about 2.5 Mb, from about 2.5 Mb to about 2.8 Mb, from about 2.8 Mb to about 3 Mb. In other embodiments, the given insert polynucleotide and/or the region of the rat, eukaryotic, non-rat eukaryotic, non-human mammal, mammal, human, rodent, non-rat rodent, mouse or hamster locus being replaced is at least 100, 200, 300, 400, 500, 600, 700, 800, or 900 nucleotides or at least 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb, 12 kb, 13 kb, 14 kb, 15 kb, 16 kb or greater.

i. Methods for Modifying a Target Locus of a Nucleic Acid via Bacterial Homologous Recombination (BHR)

[00304] Methods and compositions are provided for modifying a target locus of a eukaryotic, non-rat eukaryotic, a mammalian, a human or a non-human mammalian nucleic acid, via bacterial homologous recombination (BHR) in a prokaryotic cell. Such methods find use in utilizing bacterial homologous recombination in a prokaryotic cell to genetically modify a target locus of a eukaryotic, non-rat eukaryotic, a mammalian, a human or a non-human mammalian nucleic acid in order to create a targeting vector. Such a targeting vector comprising the genetically modified target locus can be introduced into a eukaryotic cell, for example, a eukaryotic cell, non-rat eukaryotic cell, a mammalian cell, a human cell, a non-human mammalian cell, a pluripotent cell, a non-

pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, a human iPS cell, a human cell, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell, a fibroblast, or a CHO cell. "Homologous recombination" includes the exchange of DNA fragments between two DNA molecules at cross-over sites within regions of homology. Thus, "bacterial homologous recombination" or "BHR" includes homologous recombination that occurs in bacteria.

[00305] Methods for modifying a target locus of a nucleic acid from a eukaryotic cell, non-rat eukaryotic cell, a mammalian cell, a human cell, a non-human mammalian cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, a human iPS cell, a human cell, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell, a fibroblast, or a CHO cell via bacterial homologous recombination (BHR) are provided. The methods comprise introducing into a prokaryotic cell a targeting vector comprising an insert nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the prokaryotic cell comprises a target locus of a nucleic acid and is capable of expressing a recombinase that mediates the BHR at the target locus. Such targeting vectors can include any of the large targeting vectors described herein.

[00306] In one embodiment, the method comprises introducing into a prokaryotic cell: (i) a first construct comprising a nucleic acid having a DNA sequence of interest; (ii) a second targeting construct comprising an insert nucleic acid flanked with a 5' homology arm and a 3' homology arm, and (iii) a third construct encoding a recombinase that mediates bacterial homologous recombination. In one embodiment, the first, the second, and the third construct are introduced into the prokaryotic cell separately over a period of time. In one embodiment, the prokaryotic cell comprises a nucleic acid that encodes the recombinase, and the method does not require introduction of the third construct. In one embodiment, the recombinase is expressed under the control of an inducible promoter.

[00307] In one embodiment the first construct comprising the nucleic acid, is derived from a bacterial artificial chromosome (BAC) or yeast artificial chromosome

(YAC). A prokaryotic cell comprising the insert nucleic acid at the target genomic locus can be selected. This method can be serially repeated as disclosed herein to allow the introduction of multiple insert nucleic acids at the targeted locus in the prokaryotic cell. Once the target nucleic acid locus is "built" within the prokaryotic cell, a targeting vector comprising the modified target locus can be isolated from the prokaryotic cell and introduced into a target genomic locus within a eukaryotic cell, non-rat eukaryotic cell, a mammalian cell, a human cell, a non-human mammalian cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally-restricted human progenitor cell, a human iPS cell, a human cell, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell, a fibroblast, or a CHO cell.

[00308] Preferred rat cells for receiving targeting vectors are described in U.S. Application 14/185,703, filed February 20, 2014, the contents of which are summarized herein. These rat cells are pluripotent rat cells capable of sustaining their pluripotency following one or more targeted genetic modifications *in vitro*, and are capable of transmitting the targeted genetic modifications through the germline.

[00309] Electroporated pluripotent cells, for example, are plated at a high density for the selection of drug-resistant cells comprising the targeting vector. The drug selection process removes the majority of the plated cells (~99%), leaving behind individual colonies, each of which is a clone derived from a single cell. Of the remaining cells, most cells (~ 80-100%) contain the targeting vector (comprising a drug selection cassette) integrated at a random location in the genome. Therefore, the colonies are picked individually and genotyped to identify ES cells harboring the targeting vector at the correct genomic location (e.g., using the modification of allele assay described below).

[00310] A high-throughput quantitative assay, namely, modification of allele (MOA) assay, can be used for genotyping. Such an assay 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. In one embodiment, the quantitative assay is carried out via Invader Probes®. In one embodiment, the quantitative assay is carried out via MMP assays®. In one embodiment, the quantitative assay is carried out via TaqMan® Molecular Beacon. In one embodiment, the quantitative assay is carried out via Eclipse™ probe technology. (See, for example, US2005/0144655, which is incorporated by reference herein in its entirety).

[00311] The selected pluripotent cell (i.e., a non-human pluripotent cell, a non-human ES cell) comprising the targeted genetic modification can then be introduced into a host embryo, for example, a pre-morula stage or blastocyst stage embryo and implanted in the uterus of a surrogate mother to generate a founder non-human animal (F0 animal). Subsequently, the founder animal, for example, can be bred to a wild-type animal to create F1 progeny heterozygous for the genetic modification. Mating of the heterozygous F1 animal can produce progeny homozygous for the genetic modification. Mating of the heterozygous F1 animal can produce progeny homozygous for the genetic modification. In some embodiments, various genetic modifications of the target loci described herein can be carried out using a large targeting vector (LTVEC) as described in detail elsewhere herein. For example, an LTVEC can be 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).

[00312] Use of bacterial homologous recombination (BHR) to generate a large targeting vector (LTVEC) circumvents the limitations of plasmids in accommodating a large genomic DNA fragment and consequent low efficiency of introducing a targeted modification into an endogenous locus in pluripotent or non-pluripotent cells. One or more targeted genetic modifications can be performed in generating a LTVEC. An exemplary LTVEC produced in the prokaryotic cell can comprises an insert nucleic acid that carries a genomic sequence with one or more genetic modifications or an exogenous

nucleic acid (e.g., a homolog or ortholog of a rat nucleic acid), which is flanked by homologous arms, complementary to specific genomic regions.

[00313] Host prokaryotic cells comprising the various targeting vectors described herein are also provided. Such prokaryotic cells include, but are not limited to, bacteria such as *E. coli.* In one embodiment, a host prokaryotic cell comprises a targeting vector comprising an insert nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the insert nucleic acid ranges from about 5 kb to about 200 kb.

[00314] The host prokaryotic cell can further comprise a nucleic acid that encodes a recombinase polypeptide or the nucleic acid that encodes the recombinase polypeptide is operably linked to an inducible promoter.

[00315] Further provided are various methods and compositions, which employ the LTVEC as described herein in combination with a prokaryotic cell in order to produce targeted genetic modifications. Such compositions and methods are discussed elsewhere herein.

[00316] Methods for modifying a target locus of a nucleic acid via bacterial homologous recombination (BHR) are provided that comprise introducing into a prokaryotic cell a targeting vector comprising an insert nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the prokaryotic cell comprises nucleic acids corresponding to the 5' and 3' homology arms and the prokaryotic cell is capable of expressing a recombinase that mediates the BHR at the target locus. Such targeting vectors can include any of the large targeting vectors described herein. Such methods can employ a LTVEC as discussed in detail herein and further employ the CRISPR/Cas system as discussed elsewhere herein.

[00317] In one embodiment, the CRISPR/Cas system can be controlled by a promoter active in a prokaryotic cell, such as, for example, *E.coli.*

ii. Methods for Modifying a Target Locus of Interest in a Pluripotent Cell or Non-Pluripotent Cell.

[00318] Further provided is a method for modifying a target locus of interest in a pluripotent cell or non-pluripotent cell via targeted genetic modification, comprising (a) introducing into the pluripotent cell or non-pluripotent cell a targeting vector comprising

an insert nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the sum total of the 5' homology arm and the 3' homology arm is at least 10 kb; and (b) identifying a genetically modified pluripotent or non-pluripotent cell comprising the targeted genetic modification at the target locus of interest. In one embodiment, the sum total of the 5' homology arm and the 3' homology arm is at least about 16 kb to about 30 kb. In specific embodiments, the targeted genetic modification is capable of being transmitted through the germline. Such targeting vectors can include any of the large targeting vectors described herein.

[00319] Various cells can also be used in the methods for modifying a target locus of interest provided herein. In specific embodiments, the cell is a eukaryotic cell, non-rat eukaryotic cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally restricted human progenitor cell, a human induced pluripotent cell (iPS) cell, a mammalian cell, a human cell, a fibroblast, a rodent cell, a non-rat rodent cell, a mouse cell, a hamster cell or a CHO cell.

[00320] In one aspect, a method for modifying a genomic locus of interest in a pluripotent cell via targeted genetic modification is provided, comprising: (a) providing a pluripotent cell that is able to sustain its pluripotency following at least one targeted genetic modification of its genome and is able to transmit the targeted modification to a germline of an F1 generation; (b) introducing a large targeting vector (LTVEC) into the pluripotent cell, wherein the LTVEC comprises an insert nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the 5' homology arm and the 3' homology arm comprise a genomic DNA fragment; and (c) identifying a genetically modified pluripotent cell comprising the targeted genetic modification.

[00321] Various methods can be used to identify cells having the insert nucleic acid integrated at the target locus of interest. Insertion of the insert nucleic acid at the target locus of interest results in a "modification of allele". The term "modification of allele" and methods for the detection of the modified allele are discussed in further detail elsewhere herein.

[00322] In one aspect, a method for modifying a genomic locus of interest in a non-pluripotent cell or a pluripotent cell via endonuclease-mediated gene targeting is

provided, the method comprising: (a) providing an isolated non-pluripotent cell or an isolated pluripotent cell that is able to transmit the genetically modified genome to a germline of an F1 generation; (b) introducing into the non-pluripotent cell or the pluripotent cell an endonuclease agent; wherein the endonuclease agent makes a nick or a double strand break at a target DNA sequence located in the genomic locus of interest, and wherein the nick or the double strand break at the target DNA sequence in the non-pluripotent cell or the pluripotent cell induces: (i) non-homologous end joining (NHEJ)-mediated DNA repair of the nick or the double strand break, wherein the NHEJ-mediated DNA repair generates a mutant allele comprising an insertion or a deletion of a nucleic acid sequence at the target DNA sequence; or (ii) homologous recombination-mediated DNA repair that results in restoration of a wild-type nucleic acid sequence; and (c) identifying the modified genomic locus of interest.

[00323] In one aspect, a method for modifying a genomic locus of interest in an isolated embryonic stem cell (ES) via a nuclease agent is provided, comprising: (a) providing an isolated ES cell that is able to transmit the targeted genetic modification to a germline of an F1 generation; (b) introducing into the ES cell: (i) a large targeting vector (LTVEC) comprising an insert nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the insert is a nucleic acid sequence that is at least 5 kb; and (ii) an endonuclease agent, wherein the endonuclease agent makes a nick or a double strand break at a target DNA sequence located in the genomic locus of interest, and wherein the target sequence is not present in the insert nucleic acid; and (c) identifying the targeted genetic modification in the embryonic stem (ES) cell.

[00324] In one aspect, a method for modifying a genomic locus of interest in a non-pluripotent cell or a pluripotent cell via RNA-guided genome engineering is provided, the method comprising: (a) providing a non-pluripotent cell or a pluripotent cell that is able to transmit the genetically modified genome to a germline of an F1 generation; (b) introducing into the non-pluripotent cell or the pluripotent cell: (i) a first expression construct comprising a first promoter operably linked to a first nucleic acid sequence encoding a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) protein, (ii) a second expression construct comprising a second promoter operably linked to a genomic target sequence linked to a guide RNA

(gRNA), wherein the genomic target sequence is flanked by a Protospacer Adjacent Motif (PAM) sequence. Optionally the genomic target sequence is flanked on the 3'end by a Protospacer Adjacent Motif (PAM) sequence. In one embodiment, the Cas protein and the CRISPR RNA and/or tracrRNA do not naturally occur together (e.g., the Cas protein and CRISPR RNA do not naturally occur together). In one embodiment, the genomic target sequence comprises the nucleotide sequence of GNNNNNNNNNNNNNNNNNNGG ($GN_{1-20}GG$; SEQ ID NO: 1). In one embodiment, the genomic target sequence comprises SEQ ID NO: 1, wherein N is between 14 and 20 nucleotides in length. In one embodiment, the gRNA comprises a third nucleic acid sequence encoding a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNA (crRNA) and a fourth nucleic acid sequence encoding a trans-activating CRISPR RNA (tracrRNA). In one embodiment, upon expression, the Cas protein forms a CRISPR-Cas complex comprising the crRNA and the tracrRNA, and the CRISPR-Cas complex makes a nick or a double strand break at a target DNA sequence located in the genomic locus of interest, and wherein the nick or the double strand break at the target DNA sequence in the non-pluripotent cell or the pluripotent cell induces: (i) non-homologous end joining (NHEJ)-mediated DNA repair of the nick or the double strand break created by the CRISPR-Cas complex, wherein the NHEJ generates a mutant allele comprising an insertion or a deletion of a nucleic acid sequence at the target DNA sequence; or (ii) homologous recombination-mediated DNA repair that results in restoration of a wild-type nucleic acid sequence; and (c) identifying the modified the genomic locus of interest.

[00325] In one aspect, a method for modifying a genomic locus of interest in a non-pluripotent cell or a pluripotent cell via RNA-guided genome engineering is provided, the method comprising introducing into the non-pluripotent cell or the pluripotent cell that is able to transmit the modified genome through the germline: (i) a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) protein or a nucleic acid encoding the Cas protein; and (ii) a gRNA or a DNA encoding the gRNA, wherein the gRNA comprises a nucleotide sequence that hybridizes to a genomic target sequence and a trans-activating CRISPR RNA (tracrRNA); wherein the genomic target sequence is flanked by a Protospacer Adjacent Motif (PAM) sequence.

[00326] In some embodiments, the Cas protein can be introduced into the non-pluripotent cell or the pluripotent cell as an isolated protein. In some embodiments, the Cas protein can further comprise a cell-penetrating domain that facilitates cellular uptake of the protein. In other embodiments, the Cas protein can be introduced into the cell as a messenger RNA (mRNA) molecule encoding the Cas protein. In other embodiments, the Cas protein can be introduced into the cell as a DNA molecule encoding the Cas protein. For example, the DNA molecule encoding the Cas protein can be provided in a construct and be operably linked to a promoter capable of expressing in the non-pluripotent cell or the pluripotent cell. In certain embodiments, the nucleic acid encoding the Cas protein is codon-optimized for expression in the non-pluripotent cell or the pluripotent cell.

[00327] In some embodiments, the gRNA can be introduced into the non-pluripotent cell or the pluripotent cell as a RNA molecule. For example, the gRNA molecule can be transcribed *in vitro*. In other embodiments, the gRNA can be introduced into the non-pluripotent cell or the pluripotent cell as a DNA molecule encoding the gRNA. For example, the DNA molecule encoding the gRNA can be in a construct and be operably linked to a promoter capable of expressing the gRNA in the non-pluripotent cell or the pluripotent cell. In other embodiments, the gRNA can be chemically synthesized.

[00328] In some embodiments, the gRNA can be introduced into the non-pluripotent cell or the pluripotent cell as a fused crRNA-tracrRNA molecule (i.e., a single transcript). In other embodiments, the gRNA can be introduced into the non-pluripotent cell or the pluripotent cell as separate crRNA and tracrRNA molecules (i.e., separate transcripts). In other embodiments, the gRNA can be introduced into the non-pluripotent cell or the pluripotent cell as separate DNA molecules encoding the crRNA and tracrRNA, respectively. For example, the separate DNA molecules encoding the crRNA and tracrRNA can be in separate constructs and be operably linked to promoters capable of expressing in the non-pluripotent cell or the pluripotent cell. In any of the above embodiments, any combination of the constructs can be in separate nucleic acid molecules or together in a single nucleic acid molecule

[00329] In some embodiments, the Cas protein and the gRNA can be introduced into the non-pluripotent cell or the pluripotent cell simultaneously or sequentially. Likewise, the crRNA and the tracrRNA of the gRNA can be introduced into the non-pluripotent cell or the pluripotent cell simultaneously or sequentially. The ratio of the Cas protein (or encoding nucleic acid) to the gRNA (or encoding DNA) and/or the ratio of the crRNA to the tracrRNA can be about stoichiometric such that they can form an RNA-protein complex.

[00330] In certain embodiments, the Cas protein can be introduced into the non-pluripotent cell or the pluripotent cell in the form of a complex with the gRNA.

[00331] In one embodiment, the pluripotent cell is an induced pluripotent stem cell (iPS). In one embodiment, the pluripotent cell is a developmentally restricted progenitor cell.

[00332] The presence of a nick or a double-strand break in the recognition site within the selection marker, in various embodiments, increases the efficiency and/or frequency of recombination between a targeting vector (such as a LTVEC) and the targeted locus of interest. 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 nick or double strand break, targeting efficiency of a targeting vector (such as a LTVEC) 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 nick or double-strand break (using, e.g., the same targeting vector and the same homology arms and corresponding target sites at the genomic locus of interest but in the absence of an added nuclease agent that makes the nick or double strand break).

[00333] In one embodiment, the targeted genetic modification at the target locus is biallelic. By "biallelic" is meant that both alleles of a gene comprise the targeted genetic modification. The targeted genetic modification can be the same or different in each allele. For example, a biallelic modification can result from the same modification being made to corresponding alleles on corresponding homologous chromosomes, or from different modifications being made to corresponding alleles on corresponding homologous chromosomes. Thus, a biallelic modification can result, for example, in

homozygosity for a specific modification at a genomic locus of interest (i.e., the specific modification in both alleles), compound heterozygosity at a genomic locus of interest (e.g., the specific modification in one allele and inactivation or disruption of the other allele), or hemizyogosity at a genomic locus of interest (e.g., the specific modification in one allele and loss of the other allele). In certain embodiments, the combined use of a targeting vector (including, for example, an LTVEC) with a nuclease agent results in biallelic targeted genetic modification of the genomic locus of interest in a cell as compared to use of the targeting vector alone. When the targeting vector is used in conjunction with a nuclease agent, biallelic targeting efficiency is increased at least by two-fold, at least three-fold, at least 4-fold or more as compared to when the targeting vector is used alone. In further embodiments, the biallelic targeting efficiency is at least 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 2%, 3%, 4% or 5% or higher.

[00334] The biallelic targeted genetic modification at the target locus can result in a homozygous genetically modified cell. By "homozygous" is meant that both alleles of the target locus (i.e., the alleles on both homologous chromosomes) have been modified in the same way. In certain embodiments, the combined use of a targeting vector (including, for example, an LTVEC) with a nuclease agent results in biallelic homozygous targeted genetic modification of the genomic locus of interest in a cell. In one embodiment, the biallelic genetic modification comprises deletion of an endogenous nucleic acid sequence at the genomic locus of interest in two homologous chromosomes (i.e., a pair of first and second homologous chromosomes) and insertion of the insert nucleic acid at the genomic locus of interest in two homologous chromosomes (i.e., the pair of first and second homologous chromosomes). In some embodiments, the insert nucleic acid replaces the endogenous nucleic acid sequence at the genomic locus of interest in both homologous chromosomes. In one embodiment, the insert nucleic acid is homologous or orthologous to the deleted endogenous nucleic acid sequence.

[00335] In one embodiment, the targeted genetic modification at the target locus results in a hemizygous genetically modified cell. By "hemizygous" is meant that only one allele (i.e., the allele on one of two homologous chromosomes) of the target locus is present or only one allele is capable of being expressed and functional. In other embodiments, the targeted genetic modification results more generally in compound

heterozygosity. Compound heterozygosity includes situations in which both alleles of the target locus (i.e., the alleles on both homologous chromosomes) have been modified, but they have been modified in different ways (e.g., an insertion in one allele and inactivation or disruption of the other allele). In certain embodiments, the combined use of a targeting vector (including, for example, an LTVEC) with a nuclease agent results in hemizygous targeted genetic modification of the genomic locus of interest in a cell. In certain embodiments, the combined use of a targeting vector (including, for example, an LTVEC) with a nuclease agent results in targeted genetic modifications that create compound heterozygosity at a genomic locus of interest in a cell. In one embodiment, the targeted genetic modification at the genomic locus of interest in one chromosome comprises deletion of an endogenous nucleic acid sequence and insertion of the insert nucleic acid. In other embodiments, the targeted genetic modification comprises: (1) deletion of an endogenous nucleic acid sequence at the genomic locus of interest in two homologous chromosomes; and (2) insertion of the insert nucleic acid into the genomic locus of interest in a first chromosome and disruption of the genomic locus of interest in a second chromosome. The first chromosome can be the first of the two homologous chromosomes, and the second chromosome can be the second of the two homologous chromosomes. In other embodiments, the targeted modification comprises: (1) deletion of an endogenous nucleic acid sequence at the genomic locus of interest and insertion of the insert nucleic acid into the genomic locus of interest in the first homologous chromosome; and (2) disruption of the genomic locus of interest in the second homologous chromosome. Disruption of the endogenous nucleic acid sequence can result, for example, when a double-strand break at the genomic locus of interest created by the nuclease agent is repaired by non-homologous end joining (NHEJ)-mediated DNA repair, which generates a mutant allele comprising an insertion or a deletion of a nucleic acid sequence at the genomic locus of interest and thereby causes disruption of the genomic locus of interest. Examples of disruption include alteration of a regulatory element (e.g., promoter or enhancer) at the genomic locus of interest, a missense mutation, a truncation mutation, a null mutation, or an insertion or deletion of small number of nucleotides (e.g., causing a frameshift mutation). Another example of

disruption is a nonsense mutation. Disruption can result in inactivation (i.e., loss of function) or loss of the allele.

[00336] Homozygous and hemizygous targeted genetic modifications are advantageous because when genetically modified cells containing these mutations are used to generate genetically modified animals as discussed below, the process for generating genetically modified animals that are non-heterozygous (i.e., homozygous or hemizygous) for the intended targeted genetic modification is more efficient and less-time consuming because fewer breeding steps are required. Targeted genetic modifications resulting in compound heterozygosity or hemizygosity (e.g., an insertion in one allele and inactivation, disruption, or loss of the other allele) can be advantageous for the same reason.

[00337] Various cell types can also be used in any of the various methods described herein above for modifying a genomic locus via a nuclease agent. In specific embodiments, the cell is a eukaryotic cell, non-rat eukaryotic cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally restricted human progenitor cell, a human induced pluripotent cell (iPS) cell, a mammalian cell, a human cell, a fibroblast, a rodent cell, a non-rat rodent cell, a mouse cell, a hamster cell or a CHO cell.

[00338] Compositions are provided which comprise a genetically modified non-human animal, having a targeted genetic modification in the interleukin-2 receptor gamma locus or in the *ApoE* locus. The various methods and compositions provided herein allows for these modified loci to be transmitted through the germline.

[00339] In specific embodiments, a genetically modified non-human animal, or a genetically modified pluripotent or non-pluripotent cell comprises a genomic locus having a targeted genetic modification in the interleukin-2 gamma receptor locus or having a targeted genetic modification in the *ApoE* locus, wherein the interleukin-2 gamma receptor genomic locus or the *ApoE* locus comprise: (i) a deletion of at least a portion of the interleukin-2 gamma receptor locus or at least a portion of the *ApoE* locus; (ii) an insertion of a heterologous nucleic acid sequence into the *ApoE* locus or into the interleukin-2 gamma receptor locus; or (iii) a combination thereof, wherein the genetically modified genomic locus is capable of being transmitted through the germline.

[00340] Methods are further provided that allow for such genetically modified non-human animals, and for such genetically modified pluripotent cells to be made. Such methods include a method for modifying an *ApoE* genomic locus or an interleukin-2 gamma receptor locus in a pluripotent cell via targeted genetic modification. The method comprises (a) introducing into the pluripotent cell a targeting vector comprising an insert nucleic acid flanked with a 5' homology arm, to the *ApoE* locus and a 3' homology arm, to the *ApoE* locus, (b) identifying a genetically modified pluripotent cell comprising the targeted genetic modification at the *ApoE* genomic locus of interest, wherein the targeted genetic modification is capable of being transmitted through germline.

[00341] Additional methods include (a) introducing into the pluripotent cell a targeting vector comprising an insert nucleic acid flanked with a 5' homology arm to the interleukin-2 receptor gamma locus and a 3' homology arm to the interleukin-2 receptor gamma locus, (b) identifying a genetically modified pluripotent cell comprising the targeted genetic modification at the interleukin-2 receptor gamma locus, wherein the targeted genetic modification is capable of being transmitted through germline.

iii. Methods of Integrating Multiple Polynucleotides of Interest at the Targeted Locus

[00342] The various methods and compositions provided herein allow for the targeted integration of multiple polynucleotides of interest with a given target locus. The various methods set forth above can be sequentially repeated to allow for the targeted integration of any number of insert nucleic acids into a given targeted locus. Thus, the various methods provide for the insertion of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more insert nucleic acids into the target locus. In particular embodiments, such sequential tiling methods allow for the reconstruction of large genomic regions from a eukaryotic cell, for example, non-rat eukaryotic cell, a mammalian cell (i.e., a human, a non-human, a rodent, a non-rat rodent, a mouse, a monkey, a rat, a hamster, a domesticated mammal or an agricultural animal) into a targeted locus. In such instances, the transfer and reconstruction of genomic regions that include both coding and non-coding regions allow for the complexity of a given region to be preserved by retaining, at least in part, the coding regions, the non-coding regions and

the copy number variations found within the native genomic region. Thus, the various methods provide, for example, methods to generate "heterologous" or "exogenous" genomic regions within any eukaryotic cell, any non-rat eukaryotic cell, any mammalian cell or animal of interest, particularly within a prokaryotic host cell or within a non-pluripotent cell, a pluripotent cell or an ES cell. In one non-limiting example, a "humanized" genomic region within a non-human animal (i.e., within a rat) is generated. Methods to generate genomic regions within any cell are provided herein. In specific embodiments, the cell is a eukaryotic cell, a non-rat eukaryotic cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally restricted human progenitor cell, a human induced pluripotent cell (iPS) cell, a mammalian cell, a human cell, a fibroblast, a rodent cell, a non-rat rodent cell, a mouse cell, a hamster cell or a CHO cell.

3. A Humanized Genomic Locus

[00343] Provided herein are various methods and compositions comprising a humanized genomic locus. As used herein, by "humanized" genomic locus is meant a region of a non-human genome comprising at least one human nucleic acid sequence. The humanized genomic locus can comprise a region of DNA from any organism that has a human DNA sequence inserted therein. In specific embodiments, the organism is a eukaryote, a non-rat eukaryote, a non-human mammal, a mammal, a human, a rodent, a non-rat rodent, a rat, a mouse or a hamster. For example, a "humanized rat locus" comprises a region of rat DNA that has a human DNA sequence inserted therein.

[00344] The human DNA sequence can be a naturally occurring human DNA sequence or it can be modified from its native form. In specific embodiments, the human DNA shares at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a native human sequence. If a human sequence is not a native human sequence it at least has greater sequence identity to a native human sequence than it does to an orthologous non-human sequence. Moreover, the human DNA sequence can comprise a cDNA, a region of human genomic DNA, a non-coding regulatory region, or any portion of a coding, genomic, or regulatory region of the human DNA. The human DNA sequence inserted into the non-human locus can comprise any of the insert

polynucleotides as described elsewhere herein. In specific embodiments, the human DNA sequence is orthologous to the non-human target locus, while in other instances, the human DNA sequence is homologous to the non-human target locus.

[00345] In one embodiment, the targeted genetic modification is an insertion or a replacement of an endogenous nucleic acid sequence, with a homologous or orthologous human nucleic acid sequence. In one embodiment, the targeted genetic modification comprises an insertion or replacement of an endogenous nucleic acid sequence with a homologous or orthologous human nucleic acid sequence at an endogenous locus that comprises the corresponding non-human nucleic acid sequence.

[00346] Methods for making a humanized locus comprise introducing into the target locus comprising a nucleic acid a human nucleic acid sequence. In one embodiment, a method of making a humanized non-human animal provided. Such a method comprises (a) modifying a genome of a non-human pluripotent cell or non-pluripotent cell with a targeting vector comprising an insert nucleic acid that comprises a human nucleic acid sequence to form a donor cell; (b) introducing the donor cell into a host embryo; and (c) gestating the host embryo in a surrogate mother; wherein the surrogate mother produces a progeny that comprises the human nucleic acid sequence. In specific embodiments, the humanized locus is capable of being transmitted through the germline. In a further embodiment, the targeting vector comprises a large targeting vector (LTVEC) and the insert nucleic acid that comprises a human nucleic acid sequence is at least 5 kb.

[00347] In other methods, the humanized genomic locus is made by modifying a target locus of a nucleic acid via bacterial homologous recombination (BHR). The method comprises introducing into a prokaryotic cell a targeting vector comprising an insert nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the insert nucleic acid comprises a human nucleic acid sequence, and wherein the prokaryotic cell comprises a nucleic acid and is capable of expressing a recombinase that mediates the BHR at the target locus.

[00348] The humanized genomic locus can comprise (a) an insertion of a homologous or orthologous human nucleic acid sequence; (b) a replacement of an endogenous nucleic acid sequence with a homologous or orthologous human nucleic acid

sequence; or (c) a combination thereof. In specific embodiments, the humanized genomic locus is capable of being transmitted through the germline. In still other embodiments, the human orthologous sequence replaces the corresponding sequence found in the non-human locus.

[00349] Any human nucleic acid sequence can be used in the methods and compositions provided herein. Non-limiting examples of human nucleic acid sequences that can be used in the methods and compositions are discussed in detail elsewhere herein.

[00350] The human nucleic acid sequence for insertion into a locus of interest can be any size. In one embodiment, the human nucleic acid sequence can be from about 500 nucleotides to about 200 kb, from about 500 nucleotides to about 5 kb, from about 5 kb to about 200 kb, from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 30 kb, from about 30 kb to about 40 kb, from about 40 kb to about 50 kb, from about 60 kb to about 70 kb, from about 80 kb to about 90 kb, from about 90 kb to about 100 kb, from about 100 kb to about 110 kb, from about 120 kb to about 130 kb, from about 130 kb to about 140 kb, from about 140 kb to about 150 kb, from about 150 kb to about 160 kb, from about 160 kb to about 170 kb, from about 170 kb to about 180 kb, from about 180 kb to about 190 kb, or from about 190 kb to about 200 kb. In a specific embodiment, the human nucleic acid sequence is at least 5 kb.

[00351] In one embodiment, a genomic locus is provided wherein the homologous or orthologous human nucleic acid sequence comprises (a) one or more unrearranged human immunoglobulin heavy chain V_H gene segments, one or more unrearranged human immunoglobulin heavy chain D gene segments, and one or more unrearranged human immunoglobulin heavy chain J_H gene segments, which are operably linked to a mammalian heavy chain constant region nucleic acid sequence; (b) a rearranged human immunoglobulin heavy chain variable region nucleic acid sequence operably linked to a mammalian immunoglobulin heavy chain constant region nucleic acid sequence; (c) one or more unrearranged human immunoglobulin V_κ or V_λ gene segments and one or more unrearranged human immunoglobulin J_κ or J_λ gene segments, which are operably linked to a mammalian, immunoglobulin λ or κ light chain light chain constant region nucleic acid sequence; or (d) a rearranged human immunoglobulin λ or κ light chain variable

region nucleic acid sequence operably linked to a mammalian immunoglobulin λ or κ light chain light chain constant region nucleic acid sequence.

[00352] In another embodiment, a genomic locus is provided wherein (a) the mammalian immunoglobulin heavy chain constant region nucleic acid sequence is a constant region nucleic acid sequence, a human constant region nucleic acid sequence, or a combination thereof; or (b) the mammalian immunoglobulin λ or κ light chain light chain constant region nucleic acid sequence is a rat constant region nucleic acid sequence, a human constant region nucleic acid sequence, or a combination thereof.

[00353] In a specific embodiment, a genomic locus is provided wherein the immunoglobulin heavy chain constant region nucleic acid sequence is selected from or comprises a CH1, a hinge, a CH2, a CH3, and/or a combination thereof.

[00354] In one embodiment, the genomic locus comprises one or more functional human V_H gene segments comprising V_H1-2, V_H1-3, V_H1-8, V_H1-18, V_H1-24, V_H1-45, V_H1-46, V_H1-58, V_H1-69, V_H2-5, V_H2-26, V_H2-70, V_H3-7, V_H3-9, V_H3-11, V_H3-13, V_H3-15, V_H3-16, V_H3-20, V_H3-21, V_H3-23, V_H3-30, V_H3-30-3, V_H 3-30-5, V_H3-33, V_H3-35, V_H3-38, V_H3-43, V_H3-48, V_H3-49, V_H3-53, V_H3-64, V_H3-66, V_H3-72, V_H3-73, V_H3-74, V_H4-4, V_H4-28, V_H4-30-1, V_H4-30-2, V_H4-30-4, V_H4-31, V_H4-34, V_H4-39, V_H4-59, V_H4-61, V_H5-51, V_H6-1, V_H7-4-1, V_H7-81, or a combination thereof.

[00355] In one embodiment, the genomic locus comprises one or more functional human D gene segments comprisingD1-1, D1-7, D1-14, D1-20, D1-26, D2-2, D2-8, D2-15, D2-21, D3-3, D3-9, D3-10, D3-16, D3-22, D4-4, D4-11, D4-17, D4-23, D5-12, D5-5, D5-18, D5-24, D6-6, D6-13, D6-19, D6-25, D7-27, or a combination thereof.

[00356] In one embodiment, the genomic locus comprises one or more functional J_H gene segments comprising J_H1, J_H2, J_H3, J_H4, J_H5, J_H6, and/or a combination thereof. In one embodiment, the insert nucleic acid comprises one or more human Vκ gene segments comprisesVκ4-1, Vκ5-2, Vκ 7-3, Vκ 2-4, Vκ1-5, Vκ1-6, Vκ3-7, Vκ1-8, Vκ1-9, Vκ2-10, Vκ3-11, Vκ1-12, Vκ1-13, Vκ2-14, Vκ3-15, Vκ1-16, Vκ1-17, Vκ2-18, Vκ2-19, Vκ3-20, Vκ6-21, Vκ1-22, Vκ1-23, Vκ2-24, Vκ3-25, Vκ2-26, Vκ1-27, Vκ2-28, Vκ2-29, Vκ2-30, Vκ3-31, Vκ1-32, Vκ1-33, Vκ3-34, Vκ1-35, Vκ2-36, Vκ1-37, Vκ2-38, Vκ1-39, Vκ2-40, or a combination thereof.

[00357] In one embodiment, the genomic locus comprises one or more human Vλ gene segments comprising Vλ3-1, Vλ4-3, Vλ2-8, Vλ3-9, Vλ3-10, Vλ2-11, Vλ3-12, Vλ2-14, Vλ3-16, Vλ2-18, Vλ3-19, Vλ3-21, Vλ3-22, Vλ2-23, Vλ3-25, Vλ3-27, or a combination thereof.

[00358] In one embodiment, the genomic locus comprises one or more human Jκ gene segments comprising Jκ1, Jκ2, Jκ3, Jκ4, Jκ5, or a combination thereof.

[00359] In yet another embodiment, the genomic locus, comprises a humanized genomic locus comprising a human interleukin-2 receptor (IL2R) nucleic acid sequence or a variant or a fragment thereof is provided. In specific embodiments, the IL2R nucleic acid sequence comprises an interleukin-2 receptor alpha, an interleukin-2 receptor beta, or an interleukin-2 receptor gamma nucleic acid sequence or variants or fragments thereof.

[00360] In further embodiments, a genomic locus, comprises a humanized genomic locus comprising of a portion of the human *ApoE* locus, the human interleukin-2 receptor gamma locus, the human *Rag2* locus, the human *Rag1* locus and/or the human *Rag2*/*Rag1* locus replacing the corresponding homologous or orthologous portion of the non-human *ApoE* locus, interleukin-2 receptor gamma locus, *Rag2* locus, *Rag1* locus and/or *Rag2*/*Rag1* locus. In one embodiment, the non-human ecto-domain of IL-2Rg is replaced with the ecto-domain of human IL-2Rg, with the remainder of the molecule being from the non-human.

[00361] In another embodiment, a genetically modified non-human animal, comprising a humanized genomic locus is provided. Such genetically modified non-human animals comprise (a) an insertion of a homologous or orthologous human nucleic acid sequence; (b) a replacement of nucleic acid sequence with a homologous or orthologous human nucleic acid sequence at an endogenous genomic locus; or (c) a combination thereof, wherein the humanized genomic locus is capable of being transmitted through the germline.

[00362] Genetically modified animals, including non-human animals) comprising any of the various humanized genomic loci provided herein and described above are also provided.

4. Polynucleotides of Interest

[00363] Any polynucleotide of interest may be contained in the various insert nucleic acids and thereby integrated at the target locus. The methods disclosed herein, provide for at least 1, 2, 3, 4, 5, 6 or more polynucleotides of interest to be integrated into the targeted genomic locus.

[00364] The polynucleotide of interest within the insert nucleic acid when integrated at the target genomic locus can introduce one or more genetic modifications into the cell. The genetic modification can comprise a deletion of an endogenous nucleic acid sequence and/or the addition of an exogenous or heterologous or orthologous polynucleotide into the target genomic locus. In one embodiment, the genetic modification comprises a replacement of an endogenous nucleic acid sequence with an exogenous polynucleotide of interest at the target genomic locus. Thus, methods provided herein allow for the generation of a genetic modification comprising 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. Such modifications may occur upon integration of the first, second, third, fourth, fifth, six, seventh, or any subsequent insert nucleic acids into the target genomic locus.

[00365] The polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus can comprise a sequence that is native to the cell it is introduced into; the polynucleotide of interest can be heterologous to the cell it is introduced to; the polynucleotide of interest can be exogenous to the cell it is introduced into; the polynucleotide of interest can be orthologous to the cell it is introduced into; or the polynucleotide of interest can be from a different species than the cell it is introduced into. As used herein "native" in reference to a sequence inserted at the target locus is a sequence that is native to the cell having the target locus or native to the cell from which the target locus was derived (i.e., from a rat). As used herein, "heterologous" in reference to a sequence includes a sequence that originates from a foreign species, or, if from the same species, is substantially different or modified from its native form in composition and/or genomic locus by deliberate human intervention. As used herein, "exogenous" in reference to a sequence is a sequence that originates from a foreign species. The

polynucleotide of interest can be from any organism of interest including, but not limited to, non-human, a rodent, a non-rat rodent, a hamster, a mouse, a rat, a human, a monkey, an agricultural mammal or a non-agricultural mammal. The polynucleotide of interest can further comprise a coding region, a non-coding region, a regulatory region, or a genomic DNA. Thus, the 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, and/or any of the subsequent insert nucleic acids can comprise such sequences.

[00366] In one embodiment, the polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus is native to a mouse nucleic acid sequence, a human nucleic acid, a non-human nucleic acid, a eukaryotic nucleic acid, a non-rat eukaryotic nucleic acid, a non-human mammalian nucleic acid, a mammalian nucleic acid, a rodent nucleic acid, a non-rat rodent nucleic acid, a rat nucleic acid, a hamster nucleic acid, a monkey nucleic acid, an agricultural mammal nucleic acid, or a non-agricultural mammal nucleic acid. In still further embodiments, the polynucleotide of interest integrated at the target locus 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, a non-human nucleic acid, a eukaryotic nucleic acid, a non-rat eukaryotic nucleic acid, a non-human mammalian nucleic acid, a mammalian nucleic acid, a rodent nucleic acid, a non-rat rodent nucleic acid, a rat nucleic acid, a hamster nucleic acid, a monkey nucleic acid, an agricultural mammal nucleic acid or a non-agricultural mammal nucleic acid or a combination thereof.

[00367] In one embodiment, the polynucleotide of interest can range from about 500 nucleotides to about 200 kb as described above. The polynucleotide of interest can be from about 500 nucleotides to about 5 kb, from about 5 kb to about 200 kb, from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 30 kb, from about 30 kb to about 40 kb, from about 40 kb to about 50 kb, from about 60 kb to about 70 kb, from about 80 kb to about 90 kb, from about 90 kb to about 100 kb, from about 100 kb to about 110 kb, from about 120 kb to about 130 kb, from about 130 kb to about 140 kb, from about 140 kb to about 150 kb, from about 150 kb to about 160 kb, from about 160 kb to about 170 kb, from about 170 kb to about 180 kb, from about 180 kb to about 190 kb, or from about 190 kb to about 200 kb, from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb

to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb.

[00368] The polynucleotide of interest within the insert nucleic acid and/or inserted at the target genomic locus can encode a polypeptide, can encode an miRNA, or it can comprise any regulatory regions or non-coding regions of interest including, for example, a regulatory sequence, a promoter sequence, an enhancer sequence, a transcriptional repressor-binding sequence, or a deletion of a non-protein-coding sequence, but does not comprise a deletion of a protein-coding sequence. In addition, the polynucleotide of interest within the insert nucleic acid and/or inserted at the target genomic locus can encode 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 polynucleotide of interest within the insert nucleic acid and/or inserted at the target genomic locus encodes a protein expressed in a bone marrow or a bone marrow-derived cell. In one embodiment, the polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus encodes a protein expressed in a spleen cell. In still further embodiments, the polynucleotide of interest within the insert nucleic acid and/or inserted at the target locus encodes a protein expressed in a B cell, encodes a protein expressed in an immature B cell or encodes a protein expressed in a mature B cell.

[00369] The polynucleotide of interest within the insert polynucleotide can comprise a portion of an *ApoE* locus, an *Il2rg* locus, a *Rag1* locus, a *Rag2* locus and/or a *Rag2*/*Rag1* locus. Such portions of these given loci are discussed elsewhere herein, as are the various homologous and orthologous regions from any organism of interest that can be employed.

[00370] In one embodiment, polynucleotide of interest within the insert nucleic acid and/or inserted at the target locus comprises a genomic nucleic acid sequence that encodes an immunoglobulin heavy chain variable region amino acid sequence. The phrase "heavy chain," or "immunoglobulin heavy chain" are described elsewhere herein.

[00371] In one embodiment, the polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus comprises a genomic nucleic acid sequence that encodes a human immunoglobulin heavy chain variable region amino acid sequence.

[00372] In one embodiment, the genomic nucleic acid sequence comprises one or more unrearranged human immunoglobulin heavy chain V_H gene segments, one or more unrearranged human immunoglobulin heavy chain D gene segments, and one or more unrearranged human immunoglobulin heavy chain J_H gene segments, which are operably linked to a mammalian heavy chain constant region nucleic acid sequence. In one embodiment, the genomic nucleic acid sequence comprises a rearranged human immunoglobulin heavy chain variable region nucleic acid sequence operably linked to a mammalian heavy chain constant region nucleic acid sequence. In one embodiment, the genomic nucleic acid sequence comprises one or more unrearranged human immunoglobulin V_κ or V_λ gene segments and one or more unrearranged human immunoglobulin J_κ or J_λ gene segments, which are operably linked to a mammalian immunoglobulin λ or κ light chain light chain constant region nucleic acid sequence. In one embodiment, the genomic nucleic acid sequence comprises a rearranged human immunoglobulin λ or κ light chain variable region nucleic acid sequence operably linked to a mammalian immunoglobulin λ or κ light chain light chain constant region nucleic acid sequence. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a rat constant region nucleic acid sequence, a human constant region nucleic acid sequence, or a combination thereof. In one embodiment, the immunoglobulin λ or κ light chain constant region nucleic acid comprises a rat constant region nucleic acid sequence, a human constant region nucleic acid sequence, or a combination thereof.

[00373] In one embodiment, the immunoglobulin heavy chain constant region nucleic acid sequence is selected from or comprises a CH1, a hinge, a CH2, a CH3, and/or a combination thereof. In one embodiment, the heavy chain constant region nucleic acid sequence comprises a CH1-hinge-CH2-CH3.

[00374] In one embodiment, the polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus comprises a genomic nucleic acid sequence that encodes an immunoglobulin light chain variable region amino acid sequence. The phrase

"light chain" includes an immunoglobulin light chain sequence from any organism, and is described elsewhere herein.

[00375] In one embodiment, the polynucleotide of interest within the insert nucleic acid and/or integrated at the target genomic locus comprises a genomic nucleic acid sequence that encodes a human immunoglobulin light chain variable region amino acid sequence.

[00376] In one embodiment, the genomic nucleic acid sequence comprises one or more unrearranged human immunoglobulin V_κ or V_λ gene segments and one or more unrearranged human immunoglobulin J_κ or J_λ gene segments, which are operably linked to a rodent immunoglobulin λ or κ light chain light chain constant region nucleic acid sequence. In one embodiment, the genomic nucleic acid sequence comprises a rearranged human immunoglobulin λ or κ light chain variable region nucleic acid sequence operably linked to a rodent immunoglobulin λ or κ light chain light chain constant region nucleic acid sequence. In one embodiment, the light chain constant region nucleic acid sequence comprises a rat constant region nucleic acid sequence, a human constant region nucleic acid sequence, or a combination thereof. In one embodiment, the immunoglobulin λ or κ light chain constant region nucleic acid comprises a rat constant region nucleic acid sequence, a human constant region nucleic acid sequence, or a combination thereof.

[00377] The polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus can encode an extracellular protein or a ligand for a receptor. In specific embodiments, the encoded ligand is a cytokine. Cytokines of interest includes a chemokine selected from or comprising CCL, CXCL, CX3CL, and/or XCL. The cytokine can also comprise a tumor necrosis factor (TNF). In still other embodiments, the cytokine is an interleukin (IL). In one embodiment, the interleukin is selected from or comprises 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/or IL-36. In one embodiment, the interleukin is IL-2. In specific embodiments, such polynucleotides of interest within the insert nucleic acid and/or integrated at the target genomic locus are from a human and, in more specific embodiments, can comprise human genomic sequence.

[00378] The polynucleotide of interest within the insert nucleic acid and/or integrated at the target genomic locus can encode Apolipoprotein E (ApoE).

[00379] The polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus can encode a cytoplasmic protein or a membrane protein. In one embodiment, the membrane protein is a receptor, such as, a cytokine receptor, an interleukin receptor, an interleukin 2 receptor-alpha, an interleukin-2 receptor beta, an interleukin-2 receptor gamma or receptor tyrosine kinase. In other instances, the polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus can comprise an orthologous or homologous region of the target locus.

[00380] The polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus can comprise a polynucleotide encoding at least a region of a T cell receptor, including the T cell receptor alpha. In specific methods each of the insert nucleic acids comprise a genomic region of the T cell receptor locus (i.e., the T cell receptor alpha locus) such that upon completion of the serial integration, a portion or the entirety of the genomic T cell receptor locus has been integrated at the target locus. Such insert nucleic acids can comprise at least one or more of a variable segment or a joining segment of a T cell receptor locus (i.e., of the T cell receptor alpha locus). In still further embodiments, the polynucleotide of interest encoding the region of the T cell receptor can be from, for example, a eukaryote, a non-rat eukaryote, a mammal, a non-human mammal, rodent, non-rat rodent, mouse, rat, a human, a monkey, a hamster, an agricultural mammal or a domestic mammal polynucleotide encoding a mutant protein.

[00381] In other embodiments, the polynucleotide of interest integrated at the target locus encodes a nuclear protein. In one embodiment, the nuclear protein is a nuclear receptor. In specific embodiments, such polynucleotides of interest within the insert nucleic acid and/or integrated at the target locus are from a human and, in more specific embodiments, can comprise human genomic sequence.

[00382] The polynucleotide of interest within the insert nucleic acid and/or integrated at the target genomic locus can comprise a genetic modification in a coding sequence. Such genetic modifications include, but are not limited to, a deletion mutation of a coding sequence or the fusion of two coding sequences.

[00383] The polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus can comprise a polynucleotide encoding a mutant protein, including, for example, a human mutant protein. In one embodiment, the mutant protein is characterized by an altered binding characteristic, altered localization, altered expression, and/or altered expression pattern. In one embodiment, the polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus comprises at least one disease allele, including for example, an allele of a neurological disease, an allele of a cardiovascular disease, an allele of a kidney disease, an allele of a muscle disease, an allele of a blood disease, an allele of a cancer-causing gene, or an allele of an immune system disease. In such instances, the disease allele can be a dominant allele or the disease allele is a recessive allele. Moreover, the disease allele can comprises a single nucleotide polymorphism (SNP) allele. The polynucleotide of interest encoding the mutant protein can be from any organism, including, but not limited to, a eukaryote, a non-rat eukaryote, a mammal, a non-human mammal, rodent, non-rat rodent, mouse, rat, a human, a hamster, a monkey, an agricultural mammal or a domestic mammal polynucleotide encoding a mutant protein.

[00384] In one embodiment, the genetic modification produces a mutant form of a protein with an altered binding characteristic, altered localization, altered expression, and/or altered expression pattern.

[00385] In one embodiment, the genetic modification produces a deletion, addition, replacement or a combination thereof of a region of the *ApoE* locus, for example, the rat *ApoE* locus, wherein the genetic modification at the *ApoE* locus results in a decrease in ApoE activity. In one embodiment, an *ApoE* knockout is generated.

[00386] In one embodiment, the genetic modification produces a deletion, addition, replacement or a combination thereof of a region of the *Rag1* locus, for example, the rat *Rag1* locus, wherein the genetic modification at the *Rag1* locus results in a decrease in Rag1 activity. In one embodiment, a *Rag1* knockout is generated. In one embodiment, the genetic modification produces a deletion, addition, replacement or a combination thereof of a region of the *Rag2* locus, for example, the rat *Rag2* locus, wherein the genetic modification at the *Rag2* locus results in a decrease in Rag2 activity. In one embodiment, a *Rag2* knockout is generated. In one embodiment, the genetic

modification produces a deletion, addition, replacement or a combination thereof of a region of the *Rag1*/*Rag2* locus, for example, the rat *Rag1*/*Rag2* locus, wherein the genetic modification at the *Rag1*/*Rag2* locus results in a decrease in Rag1 activity and a decrease in Rag2 activity. In one embodiment, a *Rag1*/*Rag2* knockout is generated.

[00387] In one embodiment, the genetic modification produces a deletion, addition, replacement or a combination thereof of a region of the interleukin-2 receptor gamma locus, for example, the rat interleukin-2 receptor gamma locus, wherein the genetic modification at the interleukin-2 receptor gamma locus results in a decrease in interleukin-2 receptor gamma. In one embodiment, an interleukin-2 receptor gamma knockout is generated.

[00388] As discussed elsewhere herein, further embodiments provided herein comprises one or more of the *ApoE* locus, the interleukin-2 receptor gamma locus, the *Rag2* locus, the *Rag1* locus and/or the *Rag2*/*Rag1* locus, for example, the rat *ApoE* locus, the rat interleukin-2 receptor gamma locus , the *Rag2* locus, the *Rag1* locus and/or the *Rag2*/*Rag1* locus, is modified through the replacement of a portion of the rat *ApoE* locus, the interleukin-2 receptor gamma locus , the *Rag2* locus, the *Rag1* locus and/or *Rag2*/*Rag1* locus with the corresponding orthologous portion of an *ApoE* locus, an interleukin-2 receptor gamma locus, a *Rag2* locus, a *Rag1* locus and/or a *Rag2*/*Rag1* locus from another organism.

[00389] In one embodiment, multiple genetic modifications are generated. In one embodiment, a genetic modification produces a deletion, addition, replacement or a combination thereof of a region of interleukin-2 receptor gamma locus, for example, the rat interleukin-2 receptor gamma locus, wherein the genetic modification at the interleukin-2 receptor gamma locus results in a decrease in interleukin-2 receptor gamma and a second genetic modification produces a deletion, addition, replacement or a combination thereof of a region of the rat *Rag2* locus, wherein the genetic modification at the *Rag2* locus results in a decrease in Rag2 activity. In one embodiment, an interleukin-2 receptor gamma/*Rag2* knockout is generated. Such a rat has a SCID phenotype.

[00390] In one embodiment, the mammalian 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 mammalian 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 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 DNA sequence, a rat genomic DNA sequence, a human genomic DNA sequence, or a combination thereof. In one embodiment, the genomic locus comprises, in any order, rat and human genomic DNA sequences. In one embodiment, the genomic locus comprises, in any order, mouse and human genomic DNA sequences. In one embodiment, the genomic locus comprises, in any order, mouse and rat genomic DNA sequences. In one embodiment, the genomic locus comprises, in any order, rat, mouse, and human genomic DNA sequences.

[00391] In one embodiment, the insert nucleic acid comprises a genetic modification in a coding sequence of a gene. In one embodiment, the genetic modification comprises a deletion mutation in the coding sequence. In one embodiment, the genetic modification comprises a fusion of two endogenous coding sequences.

[00392] In one embodiment, the genetic modification 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 element. In one embodiment, the genetic modification comprises an addition of a promoter. In one embodiment, the genetic modification comprises a replacement of a promoter or regulatory element. In one embodiment, the regulatory element is an enhancer. In one embodiment, the regulatory element is a transcriptional repressor-binding element.

[00393] In one embodiment, the genetic modification comprises placement of a human nucleic acid sequence encoding a mutant human protein. In one embodiment, the genetic modification comprises at least one human disease allele of a human gene. In one embodiment, the human disease is a neurological disease. In one embodiment, the human disease is a cardiovascular disease. In one embodiment, the human disease is a kidney disease. In one embodiment, the human disease is a muscle disease. In one embodiment, the human disease is a blood disease. In one embodiment, the human disease is a cancer.

In one embodiment, the human disease is 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.

[00394] The polynucleotide of interest within the insert nucleic acid and/or integrated at the target locus can also comprise a regulatory sequence, including for example, a promoter sequence, an enhancer sequence, or a transcriptional repressor-binding sequence. In specific embodiments, the polynucleotide of interest within the insert nucleic acid and/or integrated at the target genomic locus comprises a polynucleotide having 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 another 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. Such a polynucleotide of interest can be from any organism, including, but not limited to, a eukaryote, a non-rat eukaryote, a mammal, a non-human mammal, rodent, non-rat rodent, mouse, rat, a human, a monkey, an agricultural mammal or a domestic mammal polynucleotide encoding a mutant protein.

5. Methods of Introducing Sequences and Generation of Transgenic Animals

[00395] As outlined above, methods and compositions are provided herein to allow for the targeted integration of one or more polynucleotides of interest into a target locus. Such systems employ a variety of components and for ease of reference, herein the term "targeted integration system" generically comprises all the components required for an integration event (i.e., in non-limiting examples, the various nuclease agents, recognition sites, insert DNA polynucleotides, targeting vectors, target genomic locus, and/or polynucleotides of interest).

[00396] The methods provided herein comprise introducing into a cell one or more polynucleotides or polypeptide constructs comprising the various components of the targeted genomic integration system. "Introducing" means presenting to the cell the sequence (polypeptide or polynucleotide) in such a manner that the sequence gains access

to the interior of the cell. The methods provided herein do not depend on a particular method for introducing any component of the targeted genomic integration system into the cell, only that the polynucleotide gains access to the interior of a least one cell. Methods for introducing polynucleotides into various cell types are known in the art and include, but are not limited to, stable transfection methods, transient transfection methods, and virus-mediated methods.

[00397] Any cells from any organism can be used in the methods provided herein. In specific embodiments the cells are from a eukaryote, a non-rat eukaryote, a mammal, a non-human mammal, a human, a rodent, a non-rat rodent, a rat, a mouse or a hamster. In specific embodiments, the cells are a eukaryotic cell, a non-rat eukaryotic cell, a pluripotent cell, a non-pluripotent cell, a non-human pluripotent cell, a non-human mammalian cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally restricted human progenitor cell, a human induced pluripotent cell (iPS) cell, a mammalian cell, a human cell, a fibroblast, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell or a CHO cell.

[00398] In some embodiments, the cells employed in the methods and compositions have a DNA construct stably incorporated into their genome. "Stably incorporated" or "stably introduced" means the introduction of a polynucleotide into the cell such that the nucleotide sequence integrates into the genome of the cell and is capable of being inherited by progeny thereof. Any protocol may be used for the stable incorporation of the DNA constructs or the various components of the targeted genomic integration system.

[00399] Transfection protocols as well as protocols for introducing polypeptides or polynucleotide sequences into cells may vary. Non-limiting transfection methods include chemical-based transfection methods include the use of liposomes; nanoparticles; calcium phosphate (Graham et al. (1973). Virology 52 (2): 456–67, Bacchetti et al. (1977) Proc Natl Acad Sci USA 74 (4): 1590–4 and, Kriegler, M (1991). Transfer and Expression: A Laboratory Manual. New York: W. H. Freeman and Company. pp. 96–97); dendrimers; or cationic polymers such as DEAE-dextran or polyethylenimine. Non chemical methods include electroporation; Sono-poration; and optical transfection . Particle-based transfection include the use of a gene gun, magnet assisted transfection

(Bertram, J. (2006) Current Pharmaceutical Biotechnology 7, 277–28). Viral methods can also be used for transfection.

[00400] In one embodiment, the introducing one or more of the polynucleotides into a cell is mediated by electroporation, by intracytoplasmic injection, by a viral infection, by an adenovirus, by lentivirus, by retrovirus, by transfection, by lipid-mediated transfection or is mediated via Nucleofection™.

[00401] In one embodiment, introduction one or more of the polynucleotides into a cell further comprises: introducing an expression construct comprising a nucleic acid sequence of interest 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 a stem cell, for example, an embryonic stem cell.

[00402] In one embodiment, the expression construct is introduced together with the LTVEC. In one embodiment, the expression construct is introduced separately from the LTVEC over a period of time.

[00403] In one embodiment, the introduction of the one or more polynucleotides into the cell can be performed multiple times over a period of time. In one embodiment, the introduction of the one or more polynucleotides into the cell are performed at least two times over a period of time, at least three times over a period of time, at least four times over a period of time, at least five times over a period of time, at least six times over a period of time, at least seven times over a period of time, at least eight times over a period of time, at least nine times over a period of times, at least ten times over a period of time, at least eleven times, at least twelve times over a period of time, at least thirteen times over a period of time, at least fourteen times over a period of time, at least fifteen times over a period of time, at least sixteen times over a period of time, at least seventeen times over a period of time, at least eighteen times over a period of time, at least nineteen times over a period of time, or at least twenty times over a period of time.

[00404] In one embodiment, the nuclease agent is introduced into the cell simultaneously with the targeting vector or the large targeting vector (LTVEC). Alternatively, the nuclease agent is introduced separately from the targeting vector or the LTVEC over a period of time. In one embodiment, the nuclease agent is introduced prior

to the introduction of the targeting vector or the LTVEC, while in other embodiments, the nuclease agent is introduced following introduction of the targeting vector or the LTVEC.

[00405] In one embodiment, screening step comprises a quantitative assay for assessing modification of allele (MOA) of a parental chromosome. In one embodiment, the quantitative assay is carried out via a quantitative PCR. In one embodiment, the quantitative PCR is a real-time PCR (qPCR). In one embodiment, 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 one embodiment, the primer set comprises a fluorescent probe that recognizes the amplified sequence. In one embodiment, the quantitative assay is carried out via fluorescence-mediated in situ hybridization (FISH). In one embodiment, the quantitative assay is carried out via comparative genomic hybridization. In one embodiment, the quantitative assay is carried out via isothermic DNA amplification. In one embodiment, the quantitative assay is carried out via isothermic DNA amplification. In one embodiment, the quantitative assay is carried out via quantitative hybridization to an immobilized probe(s). In one embodiment, the quantitative assay is carried out via Invader Probes®. In one embodiment, the quantitative assay is carried out via MMP assays®. In one embodiment, the quantitative assay is carried out via TaqMan® Molecular Beacon. In one embodiment, the quantitative assay is carried out via Eclipse™ probe technology. (See, for example, US2005/0144655, which is incorporated by reference herein in its entirety).

[00406] Further provided is a method for making a humanized non-human animal, comprising: (a) modifying a genome of a pluripotent cell with a targeting vector comprising an insert nucleic acid that comprises a human nucleic acid sequence to form a donor cell; (b) introducing the donor cell into a host embryo; and (c) gestating the host embryo in a surrogate mother; wherein the surrogate mother produces a progeny that comprises the human nucleic acid sequence. In one embodiment, the donor cell is introduced into a host embryo that is at the blastocyst stage or at a pre-morula stage (i.e., a 4 cell stage or an 8 cell stage). Moreover, step (a) can also be performed with a large targeting vector (LTVEC) and/or a human nucleic acid sequence at least 5 kb in length. In still further embodiments, the genetic modification is capable of being transmitted through the germline.

[00407] Genetically modified non-human animals can be generated employing the various methods disclosed herein. Such methods comprise (1) integrating one or more polynucleotide of interest at the target locus of a pluripotent cell to generate a genetically modified pluripotent cell comprising the insert nucleic acid in the targeted genomic locus employing the methods disclosed herein; (2) selecting the genetically modified pluripotent cell having the one or more polynucleotides of interest at the target genomic locus; (3) introducing the genetically modified pluripotent cell into a host embryo; and (4) implanting the host embryo comprising the genetically modified pluripotent cell into a surrogate mother. A progeny from the genetically modified pluripotent cell is generated. In one embodiment, the donor cell is introduced into a host embryo at the blastocyst stage or at the pre-morula stage (i.e., the 4 cell stage or the 8 cell stage). Progeny that are capable of transmitting the genetic modification though the germline are generated. The pluripotent cell can be an ES cell as discussed elsewhere herein.

[00408] Nuclear transfer techniques can also be used to generate the genetically modified non-human animals. Briefly, methods for nuclear transfer include the steps of: (1) enucleating an oocyte; (2) isolating a donor cell or nucleus to be combined with the enucleated oocyte; (3) inserting the cell or nucleus into the enucleated oocyte to form a reconstituted cell; (4) implanting the reconstituted cell into the womb of an animal to form an embryo; and (5) allowing the embryo to develop. In such methods oocytes are generally retrieved from deceased animals, although they may be isolated also from either oviducts and/or ovaries of live animals. Oocytes can be matured in a variety of medium known to those of ordinary skill in the art prior to enucleation. Enucleation of the oocyte can be performed in a number of manners well known to those of ordinary skill in the art. Insertion of the donor cell or nucleus into the enucleated oocyte to form a reconstituted cell is usually by microinjection of a donor cell under the zona pellucida prior to fusion. Fusion may be induced by application of a DC electrical pulse across the contact/fusion plane (electrofusion), by exposure of the cells to fusion-promoting chemicals, such as polyethylene glycol, or by way of an inactivated virus, such as the Sendai virus. A reconstituted cell is typically activated by electrical and/or non-electrical means before, during, and/or after fusion of the nuclear donor and recipient oocyte. Activation methods include electric pulses, chemically induced shock, penetration by

sperm, increasing levels of divalent cations in the oocyte, and reducing phosphorylation of cellular proteins (as by way of kinase inhibitors) in the oocyte. The activated reconstituted cells, or embryos, are typically cultured in medium well known to those of ordinary skill in the art and then transferred to the womb of an animal. See, for example, US20080092249, WO/1999/005266A2, US20040177390, WO/2008/017234A1, and US Patent No. 7,612,250, each of which is herein incorporated by reference.

[00409] In one aspect, a method for making a genetically modified non-human animal is provided, comprising modifying a genomic locus of interest in a pluripotent cell employing endonuclease-mediated gene targeting to introduce a modification at a genomic locus of interest to form a modified pluripotent cell, maintaining the modified pluripotent cell under conditions sufficient to maintain pluripotency, employing the modified pluripotent cell as a donor cell in a host embryo, and gestating the host embryo comprising the modified pluripotent cell in a surrogate mother, wherein the host embryo is gestated by the surrogate mother and a genetically modified progeny is born.

[00410] In one embodiment, the target sequence 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 located in a promoter regulatory region. In one embodiment, the target sequence is located in an enhancer region.

[00411] In one embodiment, introducing step is performed multiple times over a period of time using a plurality of endonucleases that recognize distinct target sequences. In one embodiment, step is performed at least two times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least three times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least four times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least five times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least six times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least seven times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least eight times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least nine times

over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least ten times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least eleven times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least twelve times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least thirteen times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least fourteen times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least fifteen times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least sixteen times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least seventeen times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least eighteen times over a period of time using a plurality of endonucleases that recognize distinct target sequences, at least nineteen times over a period of time using a plurality of endonucleases that recognize distinct target sequences, or at least twenty times over a period of time using a plurality of endonucleases that recognize distinct target sequences.

[00412] In one embodiment, introducing step is mediated by electroporation, by intracytoplasmic injection, by an adenovirus, by lentivirus, by retrovirus, by transfection, by lipid-mediated transfection or is mediated via Nucleofection™.

[00413] In one embodiment, the method further comprises introducing an exogenous nucleic acid into the genetically modified pluripotent cell. In one embodiment, the exogenous nucleic acid is a transgene. In one embodiment, the exogenous nucleic acid is introduced into an endogenous locus. In one embodiment, the exogenous nucleic acid is introduced ectopically (e.g., at a locus different from its endogenous locus).

[00414] In one aspect, a method for making a genetically modified non-human animal is provided, comprising modifying a genomic locus of interest in a pluripotent cell employing RNA-guided genome engineering to introduce a modification at a genomic locus of interest to form a modified pluripotent cell, maintaining the modified pluripotent cell under conditions sufficient to maintain pluripotency, employing the modified

pluripotent cell as a donor cell in a host embryo and gestating the host embryo comprising the modified pluripotent cell in a surrogate mother, wherein the host embryo is gestated by the surrogate mother and a genetically modified progeny is born.

[00415] In one embodiment, the method has a targeting rate ranging from about 2% to about 80%.

[00416] In one embodiment, the method comprises co-introducing a plurality of the second expression construct comprising distinct genomic target sequences for multiplex editing of distinct genomic loci. In one embodiment, the method comprises introducing a plurality of the second expression construct comprising distinct genomic target sequences for multiplex editing of distinct genomic loci over a period of time.

[00417] In one embodiment, introducing step is performed multiple times over a period of time. In one embodiment, introducing step (b) is performed at least two times over a period of time, at least three times over a period of time, at least four times over a period of time, at least five times over a period of time, at least six times over a period of time, at least seven times over a period of time, at least eight times over a period of time, at least nine times over a period of time, at least ten times over a period of time, at least eleven times over a period of time, at least twelve times over a period of time, at least thirteen times over a period of time, at least fourteen times over a period of time, at least fifteen times over a period of time, at least sixteen times over a period of time, at least seventeen times over a period of time, at least eighteen times over a period of time, at least nineteen times over a period of time, at least twenty times over a period of time.

[00418] In one embodiment, the first expression construct and the second expression construct are expressed from a same plasmid.

[00419] In one embodiment, introducing step is mediated by electroporation, by intracytoplasmic injection, by an adenovirus, by lentivirus, by retrovirus, by transfection, by lipid-mediated transfection or is mediated via Nucleofection™.

[00420] In one embodiment, the method further comprises introducing an exogenous nucleic acid into the pluripotent cell comprising the mutant allele.

[00421] In one embodiment, the exogenous nucleic acid is a transgene. In one embodiment, the exogenous nucleic acid is introduced into an endogenous locus. In one

embodiment, the exogenous nucleic acid is placed ectopically (e.g., at a locus different from its endogenous locus).

[00422] In one embodiment, the method further comprises introducing an exogenous nucleic acid into the genetically modified pluripotent cell. In one embodiment, the exogenous nucleic acid is a transgene. In one embodiment, the exogenous nucleic acid is introduced into an endogenous locus. In one embodiment, the exogenous nucleic acid is introduced ectopically (e.g., at a locus different from its endogenous locus).

[00423] In one aspect, a method for making a humanized non-human animal is provided, comprising modifying a genome of a pluripotent cell with an LTVEC comprising an insert that comprises a human sequence of at least 5 kb, and employing the pluripotent cell as a donor cell, introducing the donor cell into a host embryo, and gestating the host embryo in a surrogate mother, wherein the surrogate mother births a progeny that comprises the humanization.

[00424] Other methods for making a genetically modified non-human animal comprising in its germline one or more genetic modifications as described herein is provided, comprising: (a) modifying a targeted locus contained in a prokaryotic cell employing the various methods described herein; (b) selecting a modified prokaryotic cell comprising the genetic modification at the targeted locus; (c) isolating the genetically modified targeting vector from the genome of the modified prokaryotic cell; (d) introducing the genetically modified targeting vector into a pluripotent cell to generate a genetically modified pluripotent cell comprising the insert nucleic acid at the targeted genomic locus; (e) selecting the genetically modified pluripotent cell; (f) introducing the genetically modified pluripotent cell into a host embryo at a pre-morula stage; and (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. In such methods the targeting vector can comprise a large targeting vector. The pluripotent cell can be an ES cell. In further methods, the isolating step (c) further comprises (c1) linearizing the genetically modified targeting vector (i.e., the genetically modified LTVEC). In still further embodiments, the introducing step (d) further comprises (d1) introducing a nuclease agent as described herein into the pluripotent cell. 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. In one embodiment, selecting steps (b) and/or (e) are carried out via a modification of allele (MOA) assay as described herein.

[00425] Further methods for modifying a target genomic locus of a mammalian cell via bacterial homologous recombination (BHR) in a prokaryotic cell are provided and comprise: (a) providing a prokaryotic cell comprising a target locus comprising a nucleic acid, (b) introducing into the prokaryotic cell a targeting vector comprising an insert nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the insert nucleic acid comprises a mammalian region (including, for example, a DNA insert from a human), and (c) selecting a targeted prokaryotic cell comprising the insert nucleic acid at the target locus , wherein the prokaryotic cell is capable of expressing a recombinase that mediates the BHR. Step (a1) can comprise providing a prokaryotic cell comprising a target locus comprising a nucleic acid comprising a first polynucleotide comprising a first recognition site for a first nuclease agent, and step (b1) can further comprise expressing in the prokaryotic cell a nuclease agent that makes a nick or double-strand break at or near the first recognition site. Steps (a)-(c) can be serially repeated as disclosed herein to allow the introduction of multiple insert nucleic acids at the targeted locus in the prokaryotic cell. Once the targeted genomic locus is "built" with the prokaryotic cell, a targeting vector comprising the modified target locus can be isolated from the prokaryotic cell and introduced into a target genomic locus within a pluripotent cell. Pluripotent cells (i.e., ES cells) comprising the modified genomic locus can then be made into genetically modified non-human animals.

[00426] 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).

[00427] In some embodiments, targeted ES cells comprising various genetic modifications as described herein are used as insert ES cells and introduced into a pre-morula stage embryo from a corresponding organism, 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 embryo comprising the genetically modified ES cells is incubated until the blastocyst stage and then implanted into a surrogate mother to produce an F0. Animals bearing the genetically modified genomic locus can be identified via modification of allele (MOA) assay as described herein. The resulting F0 generation non-human animal derived from the genetically modified ES cells is crossed to a wild-type non-human animal to obtain F1 generation offspring. Following genotyping with specific primers and/or probes, F1 non-human animals that are heterozygous for the genetically modified genomic locus are crossed to each other to produce animals that are homozygous for the genetically modified genomic locus. Alternatively, an F0 female non-human animal and an F0 male non-human animal each having the genetic modification can be crossed to obtain an F1 non-human animal homozygous for the genetic modification.

[00428] In one aspect, a genetically modified rat genome, for example, is provided, comprising a targeted modification of an endogenous nucleic acid sequence with a homologous or orthologous nucleic acid sequence from another organism.

[00429] In one embodiment, the homologous or orthologous nucleic acid sequence is of a length from about 5 kb to about 200 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 5 kb to about 10 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 10 kb to about 20 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 20 kb to about 30 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 30 kb to about 40 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 40 kb to about 50 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 50 kb to about 60 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from

about 60 kb to about 70 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 70 kb to about 80 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 80 kb to about 90 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 90 kb to about 100 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 100 kb to about 110 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 110 kb to about 120 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 120 kb to about 130 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 140 kb to about 150 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 150 kb to about 160 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 160 kb to about 170 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 170 kb to about 180 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 180 kb to about 190 kb. In one embodiment, the homologous or orthologous non-rat nucleic acid sequence ranges from about 190 kb to about 200 kb. Various polynucleotides of interest that can be employed in the insert nucleic acid are described elsewhere herein.

[00430] Further methods for targeted genome modification of a non-human animal are provided. Such methods can comprise (a) modifying a genomic locus of interest in a non-human pluripotent cell according to any of the various methods provided herein for modifying a genomic locus of interest, thereby producing a genetically modified non-human pluripotent cell comprising a targeted genome modification; (b) introducing the modified non-human pluripotent cell of step (a) into a non-human host embryo; and (c) gestating the non-human host embryo comprising the modified pluripotent cell in a surrogate mother, wherein the surrogate mother produces F0 progeny comprising the targeted genome modification, and wherein the targeted genome modification is capable of being transmitted through the germline.

[00431] In some embodiments, the targeted genome modification comprises simultaneous deletion of an endogenous nucleic acid sequence at the genomic locus of

interest and insertion of an exogenous nucleic acid at the genomic locus of interest (i.e., deletion and insertion in a single step). In some embodiments, the targeted genome modification comprises a biallelic genetic modification. The biallelic genetic modification can comprise deletion of an endogenous nucleic acid sequence and insertion of an exogenous nucleic acid at the genomic locus of interest in two homologous chromosomes (i.e., a pair of first and second homologous chromosomes).

[00432] In other embodiments, the targeted genome modification creates a modified pluripotent cell that is compound heterozygous at the genomic locus of interest. In other embodiments, the targeted genome modification creates a modified pluripotent cell that is hemizygous at the genomic locus of interest. In some embodiments, the targeted genetic modification at the genomic locus of interest in one chromosome comprises deletion of an endogenous nucleic acid sequence and insertion of an exogenous nucleic acid. For example, the targeted genetic modification can comprise: (1) deletion of an endogenous nucleic acid sequence at the genomic locus of interest in two homologous chromosomes; and (2) insertion of an exogenous nucleic acid into the genomic locus of interest in a first chromosome and disruption of the genomic locus of interest in a second chromosome. The first chromosome can be the first of the two homologous chromosomes, and the second chromosome can be the second of the two homologous chromosomes.

6. Cells

[00433] The various methods and compositions described herein employ a genomic locus targeting system in a cell. In one embodiment, the cell is a pluripotent cell. In one embodiment, the cell is a non-pluripotent cell. In one embodiment, the pluripotent cell is a non-human pluripotent cell. In one embodiment, the non-human pluripotent cell is a mammalian pluripotent cell. In one embodiment, the pluripotent cell is a human induced pluripotent stem (iPS) cell.

[00434] In other embodiments, the cell is a eukaryotic cell, a non-rat eukaryotic cell, a human pluripotent cell, a human ES cell, a human adult stem cell, a developmentally restricted human progenitor cell, a non-human mammalian cell, a

mammalian cell, a human cell, a fibroblast, a rodent cell, a non-rat rodent cell, a rat cell, a mouse cell, a hamster cell or a CHO cell.

[00435] In one embodiment, a eukaryotic cell is a primary cell. Primary cells include cells or cultures of cells that have been isolated directly from an organism, organ, or tissue. Primary cells include cells that are neither transformed nor immortal. They include any cell obtained from an organism, organ, or tissue which was not previously passed in tissue culture or has been previously passed in tissue culture but is incapable of being indefinitely passed in tissue culture. Such cells can be isolated by conventional techniques and include, for example, hematopoietic cells, endothelial cells, epithelial cells, fibroblasts, mesenchymal cells, keratinocytes, melanocytes, monocytes, mononuclear cells, adipocytes, preadipocytes, neurons, glial cells, hepatocytes, skeletal myoblasts, and smooth muscle cells. In some embodiments, primary cells are derived from connective tissues, muscle tissues, nervous system tissues, or epithelial tissues.

[00436] In another embodiment, a eukaryotic cell is an immortalized cell. Immortalized cells include cells from a multicellular organism that would normally not proliferate indefinitely but, due to mutation or alteration, have evaded normal cellular senescence and instead can keep undergoing division. Such mutations or alterations can occur naturally or be intentionally induced. Examples of immortalized cells include Chinese hamster ovary (CHO) cells, human embryonic kidney cells (e.g., HEK 293 cells), and mouse embryonic fibroblast cells (e.g., 3T3 cells). Numerous types of immortalized cells are well known in the art.

[00437] In some embodiments, immortalized cells are derived from cancer cells. In another embodiment, a primary or immortalized cell is one that is typically used for culturing or for expressing recombinant genes or proteins.

[00438] In other embodiments, the pluripotent cell is able to sustain its pluripotency following at least one targeted genetic modification of its genome and is able to transmit the targeted modification to a germline of an F1 generation.

[00439] In one embodiment, the pluripotent cell is a non-human fertilized egg at the single cell stage. In one embodiment, the non-human fertilized egg is a mammalian fertilized egg. In one embodiment, the mammalian fertilized egg is a rodent fertilized

egg at the single cell stage. In one embodiment, the mammalian fertilized egg is a rat or mouse fertilized egg at the single cell stage.

[00440] The various cells employed in the method and compositions disclosed herein can also comprise prokaryotic cells, such as a bacterial cell, including *E. coli.* In specific embodiments, 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, while in other instances, 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 encoding the recombinase comprises a DNA or an mRNA. In some embodiments, 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.

A. Low Osmolality Medium for Making and Maintaining Human Induced Pluripotent Stem Cells

[00441] A cell culture medium is provided for use in the methods and compositions of the invention. In one embodiment, the medium is suitable for making a population of human iPS cells. In another embodiment, the medium is suitable for maintaining human iPS cells in culture. In some embodiments, the human iPS cells are naïve or naïve-looking.

[00442] The medium provided herein comprises at least a base medium, supplements, a leukemia inhibitory factor (LIF) polypeptide, a glycogen synthase kinase 3 (GSK3) inhibitor, and a MEK inhibitor.

[00443] The present medium is a low osmolality medium. In one example, the osmolality is between about 175-280 mOsm/kg. In further examples, the osmolality of the medium is about 180-270 mOsm/kg, about 200-250 mOsm/kg, about 220-240 mOsm/kg, or about 225-235 mOsm. In a particular embodiment, the osmolality of the medium is about 233 mOsm/kg.

[00444] The base medium provided for the invention is a low osmolality base medium to which supplements are added. The present base medium differs from base media typically used to maintain human iPS cells in culture, which include Dulbecco's

Modified Eagle's Medium (DMEM), in various forms (e.g., Invitrogen DMEM, Cat. No. 1 1971 -025), and a low salt DMEM available commercially as KO-DMEM™ (Invitrogen Cat. No. 10829-018).

[00445] The base medium provided herein is a low osmolality medium but exhibits characteristics that are not limited to low osmolality. For example, the DMEM formulation shown in Table A can be made suitable for the purposes of the invention by altering the sodium chloride and/or sodium bicarbonate concentrations as provided herein, which will result in a different osmolality as compared with the standard DMEM base medium or low-salt DMEM base medium (KO-DMEM) shown in Table A.

[00446] Table A: DMEM base medium formulation.

Component	Mg/L	mM
Glycine	30	0.4
L-Arginine•HCI	84	0.398
L-Cystine•2HCI	63	0.201
L-Glutamine	584	4
L-Histidine•HCI•H2O	42	0.2
L-Isoleucine	105	0.802
L-Leucine	105	0.802
L-Lysine•HCI	146	0.798
L-Methionine	30	0.201
L-Phenylalanine	66	0.4
L-Serine	42	0.4
L-Threonine	95	0.798
L-Tryptophan	16	0.0784
L-Tyrosine disodium salt dihydrate	104	0.398
L-Valine	94	0.803
Choline chloride	4	0.0286
D-Calcium pantothenate	4	8.39×10^{-3}
Folic Acid	4	9.07×10^{-3}
Niacinamide	4	0.0328
Pyridoxine•HCI	4	0.0196
Riboflavin	0.4	1.06×10^{-3}

Thiamine•HCI	4	0.0119
i-Inositol	7.2	0.04
Calcium Chloride ($CaCl_2$) (anhydrous)	200	1.8
Ferric Nitrate ($Fe(NO_3)_3.9H_2O$)	0.1	2.48×10^{-4}
Magnesium Sulfate ($MgSO_4$) (anhyd.)	97.67	0.814
Potassium Chloride (KCI)	400	5.33
D-Glucose (Dextrose)	4500	25
Phenol Red	15	0.0399
NaCL/$NaHCO_3$ Content of DMEM		
Sodium Bicarbonate ($NaHCO_3$)	3700	44.05
Sodium Chloride (NaCl)	6400	110.34
NaCl/$NaHCO_3$ Content of Low salt DMEM (KO-DMEM)		
Sodium Bicarbonate ($NaHCO_3$)	2200	26
Sodium Chloride (NaCl)	5100	87.7
NaCl/$NaHCO_3$ Content of Low osmolality DMEM		
Sodium Bicarbonate ($NaHCO_3$)	2200	26
Sodium Chloride (NaCl)	3000	50

[00447] The present base medium can include a salt of an alkaline metal and a halide, such as sodium chloride (NaCl). Exemplary concentrations of NaCl in the base medium include 50 ± 5 mM or about 3 mg/mL.

[00448] In another embodiment, the base medium exhibits a concentration of a salt of carbonic acid. The salt of carbonic acid can be a sodium salt. In such an example, the sodium salt can be sodium bicarbonate. In a particular embodiment, sodium bicarbonate is present in the base medium at a concentration of about 26 ± 5 mM or about 2.2 mg/mL.

[00449] In yet another embodiment, the base medium is a low osmolality base medium. The osmolality of the base medium can be within a range of about 175-280 mOsm/kg, about 180-250 mOsm/kg, about 190-225 mOsm/kg, or about 195-205 mOsm/kg. An exemplary osmolality of the base medium can be 200, 214, 216, or 218 mOsm/kg. In a particular example, the osmolality of the base medium is 200 mOsm/kg. The osmolality can be determined when cells are cultured in different concentrations of CO_2. In some examples, cells are cultured at 3% CO_2 or 5% CO_2.

[00450] In a preferred embodiment, the base medium comprises NaCl at a concentration of 3.0 mg/mL, sodium bicarbonate at a concentration of about 2.2 mg/mL, and has an osmolality of 200 mOsm/kg.

[00451] Supplements formulated with the base medium of the invention are suitable for making, maintaining, or enriching populations of human iPS cells disclosed herein. Such supplements are indicated as "supplements" or "+ supplements" in this disclosure. The term "supplements" or the phrase "+ supplements," includes one or more additional elements added to the components of the base medium described in Table A. For example, supplements can include, without limitation, F-12® medium (Gibco), N2® supplement (Gibco; 100X solution), NEUROBASAL® medium (Gibco), B-27® supplement (Gibco; 50X solution), L-glutamine, glucose, 2-mercaptoethanol, a Leukemia Inhibitory Factor (LIF) polypeptide, a glycogen synthase kinase 3 inhibitor, a MEK inhibitor, or any combination thereof.

[00452] In a particular embodiment, the LIF polypeptide is a human LIF (hLIF) polypeptide. In some examples, a hLIF polypeptide is used at a concentration of about 1-1000 units/mL, about 20-800 units/mL, about 50-500 units/mL, about 75-250 units/mL, or about 100 units/mL.

[00453] In another particular embodiment, the GSK3 inhibitor comprises CHIR99021. In some examples, CHIR99021 is used at a concentration of about 0.1 to 10 μM, about 1-5 μM, about 2-4 μM, or about 3 μM.

[00454] In another particular embodiment, the MEK inhibitor comprises PD0325901. In some examples, PD0325901 is used at a concentration of about 0.1-5 μM, about 0.2-1 μM, about 0.3-0.7 μM, or about 0.5 μM.

[00455] An exemplary medium comprises a low osmolality base medium described herein at about 24.75% (v/v), F-12 medium at about 24.75% (v/v), N2 supplement at about 0.5% (v/v), NEUROBASAL medium at about 49% (v/v), B-27 supplement at about 1% (v/v), L-glutamine at about 2 mM, 2-mercaptoethanol at about 0.1 mM, hLIF at about 100 units/mL, CHIR99021 at about 3 μM, and PD0325901 at about 0.5 μM.

[00456] In another particular embodiment, the medium may or may not comprise basic fibroblast growth factor (bFGF, also known as FGF2 or FGF-β). Preferably the present medium does not comprise bFGF.

B. Human Induced Pluripotent Stem Cells

[00457] Methods and compositions are provided herein for making a population of human iPS cells. Methods and compositions are further provided for maintaining human iPS cells in culture. Human iPS cells that are produced or maintained in culture are also provided.

[00458] The term "pluripotent cell" or "pluripotent stem cell" includes an undifferentiated cell that possesses the ability to develop into more than one differentiated cell type. Such pluripotent cells can be, for example, a mammalian embryonic stem (ES cell) cell or a mammalian induced pluripotent stem cell (iPS cell). Examples of pluripotent cells include human iPS cells.

[00459] The term "embryonic stem cell" or "ES cell" means an embryo-derived totipotent or pluripotent stem cell, derived from the inner cell mass of a blastocyst, that can be maintained in an *in vitro* culture under suitable conditions. ES cells are capable of differentiating into cells of any of the three vertebrate germ layers, e.g., the endoderm, the ectoderm, or the mesoderm. ES cells are also characterized by their ability propagate indefinitely under suitable *in vitro* culture conditions. See, for example, Thomson et al. (Science (1998) Vol. 282(5391), pp. 1145-1147).

[00460] The term "induced pluripotent stem cell" or "iPS cell" includes a pluripotent stem cell that can be derived directly from a differentiated adult cell. Human iPS cells can be generated by introducing specific sets of reprogramming factors into a non-pluripotent cell which can include, for example, Oct3/4, Sox family transcription factors (e.g., Sox1, Sox2, Sox3, Sox15), Myc family transcription factors (e.g., c-Myc, l-Myc, n-Myc), Krüppel-like family (KLF) transcription factors (e.g., KLF1, KLF2, KLF4, KLF5), and/or related transcription factors, such as NANOG, LIN28, and/or Glis1. Human iPS cells can also be generated, for example, by the use of miRNAs, small molecules that mimic the actions of transcription factors, or lineage specifiers. Human iPS cells are characterized by their ability to differentiate into any cell of the

three vertebrate germ layers, e.g., the endoderm, the ectoderm, or the mesoderm. Human iPS cells are also characterized by their ability propagate indefinitely under suitable *in vitro* culture conditions. See, for example, Takahashi and Yamanaka (Cell (2006) Vol. 126(4), pp. 663-676).

[00461] The terms "naïve" and "primed" identify different pluripotency states of human iPS cells. The term "naïve-looking" identifies a cell expressing a pluripotent state that exhibits one or more characteristics of a naïve pluripotent cell. Naïve-looking human iPS cells can also be referred to as "naïve-like" human iPS cells. In some embodiments, naïve-looking human iPS cells exhibit one or more morphological characteristics of naïve human iPS cells, such as a morphology characterized by compact dome-shaped colonies. In some embodiments, naïve-looking human iPS cells express one or more of the pluripotency markers described herein. In some embodiments, naïve or naïve-looking human iPS cells are naïve human iPS cells. In other embodiments, naïve or naïve-looking human iPS cells are naïve-looking iPS cells.

[00462] Characteristics of naïve and primed iPS cells are described in the art. See, for example, Nichols and Smith (Cell Stem Cell (2009) Vol. 4(6), pp. 487-492). Naïve human iPS cells exhibit a pluripotency state similar to that of ES cells of the inner cell mass of a pre-implantation embryo. Such naïve cells are not primed for lineage specification and commitment. Female naïve iPS cells are characterized by two active X chromosomes. In culture, self-renewal of naïve human iPS cells is dependent on leukemia inhibitory factor (LIF) and other inhibitors. Cultured naïve human iPS cells display a clonal morphology characterized by rounded dome-shaped colonies and a lack of apico-basal polarity. Cultured naïve cells can further display one or more pluripotency makers as described elsewhere herein. Under appropriate conditions, the doubling time of naïve human iPS cells in culture can be between 16 and 24 hours.

[00463] Primed human iPS cells express a pluripotency state similar to that of post-implantation epiblast cells. Such cells are primed for lineage specification and commitment. Female primed iPS cells are characterized by one active X chromosome and one inactive X chromosome. In culture, self-renewal of primed human iPS cells is dependent on fibroblast growth factor (FGF) and activin. Cultured primed human iPS cells display a clonal morphology characterized by an epithelial monolayer and display

apico-basal polarity. Under appropriate conditions, the doubling time of primed human iPS cells in culture can be 24 hours or more.

[00464] In one embodiment, human iPS cells can be derived from non-pluripotent cells transformed to express a pluripotent state. Such transformed cells include, for example, cells that have been transformed to express reprogramming genes that induce pluripotency. A pluripotent state can include, for example, expression of one or more of the pluripotency markers described herein. Such cells (such as human foreskin fibroblasts) can be transformed to express reprogramming genes, or any additional genes of interest, by any means known in the art. See, for example, Takahashi and Yamanaka (Cell (2006) Vol. 126(4), pp. 663-676). For example, they can be introduced into the cells using one or more plasmids, lentviral vectors, or retroviral vectors. In some cases, the vectors integrate into the genome and can be removed after reprogramming is complete. In particular embodiments, the non-pluripotent cells are transformed with reprogramming genes comprising Oct4, Sox2, Klf4, Myc, or any combination thereof. In some examples, the transformed cells comprise primed human iPS cells.

[00465] In some embodiments, the human iPS cells cultured in the low osmolality medium described herein express one or more phenotypes, gene expression profiles, or markers characteristic of a naïve state. In one example, the human iPS cells express one or more pluripotency markers whose expression is indicative of a naïve state. Such pluripotency markers can include alkaline phosphatase, NANOG, 5T4, ABCG2, Activin RIB/ALK-4, Activin RIIB, E-Cadherin, Cbx2, CD9, CD30/TNFRSF8, CD117/c-kit, CDX2, CHD1, Cripto, DNMT3B, DPPA2, DPPA4, DPPA5/ESG1, EpCAM/TROP1, ERR beta/NR3B2, ESGP, F-box protein 15/FBXO15, FGF-4, FGF-5, FoxD3, GBX2, GCNF/NR6A1, GDF-3, Gi24/VISTA/B7-H5, integrin alpha 6/CD49f, integrin alpha 6 beta 1, integrin alpha 6 beta 4, integrin beta 1/CD29, KLF4, KLF5, L1TD1, Lefty, Lefty-1, Lefty-A, LIN-28A, LIN-28B, LIN-41, cMaf, cMyc, Oct-3/4, Oct-4A, Podocalyxin, Rex-1/ZFP42, Smad2, Smad2/3, SOX2, SSEA-1, SSEA-3, SSEA-4, STAT3, Stella/Dppa3, SUZ12, TBX2, TBX3, TBX5, TERT, TEX19, TEX19.1, THAP11, TRA-1-60(R), TROP-2, UTF1, and/or ZIC3. In a specific example, the expressed pluripotency marker is alkaline phosphatase, NANOG, or both.

[00466] In another embodiment, human iPS cells cultured in the low osmolality medium described herein display morphological characteristics indicative of a naïve state. An exemplary morphology is characterized by cells having compact dome-shaped colonies in culture.

[00467] In another embodiment, human iPS cells cultured in the low osmolality medium described herein can be mechanically or enzymatically dissociated into a single-cell suspension, passaged, and/or subcultured. In one example, enzymatic dissociation can be performed using trypsin. When cultured in the present low osmolality medium, human iPS cells can provide greater transformation efficiency due to enhanced dissociation into a single-cell suspension. With other types of medium (e.g., mTeSR™ medium or 2i medium) typically used to maintain human iPS cells in culture, dissociation of human iPS cells must be performed mechanically or with enzymes such as collagenase that are less harsh than trypsin. Consequently, the cells are not dissociated as effectively or as completely. In contrast, with the present low osmolality medium, trypsin can be used to dissociate the cells, and the enhanced dissociation results in increased transformation efficiency. Furthermore, unlike with other types of medium typically used to maintain human iPS cells in culture (e.g., mTeSR™ medium or 2i medium), enzymatic dissociation of human iPS cells cultured with the present low osmolality medium (preferably a low osmolality medium not comprising bFGF) can be performed in the absence of one or more inhibitors that are generally necessary for the passage of such cells. An exemplary inhibitor that can be omitted is a Rho-associated protein kinase (ROCK) inhibitor. A ROCK inhibitor is generally necessary when passaging human iPS cells to inhibit the activation of pro-apoptotic pathways.

[00468] In a further embodiment, subcultured human iPS cells cultured in the low osmolality medium described herein can maintain a naïve or naïve-looking state following enzymatic dissociation and subculture. In some examples, subcultured human iPS cells can continue to display a morphology characterized by compact dome-shaped colonies. Subcultured human iPS cells can also continue to express one or pluripotency markers as described herein.

C. Methods for Making and Maintaining a Population of Human Induced Pluripotent Stem Cells

[00469] Methods and compositions are provided for making human iPS cells in an *in vitro* culture. Methods and compositions are further provided for maintaining human iPS cells in an *in vitro* culture.

[00470] The term "making" includes culturing non-pluripotent cells transformed to express one or more reprogramming factors as described herein, under suitable conditions to induce a change in cell phenotype, gene expression, or both, such that the cells display a naïve or naïve-looking state, i.e., express one or more characteristics of naïve human iPS cells. A naïve or naïve-looking state can be expressed in response to particular culture conditions, e.g., culture in a low osmolality medium as described herein. In some examples, the proportion of cells expressing a naïve or naïve-looking state is at least about 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, and up to 100% of the cells in culture.

[00471] In one embodiment, the method enriches an *in vitro* culture for a population of naïve or naïve-looking human iPS cells. In such an embodiment, naïve or naïve-looking human iPS cells can be propagated in culture preferentially over cells that do not express a naïve or naïve-looking state. In another embodiment, naïve or naïve-looking human iPS cells can be selected from a culture, be enzymatically dissociated, and subcultured to produce an enriched population of naïve or naïve-looking human iPS cells.

[00472] In one embodiment, non-pluripotent cells transformed to express a pluripotent state, are cultured *in vitro* in a medium provided herein that is suitable for inducing expression of a naïve or naïve-looking state for a period of at least 1, 2, 5, 7, 10, 14, 21, or 28 days, or any period of time sufficient to induce expression of a naïve or naïve-looking state in culture. Transformed cells can be cultured in the present medium for at least 1, 2, 3, or 4 weeks. Sometimes transformed cells are cultured for 1-4 weeks. Expression of a naïve or naïve-looking state can be determined by observing morphological characteristics or the expression of pluripotency markers, characteristic of a naïve or naïve-looking state, that are described elsewhere herein.

[00473] In one embodiment, non-pluripotent cells transformed to express a pluripotent state, are cultured in the present low osmolality medium until they express characteristics of a naïve or naïve-looking state. Cells can then be cultured in the present medium to maintain a naïve or naïve-looking state. In another embodiment, non-pluripotent cells transformed to express a pluripotent state, are first cultured in a high osmolality medium prior to culturing in the present low osmolality medium. Such high osmolality medium exhibits an osmolality higher than the present low osmolality medium and can comprise bFGF. Some high osmolality medium comprises one or more of bovine serum albumin, bFGF, transforming growth factor β (TGFβ), lithium chloride, pipecolic acid, and gamma-aminobutyric acid (GABA). Examples of a high osmolality medium include mTeSR™ medium (Stemcell Technologies).

[00474] In some embodiments, non-pluripotent cells transformed to express a pluripotent state, can first be cultured in high osmolality medium comprising bFGF until they begin to express characteristics of a naïve or naïve-looking state, at which time the cells are cultured in the present low osmolality medium. In one example, cells can be cultured in high osmolality medium comprising bFGF for a period of at least 1, 2, 5, 10, 30, 60, or 90 days, a period of 1, 2, 4, 8, or 12 weeks, or a period between 1 day to 3 months. An exemplary time period for culture in a high osmolality medium comprising bFGF is 2 months.

[00475] In other embodiments, non-pluripotent cells transformed to express a pluripotent state, can first be cultured in high osmolality medium comprising bFGF until they begin to display a morphology characterized by three-dimensional cell clumps, at which time cells are cultured in the present low osmolality medium. In such embodiments, cells displaying three-dimensional clumps can be selected, dissociated (e.g., with trypsin), and transferred to a new culture in the low osmolality medium described herein.

[00476] The terms "maintain," "maintaining," and "maintenance" include the preservation of at least one or more of the characteristics or phenotypes of the human iPS cells described herein. Such characteristics can include maintaining pluripotency, cell morphology, gene expression profiles, and/or other functional characteristics of naïve cells. The terms "maintain," "maintaining," and "maintenance" can also

encompass the propagation of cells and/or an increase in the number of naïve cells being cultured. The terms include culture conditions that prevent cells from converting to a primed or non-pluripotent state. The terms further include culture conditions that permit the cells to remain pluripotent and/or naïve, while the cells may or may not continue to divide and increase in number.

[00477] In one embodiment, human iPS cells are cultured *in vitro* in a medium provided herein that is suitable for maintaining such cells in a naïve or naïve-looking state. In a particular example, human iPS cells can be cultured in a suitable medium for a period of 1, 2, 5, 7, 10, 14, 21, or 28 days, or for a period of about 2 weeks, about 3 weeks, about 4 weeks, or more, so long as the cultured cells are maintained in a naïve or naïve-looking state. Cells can be cultured for at least 1, 2, 3 or 4 weeks. Sometimes cells are cultured for 1-4 weeks. Human iPS cells can be maintained, for example, for any period of time sufficient for propagation of the cells in culture, genetic modification of the cells, and/or subculture of the cells.

[00478] In another embodiment, human iPS cells or non-pluripotent cells transformed to express a pluripotent state, can be cultured on a substrate or feeder cell layer suitable for *in vitro* culture. In a particular example, cells are cultured on MATRIGEL™ (BD Biosciences). In another example, cells are cultured on newborn human foreskin fibroblast (NuFF) feeder cells. In another example, cells are cultured on GELTREX™ (Life Technologies).

[00479] In a further embodiment, the doubling time of human iPS cells cultured in the present low osmolality medium is reduced as compared to primed human iPS cells or non-pluripotent cells transformed to express a pluripotent state. In a particular example, the doubling time of the present human iPS cells is between about 16-24 hours.

7. Sequence Identity

[00480] The methods and compositions provided herein employ a variety of different components of the targeted genomic integration system (i.e., nuclease agents, recognition sites, insert nucleic acids, polynucleotides of interest, targeting vectors, selection markers and other components). It is recognized throughout the description that some components of the targeted genomic integration system can have active variants

and fragments. Such components include, for example, nuclease agents (i.e., engineered nuclease agents), nuclease agent recognition sites, polynucleotides of interest, target sites and corresponding homology arms of the targeting vector. Biological activity for each of these components is described elsewhere herein.

[00481] As used herein, "sequence identity" or "identity" in the context of two polynucleotides or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

[00482] As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total

number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

[00483] Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 using the following parameters: % identity and % similarity for a nucleotide sequence using GAP Weight of 50 and Length Weight of 3, and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using GAP Weight of 8 and Length Weight of 2, and the BLOSUM62 scoring matrix; or any equivalent program thereof. "Equivalent program" means any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10.

[00484] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein also can be used in the practice or testing of the described invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[00485] 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.

[00486] 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.

[00487] 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.

[00488] Non-limiting embodiments include:

[00489] 1. A method for targeted modification of a genomic locus of interest in a pluripotent rat cell, comprising (a) introducing into the pluripotent rat cell a large targeting vector (LTVEC) comprising an insert nucleic acid flanked with a 5' rat homology arm and a 3' rat homology arm, wherein the sum total of the 5' and the 3' homology arms is at least 10 kb but less than 150 kb; and (b) identifying a genetically modified pluripotent rat cell comprising the targeted genetic modification at the genomic locus of interest, wherein the targeted genetic modification is capable of being transmitted through the germline.

[00490] 2. The method of embodiment 1, wherein the targeted genetic modification is biallelic.

[00491] 3. The method of embodiment 1 or 2, wherein the pluripotent rat cell is a rat embryonic stem (ES) cell.

[00492] 4. The method of embodiment 1, 2 or 3, wherein the pluripotent rat cell is derived from a DA strain or an ACI strain.

[00493] 5. The method of any one of embodiments 1-4, wherein the pluripotent rat cell is characterized by expression of at least one pluripotency marker comprising Dnmt3L, Eras, Err-beta, Fbxo15, Fgf4, Gdf3, Klf4, Lefl, LIF receptor, Lin28, Nanog, Oct4, Sox15, Sox2, Utfl, or a combination thereof.

[00494] 6. The method of any one of embodiments 1-4 wherein the pluripotent rat cell is characterized by one of more of the following characteristics:

(a) lack of expression of one or more pluripotency markers comprising c-Myc, Ecatl, and/or Rexol; (b) lack of expression of mesodermal markers comprising Brachyury and/or Bmpr2; (c) lack of expression of one or more endodermal markers comprising Gata6, Sox17 and/or Sox7; or (d) lack of expression of one or more neural markers comprising Nestin and/or Pax6.

[00495] 7. The method of any one of embodiments 1-6, wherein the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 30 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 120 kb, or from about 120 kb to 150 kb.

[00496] 8. The method of any one of embodiments 1-6, wherein the sum total of the 5' and the 3' homology arms of the LTVEC is from about 16 kb to about 150 kb.

[00497] 9. The method of any one of embodiments 1-8, wherein the targeted genetic modification comprises: (a) a replacement of an endogenous rat nucleic acid sequence with a homologous or an orthologous nucleic acid sequence; (b) a deletion of an endogenous rat nucleic acid sequence; (c) a deletion of an endogenous rat nucleic acid sequence, wherein the deletion ranges from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb; (d) an exogenous nucleic acid sequence ranging from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb; (e) an exogenous nucleic acid sequence comprising a homologous or an orthologous nucleic acid sequence; (f) a chimeric nucleic acid sequence comprising a human and a rat nucleic acid sequence; (g) a conditional allele flanked with site-specific recombinase target sequences; or (h) a reporter gene operably linked to a promoter active in a rat cell.

[00498] 10. The method of any one of embodiments 1-9, wherein the genomic locus of interest comprises (i) a first nucleic acid sequence that is complementary to the 5' rat homology arm; and (ii) a second nucleic acid sequence that is complementary to the 3' rat homology arm.

[00499] 11. The method of embodiment 10, wherein the first and the second nucleic acid sequence is separated by at least 5 kb but less than 3 Mb.

[00500] 12. The method of embodiment 10, wherein the first and the second nucleic acid sequence is separated by at least 5 kb but less than 10 kb, at least 10 kb but

less than 20 kb, at least 20 kb but less than 40 kb, at least 40 kb but less than 60 kb, at least 60 kb but less than 80 kb, at least about 80 kb but less than 100 kb, at least 100 kb but less than 150 kb, or at least 150 kb but less than 200 kb, at least about 200 kb but less than about 300 kb, at least about 300 kb but less than about 400 kb, at least about 400 kb but less than about 500 kb, at least about 500 kb but less than about 1 Mb, at least about 1 Mb but less than about 1.5 Mb, at least about 1.5 Mb but less than about 2 Mb, at least about 2 Mb but less than about 2.5 Mb, or at least about 2.5 Mb but less than about 3 Mb.

[00501] 13. The method of any one of embodiment 1-12, wherein introducing step (a) further comprises introducing a second nucleic acid encoding a nuclease agent that promotes a homologous recombination between the targeting construct and the genomic locus of interest in the pluripotent rat cell.

[00502] 14. The method of embodiment 13, wherein the nuclease agent comprises (a) a chimeric protein comprising a zinc finger-based DNA binding domain fused to a FokI endonuclease; or (b) a chimeric protein comprising a Transcription Activator-Like Effector Nuclease (TALEN) fused to a FokI endonuclease.

[00503] 15. The method of any one of embodiments 1-12, wherein introducing step (a) further comprises introducing into the pluripotent rat cell: (i) a first expression construct comprising a first promoter operably linked to a first nucleic acid sequence encoding a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) protein, (ii) a second expression construct comprising a second promoter operably linked to a genomic target sequence linked to a guide RNA (gRNA), wherein the genomic target sequence is immediately flanked on the 3' end by a Protospacer Adjacent Motif (PAM) sequence.

[00504] 16. The method of embodiment 15, wherein the genomic locus of interest comprises the nucleotide sequence of SEQ ID NO: 1.

[00505] 17. The method of embodiment 15 or 16, wherein the gRNA comprises a third nucleic acid sequence encoding a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA).

[00506] 18. The method of embodiment 15, 16 or 17, wherein the Cas protein is Cas9.

[00507] 19. The method of embodiment 15, 16, 17, or 18, wherein the gRNA comprises: (a) the chimeric RNA of the nucleic acid sequence of SEQ ID NO: 2; or (b) the chimeric RNA of the nucleic acid sequence of SEQ ID NO: 3.

[00508] 20. The method of embodiment 17, wherein the crRNA comprises SEQ ID NO: 4; SEQ ID NO: 5; or SEQ ID NO: 6.

[00509] 21. The method of embodiment 17, wherein the tracrRNA comprises SEQ ID NO: 7 or SEQ ID NO: 8.

[00510] 22. A modified rat genomic locus comprising: (i) an insertion of a homologous or orthologous human nucleic acid sequence; (ii) a replacement of an endogenous rat nucleic acid sequence with the homologous or orthologous human nucleic acid sequence; or (iii) a combination thereof, wherein the modified rat genomic locus is capable of being transmitted through the germline.

[00511] 23. The modified rat genomic locus of embodiment 22, wherein the size of the insertion or replacement is from about 5 kb to about 400 kb.

[00512] 24. The rat genomic locus of embodiment 22, wherein the size of the insertion or replacement is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb.

[00513] 25. A method for making a humanized rat, comprising: (a) targeting a genomic locus of interest in a pluripotent rat cell with a targeting construct comprising a human nucleic acid to form a genetically modified pluripotent rat cell; (b) introducing the genetically modified pluripotent rat cell into a host rat embryo; and (c) gestating the host rat embryo in a surrogate mother; wherein the surrogate mother produces rat progeny comprising a modified genomic locus that comprises: (i) an insertion of a human nucleic acid sequence; (ii) a replacement of the rat nucleic acid sequence at the genomic locus of interest with a homologous or orthologous human nucleic acid sequence; (iii) a chimeric nucleic acid sequence comprising a human and a rat nucleic acid sequence; or

(iv) a combination thereof, wherein the modified genomic locus is capable of being transmitted through the germline.

[00514] 26. The method of embodiment 25, wherein the targeting construct is a large targeting vector (LTVEC), and the sum total of the 5' and the 3' homology arms of the LTVEC is at least 10 kb but less than 150 kb.

[00515] 27. The method of embodiment 26, wherein the sum total of the 5' and the 3' homology arms of the targeting construct is from about 10 kb to about 30 kb, from about 20 kb to 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, or from about 80 kb to about 100 kb, from about 100 kb to about 120 kb, or from about 120 kb to 150 kb.

[00516] 28. The method of embodiment 25, 26 or 27, wherein the human nucleic acid sequence is at least 5 kb but less than 400 kb.

[00517] 29. The method of embodiment 25, 26, or 27, wherein the human nucleic acid sequence is at least 5 kb but less than 10 kb, at least 10 kb but less than 20 kb, at least 20 kb but less than 40 kb, at least 40 kb but less than 60 kb, at least 60 kb but less than 80 kb, at least about 80 kb but less than 100 kb, at least 100 kb but less than 150 kb, at least 150 kb but less than 200 kb, at least 200 kb but less than 250 kb, at least 250 kb but less than 300 kb, at least 300 kb but less than 350 kb, or at least 350 kb but less than 400 kb.

[00518] 30. The method of any one of embodiments 25-29, wherein the pluripotent rat cell is a rat embryonic stem (ES) cell.

[00519] 31. The method of any one of embodiments 25-30, wherein the pluripotent rat cell is derived from a DA strain or an ACI strain.

[00520] 32. The method of any one of embodiments 25-31, wherein the pluripotent rat cell is characterized by expression of at least one pluripotency marker comprising Dnmt3L, Eras, Err-beta, Fbxo15, Fgf4, Gdf3, Klf4, Lefl, LIF receptor, Lin28, Nanog, Oct4, Sox15, Sox2, Utfl, or a combination thereof.

[00521] 33. The method of any one of embodiment 25-31, wherein the pluripotent rat cell is characterized by one or more of the following features: (a) lack of expression of one or more pluripotency markers comprising c-Myc, Ecatl, and/or Rexol; (b) lack of expression of one or more mesodermal markers comprising Brachyury and/or Bmpr2; (c)

lack of expression of one or more endodermal markers comprising Gata6, Sox17, and/or Sox7; or (d) lack of expression of one or more neural markers comprising Nestin and/or Pax6.

[00522] 34. A modified rat comprising a humanized genomic locus, wherein the humanized genomic locus comprises: (i) an insertion of a homologous or orthologous human nucleic acid sequence; (ii) a replacement of a rat nucleic acid sequence at an endogenous genomic locus with a homologous or orthologous human nucleic acid sequence; (iii) a chimeric nucleic acid sequence comprising a human and a rat nucleic acid sequence or (iv) a combination thereof, wherein the humanized genomic locus is capable of being transmitted through the germline.

[00523] 35. A rat or rat cell comprising a targeted genetic modification in its genomic locus, wherein the genomic locus is an Interleukin-2 receptor gamma locus, an *ApoE* locus, a *Rag1* locus, a *Rag2* locus, or a *Rag2/Rag1* locus, wherein the targeted genetic modification comprises: (a) a deletion of an endogenous rat nucleic acid sequence at the genomic locus; (b) an insertion of a homologous nucleic acid, an orthologous nucleic acid, or a chimeric nucleic acid comprising a human and a rat nucleic acid sequence, or (c) a combination thereof, wherein the targeted genetic modification is transmissible through the germline of the rat or a rat propagated from the rat cell.

[00524] 36. The rat or rat cell of embodiment 35, wherein (a) the deletion of the endogenous rat nucleic acid at the genomic locus is at least about 10 kb; or (b) the deletion of the endogenous rat nucleic acid at the genomic locus is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb; (c) the insertion of the exogenous nucleic acid sequence at the genomic locus is at least about 5 kb; or (d) the insertion of the exogenous nucleic acid sequence at the genomic locus is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to

about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb.

[00525] 37. The rat or rat cell of embodiment 35 or 36, wherein (a) the targeted genetic modification at the Interleukin-2 receptor gamma locus results in a decrease in or absence of Interleukin-2 receptor gamma protein activity; (b) the targeted genetic modification at the *ApoE* locus results in a decrease in or absence of ApoE protein activity; (c) the targeted genetic modification at the *Rag1* locus results in a decrease in or absence of Rag1 protein activity; (d) the targeted genetic modification at the *Rag2* locus results in a decrease in or absence of Rag2 protein activity; or (e) the targeted genetic modification at the *Rag2*/*Rag1* locus results in a decrease in or absence of Rag2 protein activity and Rag1 activity.

[00526] 38. The rat or rat cell of embodiment 35, 36, or 37, wherein the targeted genetic modification of the Interleukin-2 receptor gamma locus comprises: (a) a deletion of the entire rat Interleukin-2 receptor gamma coding region or a portion thereof; (b) a replacement of the entire rat Interleukin-2 receptor gamma coding region or a portion thereof with a human Interleukin-2 receptor gamma coding region or a portion thereof; (c) a replacement of an ecto-domain of the rat Interleukin-2 receptor gamma coding region with the ecto-domain of a human Interleukin-2 receptor gamma; or (d) at least a 3 kb deletion of the Interleukin-2 receptor gamma locus.

[00527] 39. The rat or rat cell of any one of embodiments 35-37, wherein the targeted genetic modification of the *ApoE* locus comprises: (a) a deletion of the entire *ApoE* coding region or a portion thereof; or (b) at least a 1.8 kb deletion of the *ApoE* locus comprising the *ApoE* coding region.

[00528] 40. The rat or rat cell of any one of embodiments 35-37, wherein the targeted genetic modification of the *Rag2* locus comprises: (a) a deletion of the entire *Rag2* coding region or a portion thereof; (b) at least a 5.7 kb deletion of the *Rag2* locus comprising the *Rag2* coding region.

[00529] 41. The rat or rat cell of any one of embodiments 35-37, wherein the targeted genetic modification of the *Rag2*/*Rag1* locus comprises: (a) a deletion of the entire *Rag2* coding region or a portion thereof and a deletion of the entire *Rag1* coding

region or portion thereof; or (b) a deletion of at least 16 kb of the *Rag2/Rag1* locus comprising the *Rag2* coding region.

[00530] 42. The rat or rat cell of any one of embodiment 35-41, wherein the targeted genetic modification comprises an insertion of an expression cassette comprising a selective marker at the Interleukin-2 receptor gamma locus, the *ApoE* locus, the *Rag1* locus, the *Rag2* locus, or the *Rag2/Rag1* locus.

[00531] 43. The rat or rat cell of any one of embodiments 42, wherein the expression cassette comprises a lacZ gene operably linked to the endogenous promoter at the genomic locus and a human ubiquitin promoter operably linked to a selective marker.

[00532] 44. The rat or rat cell of any one of embodiments 35-43, wherein the targeted genetic modification in the Interleukin-2 receptor gamma locus, the *ApoE* locus, the *Rag1* locus, the *Rag2* locus or the *Rag2/Rag1* locus comprises the insertion of a self-deleting selection cassette.

[00533] 45. The rat or rat cell of embodiment 44, wherein the self-deleting selection cassette comprises a selective marker gene operably linked to a promoter active in the rat cell and a recombinase gene operably linked to a male germ cell-specific promoter, wherein the self-deleting cassette is flanked by recombination recognition sites recognized by the recombinase.

[00534] 46. The rat or rat cell of embodiment 45, wherein (a) the male germ cell-specific promoter is a Protamine-1 promoter; or (b) the recombinase gene encodes Cre, and the recombination recognition sites are loxP sites.

[00535] 47. The rat or rat cell of any one of embodiments 35-46, wherein the insertion of the exogenous nucleic acid sequence at the genomic locus comprises a reporter nucleic acid operably linked to an endogenous Interleukin-2 receptor gamma promoter, an endogenous *ApoE* promoter, an endogenous *Rag1* promoter, or an endogenous *Rag2* promoter.

[00536] 48. The rat or rat cell of embodiment 47, wherein the reporter nucleic acid encodes a reporter comprising β-galactosidase, 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, or a combination thereof.

[00537] 49. The rat cell of any one of embodiments 35-48, wherein the rat cell is a pluripotent rat cell or a rat embryonic stem (ES) cell.

[00538] 50. The rat cell of embodiment 49, wherein the pluripotent rat cell or the rat embryonic stem (ES) cell (a) is derived from a DA strain or an ACI strain; (b) is characterized by expression of at least one pluripotency marker comprising Dnmt3L, Eras, Err-beta, Fbxo15, Fgf4, Gdf3, Klf4, Lefl, LIF receptor, Lin28, Nanog, Oct4, Sox15, Sox2, Utfl, or a combination thereof; or (c) is characterized by one or more of the following characteristics: (i) lack of expression of one or more pluripotency markers comprising c-Myc, Ecatl, and/or Rexol; (ii) lack of expression of mesodermal markers comprising Brachyury and/or Bmpr2; (iii) lack of expression of one or more endodermal markers comprising Gata6, Sox17 and/or Sox7; or (iv) lack of expression of one or more neural markers comprising Nestin and/or Pax6.

[00539] 51. A method for modifying a target genomic locus in an Interleukin-2 receptor gamma locus, an *ApoE* locus, a *Rag1* locus, a *Rag2* locus or a *Rag2*/*Rag1* locus in a pluripotent rat cell, the method comprising: (a) introducing into the pluripotent rat cell a targeting vector comprising an insert nucleic acid flanked with 5' and 3' rat homology arms homologous to the target genomic locus, (b) identifying a genetically modified pluripotent rat cell comprising a targeted genetic modification at the target genomic locus, wherein the targeted genetic modification is capable of being transmitted through the germline of a rat propagated from the pluripotent rat cell.

[00540] 52. The method of embodiment 51, wherein the targeting vector is a large targeting vector (LTVEC) wherein the sum total of the 5' and the 3' rat homology arms is at least about 10 kb but less than about 150 kb.

[00541] 53. The method of embodiment 51 or 52, wherein introducing the targeting vector into the pluripotent rat cell leads to: (i) a deletion of an endogenous rat nucleic acid sequence at the target genomic locus; (ii) an insertion of an exogenous nucleic acid sequence at the target genomic locus; or (iii) a combination thereof.

[00542] 54. The method of embodiment 53, wherein (a) the deletion of the endogenous rat nucleic acid at the genomic locus is at least about 10 kb; or (b) the

deletion of the endogenous rat nucleic acid at the genomic locus is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb; (c) the insertion of the exogenous nucleic acid sequence at the genomic locus is at least about 5 kb; or. (d) the insertion of the exogenous nucleic acid sequence at the genomic locus is from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb.

[00543] 55. The method of any one of embodiment 51-54, wherein (a) the targeted genetic modification at the Interleukin-2 receptor gamma locus results in a decrease in or absence of Interleukin-2 receptor gamma protein activity; (b) the targeted genetic modification at the *ApoE* locus results in a decrease in or absence of ApoE protein activity; (c) the targeted genetic modification at the *Rag1* locus results in a decrease in or absence of Rag1 protein activity; (d) the targeted genetic modification at the *Rag2* locus results in a decrease in or absence of Rag2 protein activity; or (e) the targeted genetic modification at the *Rag2*/*Rag1* locus results in a decrease in or absence of Rag2 protein activity and i Rag1 protein activity.

[00544] 56. The method of any one of embodiment 51-54, wherein the targeted genetic modification of the Interleukin-2 receptor gamma locus comprises (a) a deletion of the entire rat Interleukin-2 receptor gamma coding region or a portion thereof; (b) a replacement of the entire rat Interleukin-2 receptor gamma coding region or a portion thereof with a human Interleukin-2 receptor gamma coding region or a portion thereof; (c) a replacement of an ecto-domain of the rat Interleukin-2 receptor gamma coding region with the ecto-domain of a human Interleukin-2 receptor gamma; or (d) at least a 3

kb deletion of the Interleukin-2 receptor gamma locus comprising the Interleukin-2 receptor gamma coding region.

[00545] 57. The method of any one of embodiment 51-55, wherein the targeted genetic modification of the *ApoE* locus comprises: (a) a deletion of the entire *ApoE* coding region or a portion thereof; or (b) at least a 1.8 kb deletion of the *ApoE* locus comprising the *ApoE* coding region.

[00546] 58. The method of any one of embodiment 51-55, wherein the targeted genetic modification of the *Rag2* locus comprises: (a) a deletion of the entire *Rag2* coding region or a portion thereof; or (b) at least a 5.7 kb deletion of the *Rag2* locus comprising the *Rag2* coding region.

[00547] 59. The method of any one of embodiment 51-55, wherein the targeted genetic modification of the *Rag1*/*Rag2* locus comprises: (a) a deletion of the entire *Rag2* coding region or a portion thereof and a deletion of the entire *Rag1* coding region or portion thereof; or (b) a deletion of at least 16 kb of the *Rag2*/*Rag1* locus comprising the *Rag2* and *Rag1* coding regions.

[00548] 60. The method of any one of embodiment 51-59, wherein the insert nucleic acid comprises an expression cassette comprising a polynucleotide encoding a selective marker.

[00549] 61. The method embodiment 60, wherein the expression cassette comprises a lacZ gene operably linked to an endogenous promoter at the genomic locus and a human ubiquitin promoter operably linked to a selective marker gene.

[00550] 62. The method of any one of embodiments 51-60, wherein the insert nucleic acid comprises a self-deleting selection cassette.

[00551] 63. The method of embodiment 62, wherein the self-deleting selection cassette comprises a selective marker operably linked to a promoter active in the rat pluripotent cell and a polynucleotide encoding a recombinase operably linked to a male germ cell-specific promoter, wherein the self-deleting cassette is flanked by recombination recognition sites recognized by the recombinase.

[00552] 64. The method of embodiment 63, wherein (a) the male germ cell-specific promoter is a Protamine-1 promoter; or (b) the recombinase gene encodes Cre and the recombination recognition sites are loxP sites.

[00553] 65. The method of embodiment 53, wherein the insertion of the exogenous nucleic acid sequence at the genomic locus comprises a reporter nucleic acid sequence operably linked to an endogenous Interleukin-2 receptor gamma promoter, an endogenous *ApoE* promoter, an endogenous *Rag1* promoter, or an endogenous *Rag2* promoter.

[00554] 66. The method of embodiment 65, wherein the reporter nucleic acid sequence encodes a reporter comprising β-galactosidase, 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, or a combination thereof.

[00555] 67. The method of any one of embodiment 51-66, wherein the pluripotent rat cell is a rat embryonic stem (ES) cell.

[00556] 68. The method of any one of embodiment 51-67, wherein the pluripotent rat cell (a) is derived from a DA strain or an ACI strain; or (b) is characterized by expression of a pluripotency marker comprising Oct-4, Sox-2, alkaline phosphatase, or a combination thereof; or (c) is characterized by one or more of the following characteristics: (i) lack of expression of one or more pluripotency markers comprising c-Myc, Ecat1, and/or Rexo1; (ii) lack of expression of mesodermal markers comprising Brachyury and/or Bmpr2; (iii) lack of expression of one or more endodermal markers comprising Gata6, Sox17 and/or Sox7; or (iv) lack of expression of one or more neural markers comprising Nestin and/or Pax6.

[00557] 69. The method of any one of embodiment 51-68, further comprising identifying the targeted genetic modification at the target genomic locus, wherein the identification step employs a quantitative assay for assessing a modification of allele (MOA) at the target genomic locus.

[00558] 70. The method of any one of embodiment 51-69, wherein introducing step (a) further comprises introducing a second nucleic acid encoding a nuclease agent that promotes a homologous recombination between the targeting vector and the target genomic locus in the pluripotent rat cell.

[00559] 71. The method of embodiment 70, wherein the nuclease agent comprises a chimeric protein comprising a zinc finger-based DNA binding domain fused to a FokI endonuclease.

[00560] 72. The method of embodiment 71, wherein the method results in bi-allelic modification of the target genomic locus.

[00561] 73. The method of any one of embodiment 51-70, wherein introducing step (a) further comprises introducing into the pluripotent rat cell: (i) a first expression construct comprising a first promoter operably linked to a first nucleic acid sequence encoding a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) protein, (ii) a second expression construct comprising a second promoter operably linked to a genomic target sequence linked to a guide RNA (gRNA), wherein the genomic target sequence is immediately flanked on the 3' end by a Protospacer Adjacent Motif (PAM) sequence.

[00562] 74. The method of embodiment 73, wherein the genomic locus of interest comprises the nucleotide sequence of SEQ ID NO: 1.

[00563] 75. The method of embodiment 73 or 74, wherein the gRNA comprises a third nucleic acid sequence encoding a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA).

[00564] 76. The method of embodiment 73, wherein the Cas protein is Cas9.

[00565] 77. The method of embodiment 73, 74, or 75, wherein the gRNA comprises: (a) the chimeric RNA of the nucleic acid sequence of SEQ ID NO: 2; or (b) the chimeric RNA of the nucleic acid sequence of SEQ ID NO: 3.

[00566] 78. The method of embodiment 75, wherein the crRNA comprises SEQ ID NO: 4; SEQ ID NO: 5; or SEQ ID NO: 6.

[00567] 79. The method of embodiment 75, wherein the tracrRNA comprises SEQ ID NO: 7 or SEQ ID NO: 8.

[00568] 80. The rat or rat cell of any one of embodiments 35-50, wherein the rat or rat cell comprises targeted genetic modifications at the Interleukin-2 receptor gamma locus, the *ApoE* locus, the *Rag1* locus, the *Rag2* locus, and/or the *Rag2*/*Rag1* locus.

[00569] 81. The rat or rat cell of embodiment 80, wherein the rat or rat cell comprises targeted genetic modifications at the Interleukin-2 receptor gamma locus and the *Rag2/Rag1* locus.

[00570] Additional non-limiting embodiments include:

[00571] 1. A method for modifying a genomic locus of interest in a eukaryotic cell, comprising: (a) introducing into the eukaryotic cell: (i) a large targeting vector (LTVEC) comprising a first nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein the LTVEC is at least 10 kb; (ii) a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding a Cas protein, (iii) a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding a guide RNA (gRNA) comprising a nucleotide sequence that hybridizes to a target sequence and a trans-activating CRISPR RNA (tracrRNA), wherein the first and the second promoters are active in the eukaryotic cell; and (b) identifying a modified eukaryotic cell comprising a targeted genetic modification at the genomic locus of interest.

[00572] 2. The method of embodiment 1, wherein the targeted genetic modification is a biallelic genetic modification.

[00573] 3. The method of embodiment 1, wherein the LTVEC is at least 15 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, or at least 90 kb.

[00574] 4. The method of embodiment 1, wherein the LTVEC is at least 100 kb, at least 150 kb, or at least 200 kb.

[00575] 5. The method of embodiment 1, wherein the eukaryotic cell is a mammalian cell.

[00576] 6. The method of embodiment 5, wherein the mammalian cell is a fibroblast.

[00577] 7. The method of embodiment 1, wherein the eukaryotic cell is a pluripotent cell.

[00578] 8. The method of embodiment 7, wherein the pluripotent cell is a human pluripotent cell.

[00579] 9. The method of embodiment 8, wherein the human pluripotent cell is a human embryonic stem (ES) cell or a human adult stem cell.

[00580] 10. The method of embodiment 8, wherein the human pluripotent cell is a developmentally restricted human progenitor cell.

[00581] 11. The method of embodiment 8, wherein the human pluripotent cell is a human induced pluripotent stem (iPS) cell.

[00582] 12. The method of embodiment 1, wherein the Cas protein is Cas9.

[00583] 13. The method of embodiment 1, wherein the target sequence is immediately flanked on the 3' end by a Protospacer Adjacent Motif (PAM) sequence.

[00584] 14. The method of embodiment 1, wherein the sum total of the 5' and the 3' homology arms is from about 10 kb to about 150 kb.

[00585] 15. The method of embodiment 1, wherein the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 120 kb, or from about 120 kb to 150 kb.

[00586] 16. The method of embodiment 1, wherein the targeted genetic modification comprises:(a) a replacement of an endogenous nucleic acid sequence with a homologous or an orthologous nucleic acid sequence; (b) a deletion of an endogenous nucleic acid sequence; (c) a deletion of an endogenous nucleic acid sequence, wherein the deletion ranges from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb; (d) insertion of an exogenous nucleic acid sequence; (e) insertion of an exogenous nucleic acid sequence ranging from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from

about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb; (f) insertion of an exogenous nucleic acid sequence comprising a homologous or an orthologous nucleic acid sequence; (g) insertion of a chimeric nucleic acid sequence comprising a human and a non-human nucleic acid sequence; (h) insertion of a conditional allele flanked with site-specific recombinase target sequences; (i) insertion of a selectable marker or a reporter gene operably linked to a third promoter active in the pluripotent cell; or (j) a combination thereof.

[00587] 17. The method of embodiment 1, wherein the genomic locus of interest comprises (i) a 5' target sequence that is homologous to the 5' homology arm; and (ii) a 3' target sequence that is homologous to the 3' homology arm.

[00588] 18. The method of embodiment 17, wherein the 5' target sequence and the 3' target sequence is separated by at least 5 kb but less than 3 Mb.

[00589] 19. The method of embodiment 17, wherein the 5' target sequence and the 3' target sequence is separated by at least 5 kb but less than 10 kb, at least 10 kb but less than 20 kb, at least 20 kb but less than 40 kb, at least 40 kb but less than 60 kb, at least 60 kb but less than 80 kb, at least about 80 kb but less than 100 kb, at least 100 kb but less than 150 kb, or at least 150 kb but less than 200 kb, at least about 200 kb but less than about 300 kb, at least about 300 kb but less than about 400 kb, at least about 400 kb but less than about 500 kb, at least about 500 kb but less than about 1 Mb, at least about 1 Mb but less than about 1.5 Mb, at least about 1.5 Mb but less than about 2 Mb, at least about 2 Mb but less than about 2.5 Mb, or at least about 2.5 Mb but less than about 3 Mb.

[00590] 20. The method of embodiment 1, wherein the genomic locus of interest comprises the Interleukin-2 receptor gamma locus, the *ApoE* locus, the *Rag1* locus, the *Rag2* locus, or both of the *Rag1* and the *Rag2* loci.

[00591] 21. The method of embodiment 1, wherein the first and the second expression constructs are on a single nucleic acid molecule.

[00592] 22. A method for modifying a genome, comprising exposing the genome to a Cas protein and a CRISPR RNA in the presence of a large targeting vector (LTVEC) comprising a nucleic acid sequence of at least 10 kb, wherein following exposure to the

Cas protein, the CRISPR RNA, and the LTVEC, the genome is modified to contain at least 10 kb nucleic acid sequence.

[00593] 23. The method of embodiment 22, wherein the LTVEC comprises a nucleic acid sequence of at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, or at least 90 kb.

[00594] 24. The method of embodiment 22, wherein the LTVEC comprises a nucleic acid sequence of at least 100 kb, at least 150 kb, or at least 200 kb.

[00595] 25. A method for modifying a genome, comprising contacting the genome with a Cas protein, a CRISPR RNA that hybridizes to a target sequence, and a tracrRNA in the presence of a large targeting vector (LTVEC), wherein the LTVEC is at least 10 kb and comprises a first nucleic acid flanked with a 5' homology arm and a 3' homology arm, wherein following contacting with the Cas protein, CRISPR RNA, and tracrRNA in the presence of the LTVEC, the genome is modified at a genomic locus of interest to contain the first nucleic acid.

[00596] 26. The method of embodiment 25, wherein the genome is in a eukaryotic cell, and the Cas protein, the CRISPR RNA, the tracrRNA, and the LTVEC are introduced into the eukaryotic cell

[00597] 27. The method of embodiment 26, further comprising identifying a modified eukaryotic cell comprising a targeted genetic modification at the genomic locus of interest.

[00598] 28. The method of embodiment 26 or 27, wherein the CRISPR RNA and the tracrRNA are introduced together in the form of a single guide RNA (gRNA).

[00599] 29. The method of embodiment 26 or 27, wherein the CRISPR RNA and the tracrRNA are introduced separately.

[00600] 30. The method of any one of embodiments 26-29, wherein: (a) the Cas protein is introduced into the eukaryotic cell in the form of a protein, a messenger RNA (mRNA) encoding the Cas protein, or a DNA encoding the Cas protein; (b) the CRISPR RNA is introduced into the eukaryotic cell in the form of an RNA or a DNA encoding the CRISPR RNA; and (c) the tracrRNA is introduced into the eukaryotic cell in the form of an RNA or a DNA encoding the tracrRNA.

[00601] 31. The method of embodiment 30, wherein the Cas protein, the CRISPR RNA, and the tracrRNA are introduced into the eukaryotic cell as a protein-RNA complex.

[00602] 32. The method of embodiment 30, wherein: (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; (b) the DNA encoding the CRISPR RNA is in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding the CRISPR RNA; and (c) the DNA encoding the tracrRNA is in the form of a third expression construct comprising a third promoter operably linked to a fourth nucleic acid encoding the tracrRNA, wherein the first, second, and third promoters are active in the eukaryotic cell.

[00603] 33. The method of embodiment 32, wherein the first, second, and/or third expression constructs are on a single nucleic acid molecule.

[00604] 34. The method of embodiment 30, wherein: (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; and (b) the DNA encoding the CRISPR RNA and the DNA encoding the tracrRNA are in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding a gRNA comprising the CRISPR RNA and the tracrRNA; wherein the first and second promoters are active in the eukaryotic cell.

[00605] 35. The method of embodiment 34, wherein the first and the second expression constructs are on a single nucleic acid molecule.

[00606] 36. The method of any one of embodiments 27-35, wherein the targeted genetic modification comprises simultaneous deletion of an endogenous nucleic acid sequence at the genomic locus of interest and insertion of the first nucleic acid at the genomic locus of interest.

[00607] 37. The method of any one of embodiments 27-36, wherein the targeted genetic modification is a biallelic genetic modification.

[00608] 38. The method of embodiment 37, wherein the biallelic genetic modification comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid at the genomic locus of interest in two homologous chromosomes.

[00609] 39. The method of any one of embodiments 27-36, wherein the modified eukaryotic cell is hemizygous at the genomic locus of interest.

[00610] 40. The method of embodiment 39, wherein the targeted genetic modification at the genomic locus of interest in one chromosome comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid.

[00611] 41. The method of embodiment 39, wherein the targeted genetic modification comprises: (1) deletion of an endogenous nucleic acid sequence at the genomic locus of interest in two homologous chromosomes; and (2) insertion of the first nucleic acid into the genomic locus of interest in a first chromosome and disruption of the genomic locus of interest in a second chromosome.

[00612] 42. The method of any one of embodiments 25-41, wherein the LTVEC is at least 15 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, or at least 90 kb.

[00613] 43. The method of any one of embodiments 25-42, wherein the LTVEC is at least 100 kb, at least 150 kb, or at least 200 kb.

[00614] 44. The method of any one of embodiments 25-43, wherein the first nucleic acid is at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb, at least 200 kb, at least 250 kb, or at least 300 kb.

[00615] 45. The method of any one of embodiments 26-44, wherein the eukaryotic cell is a mammalian cell.

[00616] 46. The method of embodiment 45, wherein the mammalian cell is a fibroblast.

[00617] 47. The method of any one of embodiments 26-43, wherein the eukaryotic cell is a pluripotent cell.

[00618] 48. The method of embodiment 47, wherein the pluripotent cell is a non-human pluripotent cell.

[00619] 49. The method of embodiment 48, wherein the non-human pluripotent cell is a rodent pluripotent cell.

[00620] 50. The method of embodiment 49, wherein the rodent pluripotent cell is a mouse or rat embryonic stem (ES) cell.

[00621] 51. The method of embodiment 47, wherein the pluripotent cell is a human pluripotent cell.

[00622] 52. The method of embodiment 51, wherein the human pluripotent cell is a human embryonic stem (ES) cell or a human adult stem cell.

[00623] 53. The method of embodiment 51, wherein the human pluripotent cell is a developmentally restricted human progenitor cell.

[00624] 54. The method of embodiment 51, wherein the human pluripotent cell is a human induced pluripotent stem (iPS) cell.

[00625] 55. The method of any one of embodiments 25-54, wherein the Cas protein is Cas9.

[00626] 56. The method of any one of embodiments 25-55, wherein the target sequence is immediately flanked by a Protospacer Adjacent Motif (PAM) sequence.

[00627] 57. The method of any one of embodiments 25-56, wherein the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 150 kb.

[00628] 58. The method of any one of embodiments 25-57, wherein the sum total of the 5' and the 3' homology arms of the LTVEC is from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 120 kb, or from about 120 kb to 150 kb.

[00629] 59. The method of any one of embodiments 27-58, wherein the targeted genetic modification comprises: (a) a replacement of an endogenous nucleic acid sequence with a homologous or an orthologous nucleic acid sequence; (b) a deletion of an endogenous nucleic acid sequence; (c) a deletion of an endogenous nucleic acid sequence, wherein the deletion ranges from about 5 kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, or from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb; (d) insertion of an exogenous nucleic acid sequence; (e) insertion of an exogenous nucleic acid sequence

ranging from about 5kb to about 10 kb, from about 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb; (f) insertion of an exogenous nucleic acid sequence comprising a homologous or an orthologous nucleic acid sequence; (g) insertion of a chimeric nucleic acid sequence comprising a human and a non-human nucleic acid sequence; (h) insertion of a conditional allele flanked with site-specific recombinase target sequences; (i) insertion of a selectable marker or a reporter gene operably linked to a third promoter active in the pluripotent cell; or (j) a combination thereof.

[00630] 60. The method of any one of embodiments 25-59, wherein the genomic locus of interest comprises (i) a 5' target sequence that is homologous to the 5' homology arm; and (ii) a 3' target sequence that is homologous to the 3' homology arm.

[00631] 61. The method of embodiment 60, wherein the 5' target sequence and the 3' target sequence are separated by at least 5 kb but less than 3 Mb.

[00632] 62. The method of embodiment 60, wherein the 5' target sequence and the 3' target sequence are separated by at least 5 kb but less than 10 kb, at least 10 kb but less than 20 kb, at least 20 kb but less than 40 kb, at least 40 kb but less than 60 kb, at least 60 kb but less than 80 kb, at least about 80 kb but less than 100 kb, at least 100 kb but less than 150 kb, or at least 150 kb but less than 200 kb, at least about 200 kb but less than about 300 kb, at least about 300 kb but less than about 400 kb, at least about 400 kb but less than about 500 kb, at least about 500 kb but less than about 1Mb, at least about 1 Mb but less than about 1.5 Mb, at least about 1.5 Mb but less than about 2 Mb, at least about 2 Mb but less than about 2.5 Mb, or at least about 2.5 Mb but less than about 3 Mb.

[00633] 63. The method of embodiment 60, wherein the 5' target sequence and the 3' target sequence are separated by at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 110 kb, at least 120 kb, at least 130 kb, at least 140 kb, at least 150 kb, at least 160 kb, at least 170 kb, at least 180 kb, at least 190 kb, or at least 200 kb.

[00634] 64. The method of any one of embodiments 25-63, wherein the genomic locus of interest comprises the Interleukin-2 receptor gamma locus, the *ApoE* locus, the *Rag1* locus, the *Rag2* locus, or both of the *Rag1* and the *Rag2* loci.

[00635] 65. The method of any one of embodiments 25-63, wherein the genomic locus of interest comprises the *Adamts5* locus, the *Trpa1* locus, the *Folh1* locus, or the *Erbb4* locus.

[00636] 66. The method of any one of embodiments 25-63, wherein the genomic locus of interest comprises the *Lrp5* locus.

[00637] 67. A method for producing an F0 generation non-human animal that comprises a targeted genetic modification at a genomic locus of interest, the method comprising: (a) contacting the genome in a non-human ES cell with a Cas protein, a CRISPR RNA, and a tracrRNA in the presence of a large targeting vector (LTVEC) to form a modified non-human ES cell, wherein the LTVEC is at least 10 kb and comprises a first nucleic acid flanked with a 5' homology arm and a 3' homology arm; (b) identifying the modified non-human ES cell comprising the targeted genetic modification at the genomic locus of interest; (c) introducing the modified non-human ES cell into a non-human host embryo; and (d) gestating the non-human host embryo in a surrogate mother, wherein the surrogate mother produces the F0 generation non-human animal comprising the targeted genetic modification at the genomic locus of interest.

[00638] 68. The method of embodiment 67, wherein the CRISPR RNA and the tracrRNA are introduced together in the form of a single guide RNA (gRNA).

[00639] 69. The method of embodiment 67, wherein the CRISPR RNA and the tracrRNA are introduced separately.

[00640] 70. The method of any one of embodiments 67-69, wherein: (a) the Cas protein is introduced into the non-human ES cell in the form of a protein, a messenger RNA (mRNA) encoding the Cas protein, or a DNA encoding the Cas protein; (b) the CRISPR RNA is introduced into the non-human ES cell in the form of an RNA or a DNA encoding the CRISPR RNA; and (c) the tracrRNA is introduced into the non-human ES cell in the form of an RNA or a DNA encoding the tracrRNA.

[00641] 71. The method of embodiment 70, wherein the Cas protein, the CRISPR RNA, and the tracrRNA are introduced into the non-human ES cell as a protein-RNA complex.

[00642] 72. The method of embodiment 70, wherein: (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; (b) the DNA encoding the CRISPR RNA is in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding the CRISPR RNA; and (c) the DNA encoding the tracrRNA is in the form of a third expression construct comprising a third promoter operably linked to a fourth nucleic acid encoding the tracrRNA, wherein the first, second, and third promoters are active in the non-human ES cell.

[00643] 73. The method of embodiment 72, wherein the first, second, and third expression constructs are on a single nucleic acid molecule.

[00644] 74. The method of embodiment 70, wherein: (a) the DNA encoding the Cas protein is in the form of a first expression construct comprising a first promoter operably linked to a second nucleic acid encoding the Cas protein; and (b) the DNA encoding the CRISPR RNA and the DNA encoding the tracrRNA are in the form of a second expression construct comprising a second promoter operably linked to a third nucleic acid encoding a gRNA comprising the CRISPR RNA and the tracrRNA; wherein the first and second promoters are active in the non-human ES cell.

[00645] 75. The method of embodiment 74, wherein the first and the second expression constructs are on a single nucleic acid molecule.

[00646] 76. The method of any one of embodiments 67-75, wherein the targeted genetic modification comprises simultaneous deletion of an endogenous nucleic acid sequence at the genomic locus of interest and insertion of the first nucleic acid at the genomic locus of interest.

[00647] 77. The method of any one of embodiments 67-76, wherein the targeted genetic modification is a biallelic genetic modification.

[00648] 78. The method of embodiment 77, wherein the biallelic genetic modification comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid at the genomic locus of interest in two homologous chromosomes.

[00649] 79. The method of any one of embodiments 67-76, wherein the modified non-human ES cell is hemizygous at the genomic locus of interest.

[00650] 80. The method of embodiment 79, wherein the targeted genetic modification at the genomic locus of interest in one chromosome comprises deletion of an endogenous nucleic acid sequence and insertion of the first nucleic acid.

[00651] 81. The method of embodiment 79, wherein the targeted genetic modification comprises: (1) deletion of an endogenous nucleic acid sequence at the genomic locus of interest in two homologous chromosomes; and (2) insertion of the first nucleic acid into the genomic locus of interest in a first chromosome and disruption of the genomic locus of interest in a second chromosome.

[00652] 82. The method of any one of embodiments 67-81, wherein the Cas protein is Cas9.

EXAMPLES

[00653] 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. Rat ES Cell Derivation and Characterization

1.1. Rat ES Cell Characterization

[00654] As shown in Figure 1, rat ESCs grow as compact spherical colonies, which routinely detach and float in the dish (close-up, Figure 8). Rat ESCs express pluripotency markers including Oct-4 (Figure 2A) and Sox2 (Figure 2B), and express high levels of alkaline phosphatase (Figure 3). Karyotype for line DA.2B is 42X,Y (Figure 4). Rat ESCs often become tetraploid; thus, lines were pre-screened by counting

metaphase chromosome spreads; lines with mostly normal counts were then formally karyotyped.

[00655] ACI blastocysts were collected from super-ovulated females obtained commercially. DA blastocysts were cultured from frozen 8-cell embryos obtained commercially. Zona pellucidae were removed with Acid Tyrodes; and blastocysts were plated onto mitotically inactivated MEFs. Outgrowths were picked and expanded using standard methods. All blastocysts were plated, cultured and expanded using 2i media (Li *et al.* (2008) Germline competent embryonic stem cells derived from rat blastocysts, Cell 135:1299-1310; incorporated herein by reference in its entirety).

Table 1. Rat ES Cell Derivation

	ACI	**DA**
Embryo source	Blastocysts (Superovulation)	Frozen 8-cell embryos cultured to blastocyst
Blastocysts plated:	107	22
Outgrowths:	32 (30% of blasts)	10 (45% of blasts)
Lines:	16 (50% of outgrowths)	9 (90% of outgrowths)
Karyotyped:	3; all 42X,Y	6: 3 42X,X 3 42X,Y
GLT validated:	1 (ACI.G1)	1 42X,X (DA.2C) 1 42X,Y (DA.2B)

1.2. : Rat Production

[00656] Chimeric rats were produced by blastocyst injection and transmission of the rat ESC genome. Chimeras produced by blastocyst microinjection using parental ACI.G1 rat ESCs are shown in Figure 9. F1 agouti pups with albino littermates, sired by the ACI/SD chimera labeled with an asterisk (*) in Figure 9 are shown in Figure 10.

Germline transmission of parental rat ESC.

[00657] Three euploid rat ESC lines were evaluated for pluripotency by microinjection into albino SD blastocysts. Chimeras were identified by agouti coat color, which indicates rat ESC contribution (see Figure 10). For each line, a majority of chimeras transmitted the rESC genome to F1 offspring (Table 2).

Table 2. Germline Transmission of Parental rESC

Line (Karyotype)	Chimeras bred	Germline transmitters	Total pups from GLT chimeras	rESC-derived pups	GLT efficiency (%)
ACI.G1 (XY)	5	**3 (60%)**	103	11	11
DA.2B (XY)	5	**4 (80%)**	129	11	9
DA.2C (XX)	3	**2 (66%)**	45	7	16

1.3. : Derivation of Rat Embryonic Stem Cells.

[00658] **Superovulation protocol, rats**

[00659] Day 0: injected with pregnant mare serum: IP, 20 U (0.4 ml).

[00660] Day 1: no action

[00661] Day 2: (46 hr. later): injected with hCG, IP, 50 U (1 ml).

[00662] - set up single female matings.

[00663] Day 3: checked plugs. Females were plugged. This is day 0.5.

[00664] Day 6 (e3.5): Euthanized females and flushed embryos.

[00665] **ES Cell derivation protocol (superovulation)**

[00666] Day 0:

[00667] 1) Euthanized female rat with CO_2.

[00668] 2) Swabbed ventral abdomen with 70% ethanol; using scissors, opened the ventral body wall to expose the viscera.

[00669] 3) Dissected out the oviducts and uterine horns and placed them into a tissue culture dish containing warm N2B27

media. Washed out as much blood as possible and transferred to a new dish with N2B27.

[00670] 4) Using a 1 ml syringe and a blunt 27g needle, flushed media through the uterine horns and oviducts to eject blastocysts into the media.

[00671] 5) Collected the blastocysts with a mouth pipet and transfer to embryo culture dish containing KSOM + 2i (1μMPD0325901, 3 μM CHIR99021). KSOM is a culture medium produced by Millipore. Catalog number is MR-106-D.

[00672] 6) Cultured overnight at 37°; 7.5% CO_2.

[00673] **ES Cell derivation protocol (frozen embryos)**

[00674] Day 0:

[00675] 1) Thawed frozen 8-cell embryos (commercially obtained) into M2 medium. Cultured 10 minutes at room temperature.

[00676] 2) Transferred to KSOM + 2i and culture overnight.

[00677] **ES Cell derivation protocol (same for both)**

[00678] Day 1:

[00679] 1) Transferred cavitated embryos to 2i medium & culture overnight.

[00680] 2) Continued culturing un-cavitated embryos in KSOM +2i

[00681] Day 2:

[00682] 1) Transferred all remaining embryos to 2i medium (whether or not they've cavitated).

[00683] 2) Cultured overnight; continued culturing earlier embryos in 2i medium.

[00684] Day 3:

[00685] 1) Transferred embryos for 30 – 60 seconds with Acid Tyrodes to remove the zona pellucida.

[00686] 2) Washed embryos 3x in 2i medium to remove Acid Tyrodes.

[00687] 3) Deposited each embryo into a separate well of a 96-well feeder plate (the well contains a monolayer of mitotically inactivated mouse embryonic fibroblasts (MEFs).

[00688] 4) Cultured overnight in 2i medium.

[00689] Day 4 – 5:

[00690] 1) Monitored plated embryos for the presence of an outgrowth (an amorphous undifferentiated mass of cells). Outgrowths are ready for transfer when they are approximately twice the size of the plated embryo.

[00691] 2) Each day: remove spent media with a mircropipet and replace with fresh 2i media.

[00692] 3) Transferred outgrowths to new feeder wells:

[00693] a. Removed spent media and gently wash well with PBS.

[00694] b. Removed PBS and add 30 μl 0.05% trypsin; incubate for 10 minutes.

[00695] c. Stopped trypsin reaction by adding 30μl 2i + 10% FBS.

[00696] d. Gently dissociated the cells with a micropipettor and transferred entire contents of the well to a new well in a 24-well feeder plate. This was Passage 1 (P1).

[00697] e. Cultured overnight in 2i medium.

[00698] Day 5 – 8: (timing depends on how fast each line expands)

[00699] 1) Changed media each day (2i media) and monitored for the presence of colonies with an ESC morphology.

[00700] 2) When colonies appear, continued culturing until colonies expand to ~ 50% confluency.

[00701] 3) Trypsinized and passaged colonies as before; plated on feeders, 1 well per line, in a 6-well dish. This was Passage 2 (P2).

[00702] Ongoing:

[00703] 1) Continued feeding and monitoring each line until approximately 50% confluent.

[00704] 2) Trypsinized cells as usual.

[00705] 3) stopped trypsin with 2i + 10% FBS; pelleted the cells by centrifugation (5', 1200 rpm in Beckman-Coulter tabletop centrifuge).

[00706] 4) Aspirated the supernatant and gently resuspend the cells in 400 μl Freezing Medium (70% 2i, 20% FBS, 10% DMSO).

[00707] 5) Distributed the cells into 2 vials and freeze at -80°. This was Passage 3 (P3).

[00708] 6) For long-term storage, transferred the vials to liquid N_2 storage.

[00709] The 2i media was prepared as follows in Table 3.

Reagent	Vendor	Concentration
DMEM/F12 basal media	Invitrogen/Life Technologies	1x
Neurobasal media	Invitrogen/Life Technologies	1x
Penicillin/streptomycin	Invitrogen/Life Technologies	1%
L-Glutamine	Invitrogen/Life Technologies	4 mM
2-Mercaptoethanol	Invitrogen/Life Technologies	0.1 mM
N2 supplement	Invitrogen/Life Technologies	1x
B27 supplement	Invitrogen/Life Technologies	1x
LIF	Millipore	100 U/ml
PD0325901 (MEK inhibitor).	Stemgent	1 uM
CHIR99021 (GSK inhibitor).	Stemgent	3 uM

[00710] **Materials:** Pregnant Mare's Serum Gonadotropin (PMSG)

[00711] Human Pregnancy Urine Chorionic Gonadotropin (HCG)

[00712] Female Rats (5-12 weeks old)

[00713] Male rats (12 wks. to 8 mos. old), one per cage

[00714] Syringes/needles

[00715] Animal room with lights on 6:00-18:00

[00716] **Procedure:**

[00717] Day 1: 8:00-10:00 AM

[00718] Inject females with 20 IU PMSG (0.4 ml), IP

[00719] Discard unused PMSG.

[00720] Day 3: 8:00-10:00 AM (48 hours after PMSG injection)

[00721] Inject females with 50 IU HCG (1 ml), IP

[00722] Place one female per male in mating cage.

[00723] Discard unused HCG.

[00724] Day 4: 8:00-10:00 AM (24 hrs. after HCG injection)

[00725] Check females for plugs.

[00726] **Hormone suppliers**

[00727] PMSG: Sigma #G-4877 (1000 IU). Resuspend in PBS to a final [ ] of 50 IU/ml. Store at -20° in 1 ml aliquots.

[00728] HCG: Sigma #CG-5 (5000 IU). Resuspend in PBS to a final [ ] of 50 IU/ml. Store at -20° in 1 ml aliquots.

1.4.: Karyotyping of Rat Embryonic Stem Cell Lines

[00729] The rat ES cell lines generated herein were karyotyped, and the results are summarized in Tables 4-7.

[00730] Table 4

ACI.G1 Karyotyping Results	**Number of cells**
Number of cells karyotyped	7
Number of cells analyzed	20
Number of 42, XY cells	18

Number of abnormal cells	2
40, XY, -5, -9	1
41, XY, -14	1
42, XY	18
Other notes: Two analyzed cells were missing different autosomes, which may be a sporadic occurrence due to technical artifact. 90% of analyzed cells had a normal male 42, XY karyotype.	
Figure 5 provides a photograph showing the analysis of the chromosome number of the ACI.G1 rat ES cell line.	

[00731] Table 5

DA.2B Karyotyping Results	**Number of cells**
Number of cells karyotyped	6
Number of cells analyzed	20
Number of 42, XY cells	20
Number of abnormal cells	0
42, XY	20
Other notes: All analyzed cells had a normal diploid 42, XY karyotype.	
Figure 6 provides a photograph showing the analysis of the chromosome number of the DA.2B rat ES cell line.	

[00732] Table 6

DA.2C Karyotyping Results	**Number of cells**
Number of cells karyotyped	5
Number of cells analyzed	20
Number of 42, XX cells	20
Number of abnormal cells	0
42, XX	
Other notes: 100% of analyzed cells had normal female XX rat karyotype.	
Figure 7 provides a photograph showing the analysis of the chromosome number of the DA.2C rat ES cell line.	

[00733] Table 7.

strain	Blastocysts plated	Lines established	Lines Karyotyped	Karyotypes
BN x SD F1	41	8 (20%)	5	all lines were high % complex polyploid
ACI	27	16 (60%)	3	G1: 90% 42 XY; others were 70-85% euploid
DA	20	9 (45%)	6	2B: 100% 42 XY; 2C: 100% 42 XX; others were 95-100% euploid
F344	4	1 (25%)	0	
Totals	92	34 (37%)		

1.5.: Electroporation of Vector into Rat Embryonic Stem Cell

[00734] 1. Passaged rat ES cells 24-48 hrs prior to electroporation.

[00735] 2. Changed media to RVG2i + ROCKi (10μM Y-27632) 24 hr. prior to electroporation

[00736] 3. Changed media 30' prior to trypsinization.

[00737] 4. Aliquoted DNA to be electroporated.

[00738] 5. Allowed DNA to warm at RT for >10 min.

[00739] 6. Heated DNA for 5' @ 62°C. Place DNA on ice.

[00740] 7. Trypsinized cells:

[00741] a. Collected floating colonies. Washed plate to collect as many floaters as possible.

[00742] b. Pelleted colonies: 3' @ 750 rpm.

[00743] c. Washed pellet 1x with 5-10ml PBS and re-spin/pellet

[00744] d. Aspirated supernatant; add 500λ trypsin, 0.05% + 1% chicken serum.

[00745] i. Did not pool more than 1 10 cm plate of colonies per tube. If there are too many colonies packed into the bottom of the tube during trypsinization they will clump and most of the cells will be lost.

[00746] e. 4' @ 37°. Pipeted colonies several times to minimize clumping.

[00747] f. Repeated steps 1-2 X: 4' @ 37°.

[00748] g. Stopped trypsin with 500λ RVG2i + 10% FBS.

[00749] 8. Pelleted cells: 5' @ 1200 rpm.

[00750] 9. Resuspend cells in 10 ml PBS. Count two 20λ aliquots to determine total cell number.

[00751] 10. Pelleted cells (5'/1200rpm); calculate total cell number and total resuspension volume to achieve correct cell concentration (target #/75 μl EP buffer).

[00752] 11. Resuspend in a minimal volume of EP buffer; measure total volume and adjust to target volume with EP buffer. Electroporation buffer is sold by Millipore. The catalog # is ES-003-D. See, Valenzuela et al. (2003) Nature Biotechnology 21:652-659, which is herein incorporated by reference.

[00753] 12. Add 75λ cells to 50λ DNA; transfer the 125λ cells/DNA solution to one well of a BTX 48-well cuvette.

[00754] a. Filled the empty wells in the same column with 125λ EP buffer.

[00755] 13. Pulsed the cuvette once in the BTX electroporator:

[00756] a. Settings: 400V; Ω; 100 μF (settings may vary)

[00757] 14. Placed cuvette on ice for 15' to recover.

[00758] 15. Removed cells into 5 ml RVG2i + 10 μM ROCKi.

[00759] 16. Added to 15 cm plate with 20 ml RVG2i + 10μM ROCKi. Plate has 2x neoR MEFs (or other MEFs depending on project). The neoR selectable marker is the neomycin phosphotransferase (neo) gene of Beck et al. (1982) Gene, 19:327-36 or in US Patent No, 7,205,148 or 6,596,541, each of which are herein incorporated by reference.

[00760] 17. Incubated @ 37°. Begin selection 48hrs later.

[00761] ROCK inhibitor used was Y-27632.

1.6: Selecting a Targeted Genetic Modification in a Rat Embryonic Stem Cell.

[00762] 1. Passaged cells for 24-48 hrs prior to electroporation.

[00763] 2. Changed media to RVG2i + ROCKi (10μM Y-27632) 24 hr. prior to electroporation

[00764] 3. Changed media 30' prior to trypsinization.

[00765] 4. Aliquoted DNA to be electroporated.

[00766] 5. Allowed DNA warm at RT for >10 min.

[00767] 6. Heated DNA for 5' @ 62°C. Place DNA on ice.

[00768] 7. Trypsinized cells:

[00769] a. Collected floating colonies. Washed plate to collect as many floaters as possible.

[00770] b. Pelleted colonies: 3' @ 750 rpm.

[00771] c. Washed pellet 1x with 5-10ml PBS and re-spin/pellet

[00772] d. Aspirated supernatant; add 500λ trypsin, 0.05% + 1% chicken serum.

[00773] i. Did not pool more than 1 10 cm plate of colonies per tube. If there are too many colonies packed into the bottom of the tube during trypsinization they will clump and most of the cells will be lost.

[00774] e. 4' @ 37°. Pipeted colonies several times to minimize clumping

[00775] f. Repeated 1-2 X: 4' @ 37°.

[00776] g. Stopped trypsin with 500λ RVG2i + 10% FBS.

[00777] 8. Pelleted cells: 5' @ 1200 rpm.

[00778] 9. Resuspended cells in 10 ml PBS. Count two 20λ aliquots to determine total cell number.

[00779] 10. Pelleted cells (5'/1200rpm); calculate total cell number and total resuspension volume to achieve correct cell concentration (target #/75 μl EP buffer).

[00780] 11. Resuspend in a minimal volume of EP buffer; measured total volume and adjusted to target volume with EP buffer.

[00781] 12. Added 75λ cells to 50λ DNA; transfer the 125λ cells/DNA solution to one well of a BTX 48-well cuvette.

[00782] a. Filled the empty wells in the same column with 125λ EP buffer.

[00783] 13. Pulsed the cuvette once in the BTX electroporator:

[00784] a. Settings: 400V; 100 μF (settings may vary)

[00785] 14. Placed cuvette on ice for 15' to recover.

[00786] 15. Removed cells into 5 ml RVG2i + 10μM ROCKi.

[00787] 16. Added to 15 cm plate with 20 ml RVG2i + 10μM ROCKi. Plate had 2x neoR MEFs (or other MEFs depending on project).

[00788] 17. Incubated @ 37°. Began selection 48hrs later.

[00789] 18. G418 selection protocol was as follows:

[00790] a. Day 2 (2nd day after EP): incubated cells in 2i media + G418, 75 μg/ml.

[00791] b. Day 3: incubated cells in 2i media without G418

[00792] c. Day 4: incubated cells in 2i media + G418, 75 μg/ml.

[00793] d. Day 5: incubated cells in 2i media without G418

[00794] e. Day 6: incubated cells in 2i media + G418, 75 μg/ml.

[00795] f. Day 7: incubated cells in 2i media without G418

[00796] g. Day 8: incubated cells in 2i media + G418, 75 μg/ml.

[00797] h. Day 9: incubated cells in 2i media without G418

[00798] i. Day 10: incubated cells in 2i media + G418, 75 μg/ml.

[00799] j. Day 11: incubated cells in 2i media without G418

[00800] k. Day 12: picked colonies to expand for screening. Each colony was dissociated in 0.05% trypsin + 1% chicken serum for 10 minutes and then plated into 1 well of a 96-well feeder plate.

[00801] 19. Expanded colonies for 3 days in 2i media.

[00802] 20. Passaged clones 1:1 to new 96-well feeder plates.

[00803] 21. Expanded clones for 3 days in 2i media.

[00804] 22. For each clone, dissociated colonies in trypsin. Froze 2/3 of each clone and store at -80°; plated the remaining 1/3 onto laminin plates (96-well plates coated with 10 μg/ml laminin).

[00805] 23. When the laminin plates were confluent, passed off to the screening lab for genotyping of the clones.

1.7. Molecular Signature of the Rat Embryonic Stem Cells

[00806] The genes listed in Table 8 were expressed at 20-fold lower in rat ES cells than the corresponding genes in mouse ES cells. The genes listed in Table 9 were expressed at levels 20-fold higher in rat ES cells than the corresponding genes in mouse ES cells.

[00807] The microarray data in Tables 8 and 9 were generated as follows. Rat ES cells (ACI.G2 and DA.2B) and mouse ES cells (F1H4) were cultured in 2i media for 3 passages until confluent. F1H4 cells were cultured on gelatin-coated plates in the absence of feeders. F1H4 mouse ES cells were derived from 129S6/SvEvTac and C57BL/6NTac heterozygous embryos (see, e.g., US Pat. No. 7,294,754 and Poueymirou, W.T., Auerbach, W., Frendewey, D., Hickey, J.F., Escaravage, J.M., Esau, L., Dore, A.T., Stevens, S., Adams, N.C., Dominguez, M.G., Gale, N.W., Yancopoulos, G.D., DeChiara, T.M., Valenzuela,D.M. (2007), incorporated by reference herein in its entirety).

[00808] The following protocol was used for sample prep: The 1.5mL Eppendorf tubes were labeled with the Sample ID. Cells grown on a plate were rinsed in 37°C Phosphate-Buffered Saline (PBS). PBS was removed and 300 ul of Trizol® was added. A scraper was used to break the cells in Trizol® (Life Technology). The lysed cells were collected in Trizol® in a 1.5mL Eppendorf tube. For cells grown on suspension, the cells were rinsed in 37°C PBS and collected in a 1.5mL tube. The cells were spun down; PBS was removed; and 300 ul of Trizol® was added to the cells. The cell membranes were broken by pipetting. Samples were sorted for FACS with 10 to 10^5 cells, the volume was concentrated to less than 100uL. 4 volumes of RNA Lysis buffer were added and mixed by pipetting. For sample, 320uL RNA Lysis buffer was added to 80uL sample. Samples were stored at –20°C.

[00809] RNA-Seq was used to measure the expression level of mouse and rat genes. Sequencing reads were mapped to mouse and rat reference genome by Tophat, and RPKM (fragments per kilobase of exon per million fragments mapped) were calculated for mouse and rat genes. Homology genes based on gene symbol were selected, and then used *t*-test to compare the expression level of each gene between mouse and rat. miR-32 was in the top 10 highest expressed in rat ESCs but was not expressed in mouse ES cells.

Although no comparative data exist from miR-632, based on the level of its expression compared to other genes expressed in rat ESCs and their known function in embryonic development, miR-632 was selected as a marker for rat ES cells.

[00810] Table 8. The genes listed were expressed at levels 20-fold lower in rat ES cells than the corresponding genes in mouse ES cells.

ID	Symbol	Entrez Gene Name	Location	Type(s)
Abcb1b	Abcb1b	ATP-binding cassette, sub-family B (MDR/TAP), member 1B	Plasma Membrane	transporter
Acta2	ACTA2	actin, alpha 2, smooth muscle, aorta	Cytoplasm	other
Actg2	ACTG2	actin, gamma 2, smooth muscle, enteric	Cytoplasm	other
Aebp1	AEBP1	AE binding protein 1	Nucleus	peptidase
Angptl2	ANGPTL2	angiopoietin-like 2	Extracellular Space	other
Ankrd1	ANKRD1	ankyrin repeat domain 1 (cardiac muscle)	Cytoplasm	transcription regulator
Anxa1	ANXA1	annexin A1	Plasma Membrane	other
Anxa6	ANXA6	annexin A6	Plasma Membrane	other
Anxa8	ANXA8L2	annexin A8-like 2	Plasma Membrane	other
Arhgef25	ARHGEF25	Rho guanine nucleotide exchange factor (GEF) 25	Cytoplasm	other
Axl	AXL	AXL receptor tyrosine kinase	Plasma Membrane	kinase
Basp1	BASP1	brain abundant, membrane attached signal protein 1	Nucleus	transcription regulator
Bgn	BGN	biglycan	Extracellular Space	other
Bst2	BST2	bone marrow stromal cell antigen 2	Plasma Membrane	other
Btf3	BTF3	basic transcription factor 3	Nucleus	transcription regulator
Btg2	BTG2	BTG family, member 2	Nucleus	transcription regulator
Capsl	CAPSL	calcyphosine-like	Other	other
Cav1	CAV1	caveolin 1, caveolae protein, 22kDa	Plasma Membrane	transmembrane receptor
Ccdc80	CCDC80	coiled-coil domain containing 80	Nucleus	other
Ccnd2	CCND2	cyclin D2	Nucleus	other

ID	Symbol	Entrez Gene Name	Location	Type(s)
Cd248	CD248	CD248 molecule, endosialin	Plasma Membrane	other
Cd44	CD44	CD44 molecule (Indian blood group)	Plasma Membrane	enzyme
Cd97	CD97	CD97 molecule	Plasma Membrane	G-protein coupled receptor
Cdc42ep5	CDC42EP5	CDC42 effector protein (Rho GTPase binding) 5	Cytoplasm	other
Cdh11	CDH11	cadherin 11, type 2, OB-cadherin (osteoblast)	Plasma Membrane	other
Cdkn2a	CDKN2A	cyclin-dependent kinase inhibitor 2A	Nucleus	transcription regulator
Cdo1	CDO1	cysteine dioxygenase type 1	Cytoplasm	enzyme
Clip3	CLIP3	CAP-GLY domain containing linker protein 3	Cytoplasm	other
Cln5	CLN5	ceroid-lipofuscinosis, neuronal 5	Cytoplasm	other
Cnn1	CNN1	calponin 1, basic, smooth muscle	Cytoplasm	other
Col1a1	COL1A1	collagen, type I, alpha 1	Extracellular Space	other
Col1a2	COL1A2	collagen, type I, alpha 2	Extracellular Space	other
Col3a1	COL3A1	collagen, type III, alpha 1	Extracellular Space	other
Col5a2	COL5A2	collagen, type V, alpha 2	Extracellular Space	other
Col6a2	COL6A2	collagen, type VI, alpha 2	Extracellular Space	other
Cryab	CRYAB	crystallin, alpha B	Nucleus	other
Csf1	CSF1	colony stimulating factor 1 (macrophage)	Extracellular Space	cytokine
Cth	CTH	cystathionase (cystathionine gamma-lyase)	Cytoplasm	enzyme
Cthrc1	CTHRC1	collagen triple helix repeat containing 1	Extracellular Space	other
Ctsc	CTSC	cathepsin C	Cytoplasm	peptidase
Cyr61	CYR61	cysteine-rich, angiogenic inducer, 61	Extracellular Space	other
Ddx58	DDX58	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	Cytoplasm	enzyme
Dkk3	DKK3	dickkopf WNT signaling pathway inhibitor 3	Extracellular Space	cytokine
Dmc1	DMC1	DNA meiotic recombinase 1	Nucleus	enzyme

ID	Symbol	Entrez Gene Name	Location	Type(s)
Dpysl3	DPYSL3	dihydropyrimidinase-like 3	Cytoplasm	enzyme
Dse	DSE	dermatan sulfate epimerase	Cytoplasm	enzyme
Dusp1	DUSP1	dual specificity phosphatase 1	Nucleus	phosphatase
Dusp27	DUSP27	dual specificity phosphatase 27 (putative)	Other	phosphatase
Dusp9	DUSP9	dual specificity phosphatase 9	Nucleus	phosphatase
Ece2	ECE2	endothelin converting enzyme 2	Plasma Membrane	peptidase
Ecm1	ECM1	extracellular matrix protein 1	Extracellular Space	transporter
Egr1	EGR1	early growth response 1	Nucleus	transcription regulator
Emp1	EMP1	epithelial membrane protein 1	Plasma Membrane	other
Emp3	EMP3	epithelial membrane protein 3	Plasma Membrane	other
Ephx2	EPHX2	epoxide hydrolase 2, cytoplasmic	Cytoplasm	enzyme
F3	F3	coagulation factor III (thromboplastin, tissue factor)	Plasma Membrane	transmembrane receptor
Fau	FAU	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed	Cytoplasm	other
Fbn1	FBN1	fibrillin 1	Extracellular Space	other
Fbxo15	FBXO15	F-box protein 15	Other	transcription regulator
Fhl2	FHL2	four and a half LIM domains 2	Nucleus	transcription regulator
Flnc	FLNC	filamin C, gamma	Cytoplasm	other
Fos	FOS	FBJ murine osteosarcoma viral oncogene homolog	Nucleus	transcription regulator
Fundc2	FUNDC2	FUN14 domain containing 2	Cytoplasm	other
Gjb3	GJB3	gap junction protein, beta 3, 31kDa	Plasma Membrane	transporter
Gpa33	GPA33	glycoprotein A33 (transmembrane)	Plasma Membrane	other
Gpbp1l1	GPBP1L1	GC-rich promoter binding protein 1-like 1	Other	other
Gpc3	GPC3	glypican 3	Plasma Membrane	other
Grb10	GRB10	growth factor receptor-bound protein 10	Cytoplasm	other
Gstm1	GSTM5	glutathione S-	Cytoplasm	enzyme

ID	Symbol	Entrez Gene Name	Location	Type(s)
		transferase mu 5		
Hap1	HAP1	huntingtin-associated protein 1	Cytoplasm	other
Hist1h2bc	HIST2H2BE (includes others)	histone cluster 2, H2be	Nucleus	other
Hmga2	HMGA2	high mobility group AT-hook 2	Nucleus	enzyme
Hmgn3	Hmgn3	high mobility group nucleosomal binding domain 3	Nucleus	other
Hormad1	HORMAD1	HORMA domain containing 1	Nucleus	other
Hsd17b14	HSD17B14	hydroxysteroid (17-beta) dehydrogenase 14	Cytoplasm	enzyme
Hspb1	HSPB1	heat shock 27kDa protein 1	Cytoplasm	other
Hspb8	HSPB8	heat shock 22kDa protein 8	Cytoplasm	kinase
Htra1	HTRA1	HtrA serine peptidase 1	Extracellular Space	peptidase
Ifi204	Ifi204 (includes others)	interferon activated gene 204	Nucleus	transcription regulator
Ifi44	IFI44	interferon-induced protein 44	Cytoplasm	other
Ifit1	IFIT1B	interferon-induced protein with tetratricopeptide repeats 1B	Cytoplasm	other
Ifitm3	IFITM2	interferon induced transmembrane protein 2	Cytoplasm	other
Igf2	IGF2	insulin-like growth factor 2 (somatomedin A)	Extracellular Space	growth factor
Igfbp7	IGFBP7	insulin-like growth factor binding protein 7	Extracellular Space	transporter
Il1rl1	IL1RL1	interleukin 1 receptor-like 1	Plasma Membrane	transmembrane receptor
Inhba	INHBA	inhibin, beta A	Extracellular Space	growth factor
Inhbb	INHBB	inhibin, beta B	Extracellular Space	growth factor
Irf7	IRF7	interferon regulatory factor 7	Nucleus	transcription regulator
Isg15	ISG15	ISG15 ubiquitin-like modifier	Extracellular Space	other
Itga5	ITGA5	integrin, alpha 5 (fibronectin receptor, alpha polypeptide)	Plasma Membrane	transmembrane receptor
Jun	JUN	jun proto-oncogene	Nucleus	transcription regulator

ID	Symbol	Entrez Gene Name	Location	Type(s)
Junb	JUNB	jun B proto-oncogene	Nucleus	transcription regulator
Lgals3bp	LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein	Plasma Membrane	transmembrane receptor
Lgals9	LGALS9	lectin, galactoside-binding, soluble, 9	Extracellular Space	other
Lmna	LMNA	lamin A/C	Nucleus	other
Lox	LOX	lysyl oxidase	Extracellular Space	enzyme
Loxl2	LOXL2	lysyl oxidase-like 2	Extracellular Space	enzyme
Loxl3	LOXL3	lysyl oxidase-like 3	Extracellular Space	enzyme
Lrp1	LRP1	low density lipoprotein receptor-related protein 1	Plasma Membrane	transmembrane receptor
Mageb16	MAGEB16	melanoma antigen family B, 16	Other	other
Mcam	MCAM	melanoma cell adhesion molecule	Plasma Membrane	other
Mgp	MGP	matrix Gla protein	Extracellular Space	other
Mmp2	MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	Extracellular Space	peptidase
Mxra8	MXRA8	matrix-remodelling associated 8	Other	other
Myl9	MYL9	myosin, light chain 9, regulatory	Cytoplasm	other
Mylpf	MYLPF	myosin light chain, phosphorylatable, fast skeletal muscle	Cytoplasm	other
Nab2	NAB2	NGFI-A binding protein 2 (EGR1 binding protein 2)	Nucleus	transcription regulator
Ndufb4	NDUFB4	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa	Cytoplasm	transporter
Npm1	NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	Nucleus	transcription regulator
Nr0b1	NR0B1	nuclear receptor subfamily 0, group B, member 1	Nucleus	ligand-dependent nuclear receptor
Nr4a1	NR4A1	nuclear receptor subfamily 4, group A, member 1	Nucleus	ligand-dependent nuclear receptor
Nrp2	NRP2	neuropilin 2	Plasma Membrane	kinase
Oas1a	OAS1	2'-5'-oligoadenylate	Cytoplasm	enzyme

ID	Symbol	Entrez Gene Name	Location	Type(s)
		synthetase 1, 40/46kDa		
Oasl2	Oasl2	2'-5' oligoadenylate synthetase-like 2	Other	enzyme
P4ha2	P4HA2	prolyl 4-hydroxylase, alpha polypeptide II	Cytoplasm	enzyme
Parp3	PARP3	poly (ADP-ribose) polymerase family, member 3	Nucleus	enzyme
Pcolce	PCOLCE	procollagen C-endopeptidase enhancer	Extracellular Space	other
Pcyt1b	PCYT1B	phosphate cytidylyltransferase 1, choline, beta	Cytoplasm	enzyme
Pdgfc	PDGFC	platelet derived growth factor C	Extracellular Space	growth factor
Phlda1	PHLDA1	pleckstrin homology-like domain, family A, member 1	Cytoplasm	other
Phlda2	PHLDA2	pleckstrin homology-like domain, family A, member 2	Cytoplasm	other
Pla2g1b	PLA2G1B	phospholipase A2, group IB (pancreas)	Extracellular Space	enzyme
Pla2g4a	PLA2G4A	phospholipase A2, group IVA (cytosolic, calcium-dependent)	Cytoplasm	enzyme
Porcn	PORCN	porcupine homolog (*Drosophila*)	Cytoplasm	other
Postn	POSTN	periostin, osteoblast specific factor	Extracellular Space	other
Prrx1	PRRX1	paired related homeobox 1	Nucleus	transcription regulator
Prss23	PRSS23	protease, serine, 23	Extracellular Space	peptidase
Psmb8	PSMB8	proteasome (prosome, macropain) subunit, beta type, 8	Cytoplasm	peptidase
Ptgs2	PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	Cytoplasm	enzyme
Ptn	PTN	pleiotrophin	Extracellular Space	growth factor
Ptrf	PTRF	polymerase I and transcript release factor	Nucleus	transcription regulator
Rarg	RARG	retinoic acid receptor, gamma	Nucleus	ligand-dependent nuclear receptor
Rgs16	RGS16	regulator of G-protein signaling 16	Cytoplasm	other
Rn45s	Rn45s	45S pre-ribosomal RNA	Other	other
Rpl10a	RPL10A	ribosomal protein L10a	Other	other

ID	Symbol	Entrez Gene Name	Location	Type(s)
Rpl31	RPL31	ribosomal protein L31	Other	other
Rpl37a	RPL37A	ribosomal protein L37a	Cytoplasm	other
Rps10	RPS10-NUDT3	RPS10-NUDT3 readthrough	Cytoplasm	other
Rps14	RPS14	ribosomal protein S14	Cytoplasm	translation regulator
Rps20	Rps20	ribosomal protein S20	Cytoplasm	other
Rps26	RPS26	ribosomal protein S26	Cytoplasm	other
Rps9	RPS9	ribosomal protein S9	Cytoplasm	translation regulator
S100a4	S100A4	S100 calcium binding protein A4	Cytoplasm	other
S100a6	S100A6	S100 calcium binding protein A6	Cytoplasm	transporter
Schip1	SCHIP1	schwannomin interacting protein 1	Cytoplasm	other
Sdc2	SDC2	syndecan 2	Plasma Membrane	other
Serpine1	SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	Extracellular Space	other
Serpine2	SERPINE2	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2	Extracellular Space	other
Serpinf1	SERPINF1	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	Extracellular Space	other
Sh3gl2	SH3GL2	SH3-domain GRB2-like 2	Plasma Membrane	enzyme
Slc19a2	SLC19A2	solute carrier family 19 (thiamine transporter), member 2	Plasma Membrane	transporter
Slc25a5	SLC25A5	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 5	Cytoplasm	transporter
Slc29a1	SLC29A1	solute carrier family 29 (equilibrative nucleoside transporter), member 1	Plasma Membrane	transporter
Slc35f2	SLC35F2	solute carrier family 35, member F2	Other	other
Snrpn	SNRPN	small nuclear ribonucleoprotein polypeptide N	Nucleus	other

ID	Symbol	Entrez Gene Name	Location	Type(s)
Snx22	SNX22	sorting nexin 22	Other	transporter
Sparc	SPARC	secreted protein, acidic, cysteine-rich (osteonectin)	Extracellular Space	other
Spp1	SPP1	secreted phosphoprotein 1	Extracellular Space	cytokine
Sult4a1	SULT4A1	sulfotransferase family 4A, member 1	Cytoplasm	enzyme
Tagln	TAGLN	transgelin	Cytoplasm	other
Tcea3	TCEA3	transcription elongation factor A (SII), 3	Nucleus	transcription regulator
Tgfb3	TGFB3	transforming growth factor, beta 3	Extracellular Space	growth factor
Thbs1	THBS1	thrombospondin 1	Extracellular Space	other
Thbs2	THBS2	thrombospondin 2	Extracellular Space	other
Tm4sf1	TM4SF1	transmembrane 4 L six family member 1	Plasma Membrane	other
Tmbim1	TMBIM1	transmembrane BAX inhibitor motif containing 1	Cytoplasm	other
Tmem176b	TMEM176 B	transmembrane protein 176B	Other	other
Tnc	TNC	tenascin C	Extracellular Space	other
Tpd52l1	TPD52L1	tumor protein D52-like 1	Cytoplasm	other
Tpm2	TPM2	tropomyosin 2 (beta)	Cytoplasm	other
Usp18	USP18	ubiquitin specific peptidase 18	Cytoplasm	peptidase
Vim	VIM	vimentin	Cytoplasm	other
Wfdc2	WFDC2	WAP four-disulfide core domain 2	Extracellular Space	other
Wisp2	WISP2	WNT1 inducible signaling pathway protein 2	Extracellular Space	growth factor
Ybx1	YBX1	Y box binding protein 1	Nucleus	transcription regulator

[00811] Table 9. The genes listed were expressed at levels 20-fold higher in rat ES cells than the corresponding genes in mouse ES cells.

ID	Symbol	Entrez Gene Name	Location	Type(s)
Ajap1	Ajap1	adherens junction associated protein 1	Other	other
Amd1	AMD1	adenosylmethionine decarboxylase 1	Cytoplasm	enzyme
Ankrd2	ANKRD 2	ankyrin repeat domain 2 (stretch responsive muscle)	Nucleus	transcription regulator
Arhgef9	ARHGE F9	Cdc42 guanine nucleotide exchange factor (GEF) 9	Cytoplasm	other
Atp5h	Atp5h	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	Cytoplasm	enzyme
Btg3	BTG3	BTG family, member 3	Nucleus	other
Car6	CA6	carbonic anhydrase VI	Extracellular Space	enzyme
Camk4	CAMK4	calcium/calmodulin-dependent protein kinase IV	Nucleus	kinase
Capn12	CAPN12	calpain 12	Other	peptidase
Cct6b	CCT6B	chaperonin containing TCP1, subunit 6B (zeta 2)	Cytoplasm	transporter
Cdx2	CDX2	caudal type homeobox 2	Nucleus	transcription regulator
Cldn5	CLDN5	claudin 5	Plasma Membrane	other
Clec3a	CLEC3A	C-type lectin domain family 3, member A	Other	other
Clic6	CLIC6	chloride intracellular channel 6	Plasma Membrane	ion channel
Dhrsx	DHRSX	dehydrogenase/reductase (SDR family) X-linked	Other	enzyme
Dpysl2	DPYSL2	dihydropyrimidinase-like 2	Cytoplasm	enzyme
Dusp26	DUSP26	dual specificity phosphatase 26 (putative)	Cytoplasm	enzyme
Eci3	Eci3	enoyl-Coenzyme A delta isomerase 3	Other	enzyme
Eef2k	EEF2K	eukaryotic elongation factor-2 kinase	Cytoplasm	kinase
Efna1	EFNA1	ephrin-A1	Plasma Membrane	other
Epha4	EPHA4	EPH receptor A4	Plasma Membrane	kinase
Fank1	FANK1	fibronectin type III and ankyrin repeat domains 1	Nucleus	transcription regulator
Fhit	FHIT	fragile histidine triad	Cytoplasm	enzyme
Filip1	FILIP1	filamin A interacting protein 1	Cytoplasm	other
Fmod	FMOD	fibromodulin	Extracellular Space	other
Foxe1	FOXE1	forkhead box E1 (thyroid transcription factor 2)	Nucleus	transcription regulator

ID	Symbol	Entrez Gene Name	Location	Type(s)
Fry	FRY	furry homolog (Drosophila)	Extracellular Space	other
Gjb5	GJB5	gap junction protein, beta 5, 31.1kDa	Plasma Membrane	transporter
Gpx2	GPX2	glutathione peroxidase 2 (gastrointestinal)	Cytoplasm	enzyme
Grxcr2	GRXCR 2	glutaredoxin, cysteine rich 2	Other	other
Hecw2	HECW2	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2	Extracellular Space	enzyme
Hey2	HEY2	hairy/enhancer-of-split related with YRPW motif 2	Nucleus	transcription regulator
Icos	Icos	inducible T-cell co-stimulator	Plasma Membrane	other
Ifitm1	IFITM1	interferon induced transmembrane protein 1	Plasma Membrane	transmembrane receptor
Il1f8	IL1F8 (IL36B)	Interleukin-1 family member (Interleukin 36 beta)	Extracellular space	cytokine
Il28ra	IL-28RA	Interleukin 28 receptor, alpha	Plasma membrane	Cytokine receptor
Igfbpl1	IGFBPL 1	insulin-like growth factor binding protein-like 1	Other	other
Ipcef1	IPCEF1	interaction protein for cytohesin exchange factors 1	Cytoplasm	enzyme
Lctl	Lctl	lactase-like	Cytoplasm	other
Ldhd	LDHD	lactate dehydrogenase D	Cytoplasm	enzyme
Lef1	LEF1	lymphoid enhancer-binding factor 1	Nucleus	transcription regulator
Lefty1	LEFTY1	left-right determination factor 1	Extracellular Space	growth factor
Lifr	LIFR	leukemia inhibitory factor receptor alpha	Plasma Membrane	transmembrane receptor
Lpar2	LPAR2	lysophosphatidic acid receptor 2	Plasma Membrane	G-protein coupled receptor
Mog	MOG	myelin oligodendrocyte glycoprotein	Extracellular Space	other
Morn5	MORN5	MORN repeat containing 5	Other	other
Pigz	NCBP2	nuclear cap binding protein subunit 2, 20kDa	Nucleus	other
Nptxr	NPTXR	neuronal pentraxin receptor	Plasma Membrane	transmembrane receptor
Ntm	NTM	neurotrimin	Plasma Membrane	other
Nutf2	NUTF2	nuclear transport factor 2	Nucleus	transporter
Ocln	OCLN	occludin	Plasma Membrane	enzyme
Olr1	OLR1	oxidized low density lipoprotein (lectin-like) receptor 1	Plasma Membrane	transmembrane receptor

ID	Symbol	Entrez Gene Name	Location	Type(s)
Pabpc4	PABPC4	poly(A) binding protein, cytoplasmic 4 (inducible form)	Cytoplasm	translation regulator
Pde11a	PDE11A	phosphodiesterase 11A	Cytoplasm	enzyme
Pdyn	PDYN	prodynorphin	Extracellular Space	transporter
Per3	PER3	period circadian clock 3	Nucleus	other
Pllp	PLLP	plasmolipin	Plasma Membrane	transporter
Ppp1r14c	PPP1R14C	protein phosphatase 1, regulatory (inhibitor) subunit 14C	Cytoplasm	other
Pramel6	Pramel6	preferentially expressed antigen in melanoma like 6	Other	other
Ptpn18	PTPN18	protein tyrosine phosphatase, non-receptor type 18 (brain-derived)	Nucleus	phosphatase
Pycr1	PYCR1	pyrroline-5-carboxylate reductase 1	Cytoplasm	enzyme
Rab26	RAB26	RAB26, member RAS oncogene family	Plasma Membrane	enzyme
Ramp2	RAMP2	receptor (G protein-coupled) activity modifying protein 2	Plasma Membrane	transporter
Rbm24	RBM24	RNA binding motif protein 24	Other	other
Rhag	RHAG	Rh-associated glycoprotein	Plasma Membrane	peptidase
Rpl3	RPL3	ribosomal protein L3	Cytoplasm	other
Sall3	SALL3	sal-like 3 (Drosophila)	Nucleus	other
Satb1	SATB1	SATB homeobox 1	Nucleus	transcription regulator
Scg2	SCG2	secretogranin II	Extracellular Space	cytokine
Slc15a1	SLC15A1	solute carrier family 15 (oligopeptide transporter), member 1	Plasma Membrane	transporter
Slc1a1	SLC1A1	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	Plasma Membrane	transporter
Slc24a5	Slc24a5	solute carrier family 24 (sodium/potassium/calcium exchanger), member 5	Other	other
Slc37a2	SLC37A2	solute carrier family 37 (glucose-6-phosphate transporter), member 2	Other	transporter
40424	SNTB1	syntrophin, beta 1 (dystrophin-associated protein A1, 59kDa, basic component 1)	Plasma Membrane	other
St6galnac3	ST6GALNAC3	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 3	Cytoplasm	enzyme

ID	Symbol	Entrez Gene Name	Location	Type(s)
Tex12	TEX12	testis expressed 12	Nucleus	other
Tex15	TEX15	testis expressed 15	Extracellular Space	other
Tfap2a	TFAP2A	transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)	Nucleus	transcription regulator
Tmc1	TMC1	transmembrane channel-like 1	Plasma Membrane	other
Tmem130	TMEM130	transmembrane protein 130	Other	other
Tmem30b	TMEM30B	transmembrane protein 30B	Other	other
Tomm20	TOMM20	translocase of outer mitochondrial membrane 20 homolog (yeast)	Cytoplasm	transporter
Tox3	TOX3	TOX high mobility group box family member 3	Other	other
Ttc25	TTC25	tetratricopeptide repeat domain 25	Cytoplasm	other
Tymp	TYMP	thymidine phosphorylase	Extracellular Space	growth factor
Ubb	Ubb	ubiquitin B	Cytoplasm	other
Vamp7	VAMP7	vesicle-associated membrane protein 7	Cytoplasm	transporter
Wfdc12	Wfdc12	WAP four-disulfide core domain 12	Extracellular Space	other
Wfdc15a	Wfdc15a	WAP four-disulfide core domain 15A	Other	other
Wfdc6a	Wfdc6a	WAP four-disulfide core domain 6A	Other	other

[00812] Table 10. A subset of genes from Table 9, which are expressed at levels 20-fold higher in rat ES cells than the corresponding genes in mouse ES cells.

ID	Entrez Gene Name
Ajap1	Adherens Junctions Associate Protein
Cldn5	Claudin 5
Arhgef9	Cdc42 guanine nucleotide exchange facter 9
Camk4	Calcium/calmodulin-dependent protein kinase IV
Efna1	ephrin-A1
Epha4	EPH receptor A4
Gjb5	gap junction protein beta 5
Igfbpl1	Insulin-like growth factor binding protein-like 1
Il1f8	Interleukin 36 beta
Il28ra	Interleukin 28 receptor, alpha
Lefty1	left-right determination factor 1
Lifr	Leukemia inhibitory factor receptor alpha
Lpar2	Lysophosphatidic acid receptor 2
Ntm	Neuronal pentraxin receptor
Ptpn18	Protein tyrosine phosphatase non-receptor type 18
Cdx2	Caudal type homeobox 2

ID	Entrez Gene Name
Fank1	Fibronectin type III and ankyrin repeat domains 1
Foxe1	Forkhead box E1 (thyroid transcription factor 2)
Hey2	Hairy/enhancer-of-split related with YRPW motif 2
Lef1	Lymphoid enhancer-binding factor 1
Sall3	Sal-like 3 (Drosophila)
Satb1	SATB homeobox 1

[00813] An additional molecular signature employing the pluripotency markers/genes for the rat ES cells has also been developed. Table 11 provides a gene list and their expression ranks from the RNA profiling data. mRNA was isolated from rat ES cells and the expression levels of various markers were compared relative to each other. The term "rank" means the comparative expression levels of individual genes: the higher the rank (1 is highest), the higher the expression. For example, Oct4's rank of 13 means that, of all the genes assayed, it was expressed higher than all but 12 genes. Background in this experiment was any expression value below 30; 6107 genes had expression values of 30 or higher.

[00814] Table 11. Rat ES cell molecular signature employing various pluripotency, mesodermal, endodermal, neural and trophectoderm markers/genes.

Pluripotency	Pluripotency Rank	Mesodermal	Mesodermal Rank	Endodermal	Endodermal Rank	Neural	Neural Rank	Troph-ectoderm	Troph-ectoderm Rank
c-Myc	8248	Brachyury	7542	Gata6	11195	Nestin	7761	Cdx2	739
Dnmt3L	127	Flk1	Not tested	Sox17	11418	Pax6	13570		
Dppa2	Not tested	Nodal	3050	Hhex1	4571	Sox2	681		
Dppa5	Not tested	Bmp4	3072	Nodal	3050				
Ecat1	9714	Bmpr2	6382	Ext1	6091				
Eras	2541			Sox7	10284				
Err-beta	1368								
Fbxo15	1369								
Fgf4	3440								
Fthl17	Not tested								
Gdf3	2771		Rank > 6107 = bkg expression						
Klf4	836								
Lef1	1313								
LIF receptor	724								
Lin28	828								

Nanog	774								
Oct4	13								
Rexo1	6119								
Sox15	4524								
Sox2	681								
SSEA1	Not tested								
SSEA4	Not tested								
Stella	Not tested								
Tcl1	Not tested								
Utf1	1501								

Example 2: Inactivation of Genomic Loci in Rats

2.1: Inactivation of Endogenous Genomic Loci Using an Endonuclease Agent

[00815] In order to introduce a mutant allele at an endogenous rat genomic locus, the rat ES cells described herein are electroporated with expression vectors (or mRNA) that express ZFNs 1 and 2 (or TALENs 1 and 2). These proteins bind their target sequences on opposite strands, separated by about 6 bp to about 40 bp. A double-stranded break is formed within the target locus, which the cell attempts to repair by Non-Homologous End-Joining (NHEJ). In many cases, NHEJ results in creation of a deletion, which often disrupts the function of the gene (most often by producing a frameshift mutation). In order to identify a positive clone comprising a mutant allele, the electroporated cells are plated at low density, because no drug selection is done. Colonies are picked and assayed at the target site to see if a mutation was produced (*e.g.*, using a modification of allele (MOA) assay described above). The selected ES cells comprising the mutant allele are then introduced into a host rat embryo, for example, a pre-morula stage or blastocyst stage rat embryo, and implanted in the uterus of a surrogate mother to generate a founder rat (F0 rat). Subsequently, the founder rat is bred to a wild-type rat to create F1 progeny heterozygous for the mutant allele. Mating of the heterozygous F1 rat can produce progeny homozygous for the mutant allele.

2.2.: Rat ESC Targeting for the Inactivation of the Rat Apolipoprotein E (ApoE) gene Using Zinc Finger Nucleases

[00816] Zinc finger nucleases use sequence specific modular DNA binding domains to direct endonuclease activity to unique target sequence in the genome. ZFNs are engineered as a pair of monomers. Each monomer contains nonspecific cleavage domain from *FokI* endonuclease fused to 3 or more zinc finger DNA-binding domains. Each zinc finger binds a 3 bp subsite and specificity is achieved by the combined target sites of both monomers. ZFNs produce double-stranded breaks (DSBs) in DNA, and mutations (insertions or deletions) frequently occur during non-homologous end joining (NHEJ). Figure 15 illustrates the mechanism by which genome-editing endonucleases such as ZFNs and TALENs introduce double strand breaks in a target genomic sequence and activate NHEJ in a cell. DSBs also stimulate homology-directed repair (HDR) by homologous recombination if a donor sequence is provided with ZFN.

[00817] Such ZFNs were employed in combination with the various methods and compositions described herein to improve targeting efficiency. The rat Apolipoprotein E (*ApoE*) locus was targeted as described in Example 3.2(a)(i), except expression vectors that express ZFNs 1 and 2 were also introduced into the rat ES cells. See Figure 11, which provides a schematic of the *ApoE* targeting event in combination with rTZFN1P and rTZFN2P. The targeting efficiency was determined as discussed below in Example 5 and results are shown in Table 12. To screen for heterozygous targeting, homozygous targeting, and "mixed" doubles (e.g., compound heterozygous targeting), specific primers and probes were used to determine genotype. Surprisingly, the targeting efficiency went up 8-10 fold.

[00818] Table 12. Rat *ApoE* ZFNs: Improved Targeting Efficiency.

DNA	Colonies	Screened	Heterozygous Targeted	Homozygous Targeted	"Mixed" Doubles	Cut/ Untargeted
vector	330	184	15 (8.2%)	0	0	N/A
vector + ZFN 1	560	192	132 (68.8%)	6 (3.1%)	18 (9.4%)	17 (8.9%)
vector + ZFN 2	410	192	136 (70.8%)	2 (1.0%)	6 (3.1%)	18 (9.4%)

[00819] A plasmid targeting vector was built with a self-deleting drug selection cassette and a lacZ gene as a reporter gene (see Figure 14 for an illustration of the

homologous and non-homologous recombination events that can occur upon electroporation of a targeting vector comprising a selection cassette). Good targeting efficiency was achieved and high % chimeras were produced. Zinc finger nucleases (ZFNs) were also tested in combination with targeting vectors to examine its effect on improving targeting efficiency (see Figure 16 for an illustration of the gene targeting technique utilizing ZFNs or TALENs to improve the efficiency of homologous recombination of a targeting vector). The targeting vector was co-expressed with the expression vectors for 2 ZFN pairs that cut the *ApoE* locus. The rat ESC clones electroporated with both the targeting vector and a set of the ZFNs showed a targeting efficiency of 8-10 fold higher than that of rat ESC clones electroporated with a targeting vector alone. Moreover, bi-allelic homozygous targeting in about 2% of our clones was detected. High % chimeras from two of these targeted clones were obtained.

[00820] The *ApoE*-targeted (with ZFN assistance) rat ESC clones were microinjected into SD blastocysts, which were then transferred to pseudopregnant SD recipient females, using standard techniques. Chimeras were identified by coat color (see Figure 17, showing ApoE-ZFN-AB5 chimeras (i.e., $ApoE^{-/-}$ chimeras); male F0 chimeras were bred to SD females. Germline F1 pups were genotyped for the presence of the targeted *ApoE* allele (Table 13). High % chimeras were obtained from two of these targeted clones.

[00821] Table 13. Microinjection Results.

		(% of chimera)
ApoE-ZFN1-AB5 (homozygous targeted)	12	4 (90, 90, 80, 80)
ApoE-ZFN1-AE5 (homozygous targeted)	6	3 (90, 80, 70)

[00822] An *ApoE* knockout rat provides a means to study various types of disorders and diseases. In humans, Apolipoprotein is found in chylomicron, HDL, LDL and VLDL. ApoE is essential for the normal catabolism of triglyceride-rich lipoprotein constituents. Defects in APOE result in numerous disease states including, for example, familial hypercholesterolemia, hyperlipidemia, betalipoproteinemia, familial dysbetalipoproteinemia, type III hyperlipoproteinemia (HLP III), risk of coronary artery

disease. One isoform (ApoE4) is associated with late-onset familial and sporadic Alzheimer's disease, possibly with MS as well.

[00823] In mice, ApoE is primarily found in HDL; transports cholesterol, as in humans. *ApoE*-deficient mice (2 independent KOs) have 5 times normal plasma cholesterol; developed foam cell-rich depositions in their proximal aortas by age 3 months (comparable to human syndrome).

[00824] *ApoE* knockouts in rats offer an animal model to study endothelial function, including, but not limited to, plaque formation, transcriptional changes (RNA-Seq), *ex vivo* function. Moreover, larger size of rats would facilitate all these assays and potentially improve the quality of the RNA-Seq data.

2.3. Inactivation of The Rat Interleukin-2 Receptor Gamma (IL2r-γ) Locus Using Zinc Finger Nucleases

[00825] The rat Interleukin-2 receptor gamma (IL2r-γ or *Il2rg*) locus was targeted as described in Example 3.3(a), except that expression vectors that express ZFN U (ZFN upstream) and ZFN D (ZFN downstream) were also introduced into the rat ES cells. Figure 18 provides a schematic of the IL2r-γ targeting event in combination with ZFN U and ZFN D. The sequence of the IL2r-γ locus which these zinc fingers bind is denoted in Figure 18 within SEQ ID NO: 93. The targeting efficiency was determined as discussed below in Example 3.3(a) and the results are shown in Table 14. Briefly, homozygously targeted clones were confirmed by PCR. For the ZFN1 pair: 173 mutant clones out of 192 screened (90%) and for the ZFN2 pair: 162 clones out of 192 (84%) screened.

[00826] Table 14. Targeting of Rat IL2r-γ Locus.

Plate	Colonies screened	Targeted	Efficiency
7/18: Vector alone	96	4	4.2%
7/18: Vector + ZFN	96	3	3.1%

[00827] The IL2r-γ -targeted (with ZFN assistance) rat ESC clones were microinjected into SD blastocysts, which were then transferred to pseudopregnant SD recipient females, using standard techniques. Chimeras were identified by coat color; male F0 chimeras were bred to SD females. Germline F1 pups were genotyped for the presence of the targeted IL2r-γ allele.

2.4.: Inactivation of The Rat Interleukin-2 Receptor Gamma (IL2r-γ) using CRISPR/Cas9

[00828] The rat IL2r-γ locus was targeted as described in Example 3.3(a), except that the CRISPR/Cas9 system was also introduced into the rat ES cells to aid in targeting efficiency. SBI: System Biosciences Cas9 "SmartNuclease" all-in-one vectors were employed and Cas9 expression was driven by CAG, EF1a, PGK, or CMV promoter. Custom gRNA was ligated into a vector and expressed by H1 promoter. 4 gRNAs against *Il2rg* were designed. The regions of the rat IL2r-γ locus targeted by gRNAs1-4 are shown in Figure 19. To screen for targeting (e.g., heterozygous targeting, homozygous targeting, and compound heterozygous targeting), specific primers and probes were used to determine genotype. Targeting results when employing the various guide RNAs is shown in Table 15. "Strong" and "weak" refer to the strength of the evidence based on screening that the colony has a targeted modification.

[00829] Table 15. Targeting of Rat *Il2rg* Locus with Guide RNAs.

Construct(s)	DNA (ug)	Colonies	Candidates (Potentially Targeted)
Il2rg plasmid vector	6 ug	30	3 weak
plasmid + SBI gRNA1	6 ug/40 ug	22	1 strong, 1 weak
plasmid + SBI gRNA2	6 ug/40 ug	45	2 strong, 1 weak
plasmid + SBI gRNA3	6 ug/40 ug	66	1 strong, 2 weak
plasmid + SBI gRNA4	6 ug/40 ug	59	0

2.5.: Inactivation of the Mouse Hypoxanthine Guanine Phosphoribosyl Transferase (Hprt) gene using CRISPR/Cas9

[00830] The mouse *Hprt* locus was targeted in mouse ES cells using LTVECs alone or in combination with CRISPR/Cas9. The 32.9 kb complete *Hprt* coding sequence was targeted for deletion and replacement with the pCAGG-Puro puromycin resistance selection cassette, which also expressed eGFP. The deletion end points were the start and stop codons. The guide RNA sequence used was 5'-

GACCCGCAGUCCCAGCGUCG-3′ (SEQ ID NO: 84), which targeted exon 1 of the mouse *Hprt* gene. The predicted target site cleavage position was 22 base pairs from the 5' end of the deletion. The Cas9/gRNA on-target cleavage efficiency observed in the ES cells was ≥ 93%. A summary is shown in Table 16. Use of CRISPR/Cas9 to assist in targeting of the complete 32.9 kb *Hprt* locus resulted in a five-fold enhancement of targeting over use of LTVEC alone.

[00831] Table 16. Summary of CRISPR-Assisted Deletion of *Hprt* Gene

Summary of CRISPR-Assisted Deletion of the *Hprt* Gene						
				Targeting Efficiency (%)		
Target Gene	Deletion (kb)	5′Homology Arm (kb)	3′Homology Arm (kb)	LTVEC Alone	LTVEC + CRISPR/Cas9	Fold Enhancement
Hprt	32.9	88	66	5.0	25.4	5.1

Example 3: Targeted Modification of Rat Genomic Loci

3.1: Rat ESC Targeting: The Rat Rosa26 Locus.

[00832] The rat *Rosa26* locus lies between the *Setd5* and *Thumpd3* genes as in mouse, with the same spacing. The rat *Rosa26* locus (Figure 12, Panel B) differs from the mouse *Rosa26* locus (Figure 12, Panel A). The mouse *Rosa26* transcripts consist of 2 or 3 exons. The rat locus contains a 2nd exon 1 (Ex1b) in addition to the homologous exon to mouse exon1 (Ex1a). No 3rd exon has been identified in rat. Targeting of a rat *Rosa26* allele is depicted in Figure 12C, where homology arms of 5 kb each were cloned by PCR using genomic DNA from DA rat ESC. The targeted allele contains a SA (splicing acceptor)-lacZ-hUb-neo cassette replacing a 117 bp deletion in the rat *Rosa26* intron.

[00833] Targeting efficiency at the rat *Rosa26* locus was determined (Table 17). Linearized vector was electroporated into DA or ACI rat ESCs, and transfected colonies were cultured in 2i media + G418, using standard techniques. Individual colonies were picked and screened using a Loss of Allele (LOA) assay (Valenzuela, D. *et al.* (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, Nature Biotech. 21:652-660, incorporated herein by reference).

Table 17. rat Rosa26 Targeting Efficiency			
Cell line	**Colonies picked**	**Reconfirmed positives**	**Targeting efficiency (%)**
DA.2B	192	4	2.1
ACI.G1	96	4	4.2

[00834] Chimera production and germline transmission using Rosa26-targeted rat ESC clones. Reconfirmed *Rosa26*-targeted rat ESC clones were microinjected into SD blastocysts, which were then transferred to pseudopregnant SD recipient females, using standard techniques. Chimeras were identified by coat color; male F0 chimeras were bred to SD females. Germline (agouti) F1 pups were genotyped for the presence of the targeted *Rosa26* allele; nine of 22 agouti pups genotyped as heterozygous at the *Rosa26* locus (Table 18).

Table 18. Germline Transmission Using Targeted Rosa26 rESC						
Cell line	**R26 clones injected**	**Clones producing Chimeras**	**Germline Transmitting Clones**	**Total Pups**	**rESC-derived Pups**	**ESC-derived pups (%)**
DA.2B	4	3	2	AH7: 64 AE3: 112	AH7: 41 AE3: 6	AH7: 63 AE3: 3
ACI.G1	4	4	1	DE9: 39	DE9: 4	10

[00835] To confirm that the genetically modified allele at the *Rosa26* locus was transmitted through the germline, lacZ expression was confirmed by X-gal staining in heterozygous *Rosa26*-targeted rats. X-gal staining of the brain, heart and thymus, and a lung from a 14-week-old heterozygous *Rosa26*-targeted rat showed expression of lacZ (Figure 13B, D, and F, respectively), whereas age-matched wild type controls showed a low level of background X-gal staining (Figure 13A, C, and E, respectively). X-gal

staining in E12.5 and E 14.5 heterozygous *Rosa26*-targeted rat embryos showed ubiquitous expression of lacZ (Figure 13G and I, respectively), whereas control rat embryos showed low levels of background X-gal staining (Figure 13H and J, respectively).

3.2.(a)(i) : Targeting of the Rat Apolipoprotein E (ApoE) Locus.

[00836] The rat Apolipoprotein E (*ApoE*) locus was targeted to disrupt ApoE function. Targeting of the *ApoE* locus was done using a targeting vector comprising a lacZ-hUb-neo cassette flanked with a 5' and 3' homology arms homologous to the *ApoE* locus. Figure 20 depicts a genetically modified rat *ApoE* locus that has been disrupted by a 1.8 kb deletion and the insertion of a lacZ-hUb-neo cassette, which further includes a self-deleting Cre cassette comprising a Crei gene driven by a protamine promoter. The electroporation conditions were as follows: 6 ug DNA; 2.05 x 10^6 cells; 400V; 200 uF: 342 V, 593 usec; plate on 15 cm 2x dense neoR MEFs in 2i + 10 uM ROCKi.

[00837] Targeting efficiency at the *ApoE* locus was determined and is shown in Table 19. Linearized vector was electroporated into DA.2B rat ESCs derived from the DA strain, and transfected colonies were cultured using standard techniques. Individual colonies were picked and screened using a Loss of Allele (LOA) assay.

	Table 19. rat ApoE Targeting Efficiency			
Cell line	**Vector**	**Colonies picked**	**Targeted**	**Targeting efficiency (%)**
DA.2B	ApoE-mSDC	192	7	3.7
DA.2B	ApoE-mSDC	192	15	7.8

[00838] Chimera production and germline transmission using *ApoE*-targeted rat ESC clones was performed. *ApoE*-targeted rat ESC clones were microinjected into SD blastocysts, which were then transferred to pseudopregnant SD recipient females, using standard techniques. Chimeras were identified by coat color; male F0 chimeras were bred to SD females. Germline transmission was achieved. F1 pups were genotyped for the presence of the targeted *ApoE* allele (Table 20).

Table 20 Microinjection Results

Exp	Clone	Pups	Chimeras (% of chimera)
1	ApoE-AF5	4	3 (90, 90, 90)
2	ApoE-BC4	5	0

[00839] LacZ expression driven by the endogenous *ApoE* promoter was confirmed by X-gal staining in 12-week-old *ApoE* $^{+/-}$ female rats in the brain, blood vessels, and liver (Figures 43-45, respectively). Figures 43-45 show an expression pattern for lacZ that mirrors the expression pattern of endogenous *ApoE*. Age-matched wild type controls showed a low level of background X-gal staining.

[00840] The phenotypes of *ApoE*-deleted rats were further studied. Longitudinal serum chemistry studies were performed to measure cholesterol, LDL, HDL, and triglyceride levels at three-week intervals. Figure 46A-D show serum cholesterol, LDL, HDL, and triglyceride levels in homozygous targeted, heterozygous targeted, and wild type rats at 6 weeks, 9 weeks, 12 weeks, and 15 weeks of age. Eye bleeds were performed on an age-matched cohort consisting of 2 wild type, 7 heterozygous, and 8 homozygous rats. No significant differences were seen between males and females. Homozygous *ApoE*-deleted rats showed elevated cholesterol and LDL levels and decreased HDL levels. Unlike *ApoE* $^{-/-}$ mice, no significant increase in triglycerides was observed in *ApoE*-deleted rats.

[00841] Additional phenotypic analysis that is performed includes histology/*ex vivo* imaging for aortic arch plaque formation, *in vivo* imaging for aortic arch plaque formation, and transcriptional changes (Whole Transcriptome Shotgun Sequencing (RNA-Seq)) for aortic arch endothelium. The timing of these assays depends on the timeline of plaque formation. Plaques are detectable in *ApoE* $^{-/-}$ mice at 24 weeks.

[00842] Additional targeting data for *ApoE* is also provided in Table 22.

3.2.(a)(ii). Targeting ApoE in Rats with a Targeting Vector

[00843] Figure 20 provides a schematic of the rat *ApoE* locus and a targeting plasmid. The upper schematic of Figure 20 shows the genomic structure of the rat *ApoE* locus and the genomic regions corresponding to 5' and 3' homology arms (5 kb and 5.4 kb, respectively; dark grey boxes). Exon 1 of *ApoE* is non-coding and is shown as an open box closest to the 5' homology arm. The 3 introns of *ApoE* are denoted as lines and

exons 2 and 3 comprise coding regions and are shown as stippled grey boxes. Exon 4 contains both coding and non-coding sequences as denoted by the stippled grey shading and the open box.

[00844] The lower schematic in Figure 20 is the targeting vector. The 5' and 3' homology arms (5 kb and 5.4 kb respectively) are denoted by the dark grey boxes. The targeting vector comprises a reporter gene (lacZ) and a self-deleting cassette flanked by loxP sites (open arrows). The self-deleting cassette comprises the Crei gene operably linked to a mouse *Prm1* promoter and a selection cassette comprising a neomycin resistance gene operably linked to a human ubiquitin promoter.

[00845] The Crei gene comprises two exons encoding a Cre recombinase, which are separated by an intron (Crei) to prevent its expression in a prokaryotic cell. See, for example, U.S. Patent 8,697,851 and U.S. Application Publication 2013-0312129, which describe the self-deleting cassette in detail and are hereby incorporated by reference in their entirety. By employing the *Prm1* promoter, the self-deleting cassette can be deleted specifically in male germ cells of F0 rats. The targeting vector was electroporated into the rat ES cells obtained in Example 1 and the cells were plated on 15 cm 2x dense neomycin-resistant MEFs in 2i + 10 uM ROCKi. The transformed rat ES cells were cultured, selected, and maintained as described in Example 1.

[00846] As shown in Table 44, 384 colonies were screened and 23 targeted clones were obtained. The targeting efficiency was 5.99%. 3 clones were injected into blastocysts as described herein in Example 1. 3 clones producing chimeras were obtained and 1 of the clones transmitted the targeted modification through the germline.

3.2.(a)(iii). Targeting ApoE in Rats with a Targeting Vector in Combination with Zinc Finger Nucleases

[00847] The targeting vector employed in Example 3.2(a)(ii) was used in combination with zinc finger nucleases to target the rat *ApoE* locus. Table 21 provides a summary of the genomic organization of the rat *ApoE* locus. The positions shown in the Table 21 were taken from build 5.0 of the Reference Sequence of the rat genome (ENSMBL). *ApoE* is on chromosome 1 on the (-) strand.

[00848] Table 21. Summary of the rat *ApoE* locus and the positions of the zinc finger nuclease binding sites and cutting sites.

Feature	Start	End	length	Notes
Exon 1	81881110	81881182	73	5' non-coding
Exon2	81880269	81880332	64	contains ATG
ATG	81880309	81880311	3	start codon
Exon3	81879607	81879775	169	
ZFN1a binding site	81879707	81879693	15	CAGGCCCTGAACCGC (SEQ ID NO: 10)
ZFN1 cutting site	81879692	81879687	6	TTCTGG (SEQ ID NO: 11)
ZFN1b binding site	81879686	81879671	16	GATTACCTGCGCTGGG (SEQ ID NO: 12)
Intron 3-4	81879776	81879207	400	
ZF21a binding site	81879591	81879577	15	TTCACCCTCCGCACC (SEQ ID NO: 13)
ZFN2 cutting site	81879576	81879570	7	TGCTGAG (SEQ ID NO: 14)
ZF21b binding site	81879569	81879552	18	TATCCAGATCCAGGGGTT (SEQ ID NO: 15)
Exon 4	81878371	81879208	838	contains TGA
TGA	81878482	81878484	3	
ApoE deletion	81878482	81880311	1830	

[00849] Figure 11 provides a schematic of the rat *ApoE* locus and denotes with grey bars the cutting site for ZFN1 and ZFN2. The cutting site for ZFN1 is in exon 3 and the cutting site for ZNF2 is in intron 3. The exact position of the both ZFN sites is set forth in Table 21. The genomic regions corresponding to the 5' and 3' homology arms (5 kb and 5.4 kb, respectively) are denoted by the dark grey boxes. Exon 1 of *ApoE* is non-coding and is shown as an open box closest to the 5' homology arm. The three introns of the *ApoE* gene are denoted as lines and exons 2 and 3 comprise coding regions and are shown as stippled grey boxes. Exon 4 contains both coding and non-coding sequences as denoted by the stippled grey shading and the open box.

[00850] The employed targeting vector was the same as that in Example 3.2(a)(ii) and shown in Figure 20, and Figure 21A provides a schematic for targeting the *ApoE* locus in rat ES cells using zinc-finger nucleases and the targeting vector depicted in Figure 20. The ZFNs were introduced as two expression plasmids, one for each half of the ZFN pair. 20 ug of the plasmid for ZFN1 and 20 ug of the plasmid for ZFN2 was

used. ZFNs were purchased from Sigma. The expression of each ZFN was driven by the CMV promoter.

[00851] The targeting vector were electroporated into the rat ES cells obtained in Example 1 and the cells were plated on 15 cm 2x dense neoR MEFs in 2i + 10 uM ROCKi. The transformed rat ES cells were cultured, selected and maintained as described in Example 1.

[00852] As shown in Table 22 and Table 44, 384 colonies were screened and 290 targeted clones were obtained. The targeting efficiency was 75.52%. 2 clones were injected into blastocysts as described herein in Example 1. Two clones producing chimeras were obtained and one of the clones transmitted the targeted modification through the germline.

[00853] Moreover, employing ZFN1 and ZFN2 produced 8 biallelic targeted clones with an efficiency of 2.08%.

[00854] Table 22. Targeting of *ApoE* Locus.

DNA	Heterozygous Targeted	Homozygous Targeted	Microinjected	Chimeras (% Chimerism)	Breeding
vector alone	15/192 (8%)	0			
vector + ZFN pair 1	156/192 (81%)	6/192 (3%)	2 clones	7 (70-90%)	
vector + ZFN pair 2	134/192 (70%)	2/192 (1%)			

3.2.(b)(i): Targeted Modification of the Rat Apolipoprotein E (ApoE) Locus Using a Large Targeting Vector (LTC).

[00855] Targeting of the *ApoE* locus is done using a large targeting vector (LTVEC) comprising a lacZ-mouse *Prm1*-Crei cassette flanked with a 5' homology arm to the *ApoE* locus of about 45 kb and a 3' homology arm to the *ApoE* locus of about 23 Kb. Figure 22 depicts the rat *ApoE* locus in which the *ApoE* locus has been disrupted by a 1.83 kb deletion and the insertion of the lacZ gene and a self-deleting cassette comprising m*Prm1*-Crei cassette and a hUb-neo selection cassette. Methods employed in example 3.2(a)(i) can be used to introduce this vector into rat ES cells.

Example 3.2.(b)(ii). Targeting of the Rat ApoE locus with a Large Targeting Vector (LTVEC)

[00856] Figure 22 provides a schematic of the rat *ApoE* locus and a large targeting vector (LTVEC). The upper schematic of Figure 22 shows the genomic organization of the rat *ApoE* locus and the genomic regions corresponding to the 5' and 3' homology arms (45 kb and 23 kb, respectively; dark grey boxes). Exon 1 of *ApoE* is non-coding and is shown as an open box closest to the 5' homology arm. The 3 introns of *ApoE* are denoted as lines and exons 2 and 3 comprise coding regions and are shown as stippled grey boxes. Exon 4 contains both coding and non-coding sequences as denoted by the stippled grey shading and the open box.

[00857] The lower schematic in Figure 22 is the LTVEC. The 5' and 3' homology arms (45 kb and 23 kb, respectively) are denoted by the dark grey boxes. The targeting vector comprises a reporter gene (lacZ) and a self-deleting cassette flanked by loxP sites (open arrows), which comprises the Crei gene operably linked to a mouse *Prm1* promoter and a drug selection cassette comprising a neomycin resistance gene operably linked to a human ubiquitin promoter. The Crei comprises two exons encoding the Cre recombinase which are separated by an intron (Crei) to prevent its expression in a prokaryotic cell. See, for example, U.S. Patent 8,697,851 and U.S. Application Publication 2013-0312129, which describes the self-deleting cassette in detail and is hereby incorporated by reference in their entirety. By employing a mouse *Prm1* promoter, the self-deleting cassette can be deleted specifically in male germ cells of F0 rat.

[00858] The LTVEC was electroporated into the rat ES cells obtained in Example 1 and the cells were plated on 15 cm 2x dense neoR MEFs in 2i + 10 uM ROCKi. The transformed rat ES cells were cultured, selected, and maintained as described in Example 1.

[00859] As shown in Table 44, 288 colonies were screened and 8 targeted clones were obtained. The targeting efficiency was 2.78%. 3 clones were injected into a host embryo at a blastocyst stage as described herein in Example 2 to produce chimeric rats (F0). Moreover, one biallelic targeted clone was produced providing a biallelic efficiency of 0.35%.

3.2.(b)(iii). Targeting ApoE in Rats with a Large Targeting Vector (LTVEC) In Combination with Zinc Finger Nucleases

[00860] The LTVEC employed in Example 3.2.(b)(ii) was used in combination with zinc finger nucleases to target the rat *ApoE* locus. Table 21 provides a summary of the genomic organization of the rat *ApoE* locus and the positions shown were taken from build 5.0 of the Reference Sequence of the rat genome (ENSMBL).

[00861] Figure 23 provides a schematic of the rat *ApoE* locus and denotes with grey bars the cutting site for ZFN1 and ZFN2. The cutting site for ZFN1 is in t exon 3 and the cutting site for ZNF2 is in intron 3. The exact position of the both ZFN sites is set forth in Table 21. The 5' and 3' homology arms (45 kb and 23 kb, respectively) are denoted by the dark grey boxes. Exon 1 of the *ApoE* gene is non-coding and is shown as an open box closest to the 5' homology arm. The three introns of the *ApoE* gene are denoted as lines. Exons 2 and 3 comprise coding regions and are shown as stippled grey boxes. Exon 4 contains both coding and non-coding sequences as denoted by the stippled grey shading and the open box.

[00862] The LTVEC employed was the same as that in Example 3.2(b)(ii) and shown in Figure 22. The ZFNs were introduced as two expression plasmids, one for each half of the ZFN pair. 20 ug of the plasmid for ZFN 1 and 20 ug of the plasmid for ZFN2 was used. ZFNs were purchased from Sigma. The expression of each ZFN was driven by the CMV promoter.

[00863] The targeting vector was electroporated into the rat ES cells obtained in Example 1 and the cells were plated on 15 cm 2x dense neoR MEFs in 2i + 10 uM ROCKi. The transformed rat ES cells were cultured, selected, and maintained as described in Example 1.

[00864] As shown in Table 44, 288 colonies were screened and 16 targeted clones were obtained. The targeting efficiency was 5.56%. One clone was injected into blastocysts as described herein in Example 2.

[00865] Moreover, the employment of ZFN1 and ZFN2 produced one biallelic targeted clone, with an efficiency of 0.35%.

3.2.(b)(iv). Targeting ApoE in Rats with a Large Targeting Vector (LTVEC) in Combination with CRISPR/Cas9

[00866] The LTVEC employed in Example 3.2.(b)(ii) was used in combination with CRISPR/Cas9 to target the rat *ApoE* locus. Table 23 shows a comparison of the results of experiments in which the *ApoE* LTVEC was used alone to target the rat *ApoE* locus or was used in combination with a CRISPR/Cas9 nuclease to target the rat *ApoE* locus. In each experiment, electroporated cells were plated at a high density and subjected to drug selection to find colonies that were drug-resistant. Drug-resistant colonies were picked and screened for the targeted modification using the modification of allele (MOA) assay as described herein. Specifically, 4 x 10^6 cells were electroporated with 2 ug of *ApoE* LTVEC at a voltage of 400V, a capacitance of 100 uF, and a resistance of 0. In the latter experiment, 6 ug of Cas9 expression plasmid and 3 ug of *ApoE* gRNA2 or 3 ug of *ApoE* gRNA3 were also electroporated. Selection was done using 75 ug/mL of G418. *ApoE* gRNA2 has a sequence of GCAGGCCCTGAACCGCTTCTTGG (SEQ ID NO: 87) and targets a region 67 bp 3' of the start of rat *ApoE* exon 3. *ApoE* gRNA3 has a sequence of CCTGCGCTGGGTGCAGACGCTTT (SEQ ID NO: 88) and targets a region 97 bp 3' of the start of rat *ApoE* exon 3 (see Figure 47). As shown in Table 23, when Cas9 and either of the gRNAs were introduced into the cells together with the *ApoE* LTVEC, targeting efficiency increased (from 43% to 53% or 47%). Biallelic targeting was observed in five colonies targeted with the *ApoE* LTVEC in combination with *ApoE* gRNA2 or 3, but no biallelic targeting was observed with *ApoE* LTVEC alone.

[00867] Table 23. Comparison of Rag2 LTVEC Targeting with and without CRISPR/Cas9

Vector	Cas9	gRNA	Colonies Screened	Targeted Clones	Biallelic Targeted	Targeting Efficiency
ApoE LTVEC	NO	NO	75	32	0	43%
ApoE LTVEC	Yes	ApoE gRNA2	80	42	1	53%
ApoE LTVEC	Yes	ApoE gRNA3	89	42	4	47%

3.3(a): Targeting of the Rat Interleukin-2 Receptor Gamma (IL2r-γ) Locus

[00868] The rat Interleukin-2 receptor gamma (IL2r-γ or *Il2rg*) locus was targeted to disrupt IL2r-γ function. IL2r-γ plays an important role for signaling by IL-2, IL-4, IL-

7, IL-9, IL-15, IL-21 and mutations in IL2r-γ are associated with severe defects in T, B and NK cell development.

[00869] Targeting of the IL2r-γ locus was done using a targeting vector comprising an eGFP-hUb-neo cassette flanked with a 5' and 3' homology arms homologous to the IL2r-γ locus, as depicted in Figure 24. Figure 25 depicts the genomic structure of the rat IL2r-γ locus in which the IL2r-γ locus has been disrupted by a 3.2 kb deletion. The targeted IL2r-γ locus also comprised an eGFP gene and a self-deleting cassette containing Crei operably linked to a mouse Protamine1 promoter and a drug selection cassette comprising a hUb promoter operably linked to a neomycin resistance gene.

[00870] Targeting efficiency at the IL2r-γ locus was determined and shown in Table 24. Linearized vector was electroporated into DA.2B rat ESCs, and transfected colonies were cultured using standard techniques. Individual colonies were picked and screened using a Loss of Allele (LOA) assay.

	Table 24. rat IL2r-γ Targeting Efficiency				
Cell line	**Vector**	**Colonies picked**	**Targeted**	**Targeting efficiency (%)**	**Chimeras (% Chimerism)**
DA.2B	Il2rg-floxed neo	136	1	0.7	5 (70-90%)
DA.2B	Il2rg-mSDC	96	4	4.2	

[00871] Chimera production and germline transmission using IL2r-γ-targeted rat ESC clones was performed. IL2r-γ-targeted rat ESC clones were microinjected into SD blastocysts, which were then transferred to pseudopregnant SD recipient females, using standard techniques. Chimeras were identified by coat color; male F0 chimeras were bred to SD females. Germline F1 pups were genotyped for the presence of the targeted IL2r-γ allele (Table 25). In another microinjection experiment with clone *Il2rg*-CG12, germline transmission was also confirmed by coat colors and genotyping.

[00872] Table 25. Microinjection Results

Exp	Clone	pups	Chimeras (% of chimera)

1	Il2rg-AA1	5	2 (90, 70)
2	Il2rg-AA1	10	3 (90, 90, 80)
3	Il2rg-CG12	11	7 (95, 90, 90, 90, 80, 80, 80)

[00873] The phenotype of *Il2rg*$^{-/Y}$ chimera #3 was further studied. The peripheral blood mononuclear cells (PBMCs) were stained with antibodies that recognize antigens in several lymphoid lineages. GFP-positive PBMCs were detected from 2 of the chimeras, as shown in Figure 30. Moreover, the GFP+ cells were negative for the T-cell marker CD3 (Figure 29A), and were mostly negative for the B-cell marker B220 and the NK cell marker CD161a (Figure 29B and C, respectively). PBMCs from a wild type rat were used as negative controls for GFP expression. See Figure 29D-F. The small double-positive populations are consistent with the published *Il2rg* knockout phenotype in mice. These data were obtained from a chimeric rat, which contains IL2 receptor gamma-positive cells, and this may complicate the analysis of the phenotype. Flow cytometry analysis can also be performed on cell populations from bone marrow and spleen to reveal corresponding decreases in the number of lymphocytes. See Mashimo et al. (2010) PLoS One 5(1):e8870.

3.3(b): Targeted Modification of The Rat Interleukin-2 Receptor Gamma (IL2r-γ) Locus

[00874] The rat Interleukin-2 receptor gamma (IL2r-γ) locus was targeted to disrupt the IL2r-γ function in rats. Figure 25 shows the genomic structure of the rat *Il2rg* locus (upper panel of Figure 25) and the targeting vector introduced into the locus (lower panel of Figure 25). eGFP was chosen as a reporter so that the immunophenotype of the genetically modified rats could be examined using FACS. The self-deleting cassette (hUb-Neo; *Prm1*-Cre) was used to delete the drug section cassette and the Cre gene specifically in male germ cells of the F0 rat. Additionally, the targeting vector was designed to delete the entire coding region (about 3.2 kb) of the rat *Il2rg* gene.

[00875] The size of the deletion in rat ESCs was confirmed by PCR using primers specific to the rat *Il2rg* locus. Upon microinjection of the targeted clones into host embryos at a blastocyst stage, high percentage chimeras were obtained. Those chimeras have been set up for breeding. To determine if the targeting worked as expected, the

peripheral blood from the chimeras were collected prior to breeding, and the phenotype of the immune cells in the peripheral blood was analyzed via FACS. As shown in Figure 30, GFP-positive cells were detected in the peripheral blood in 2 of the 3 chimeras examined, and the chimeric rats contained less than 1% of T cells, less than 1% of B cells, and less than 1% of NK-cells, which are positive for GFP (i.e., *Il2rg* KO cells) (Figure 29A-C).

3.4(a)(i). Targeting the Rag2 Locus in Rats with a Large Targeting Vector (LTVEC)

[00876] Table 26 provides a summary of the genomic organization of the rat *Rag2* locus and the positions shown were taken from build 5.0 of the Reference Sequence of the rat genome (ENSMBL). *Rag2* is on chromosome 3 on the (+) strand.

[00877] Table 26. Genomic organization summary of the rat *Rag2* locus.

Feature	Start	End	Length	Notes
Exon 1	97,851,317	97,851,448	132	
Exon 2	97,854,635	97,854,693	59	
Exon 3	97,858,260	97,859,615	1,356	contains entire coding sequence
ATG	97,856,286	97,856,288	3	start codon
TGA	97,857,867	97,857,869	3	stop codon
Rag2 deletion	97,856,289	97,859,784	3,496	

[00878] Figure 26 provides a schematic of the rat *Rag2* locus and a large targeting vector (LTVEC). The LTVEC is 140 kb and targets an approximately 5.7 kb portion of the rat *Rag2* locus for deletion. The upper schematic of Figure 26 shows the genomic organization of the rat *ApoE* locus and the genomic regions corresponding to the 5' and 3' homology arms (48 kb and 84 kb, respectively; dark grey boxes). *Rag2* comprises a single exon denoted by the stippled grey shading.

[00879] The lower schematic in Figure 26 is the LTVEC. The 5' and 3' homology arms (48 kb and 84 kb, respectively) are denoted by the dark grey boxes. The LTVEC comprises a reporter gene (lacZ) and a self-deleting cassette flanked by loxP sites (open

arrows). The self-deleting cassette comprises a mouse *Prm1* promoter operably linked to the Crei gene and a drug selection cassette comprising a human ubiquitin promoter operably linked to a neomycin resistance gene. Another version of the LTVEC was generated in which the neomycin resistance gene was replaced with a hygromycin resistance gene to enable retargeting of *Il2rg*-targeted rat ES cells. The Crei comprises two exons encoding the Cre recombinase that are separated by an intron (Crei) to prevent its expression in a prokaryotic cell. See, for example, U.S. Patent 8,697,851 and U.S. Application Publication 2013-0312129, which describe the self-deleting cassette in detail and are hereby incorporated by reference in their entirety. By employing a mouse *Prm1* promoter, the self-deleting cassette can be deleted specifically in male germ cells of F0 rats.

[00880] The LTVEC was electroporated into the rat ES cells obtained in Example 1 and the cells were plated on 15 cm 2x dense neoR MEFs in 2i + 10 uM ROCKi. The transformed rat ES cells were cultured and maintained as described in Example 1.

[00881] Colonies are screened as described elsewhere herein and targeted clones are obtained. The targeted clones are then injected into a host embryo as described elsewhere herein to produce an F0 rat.

3.4(a)(ii). Targeting the Rag2 Locus in Rats with a Large Targeting Vector (LTVEC) and CRISPR/Cas9

[00882] Table 27 shows a comparison of the results of experiments in which a version of the *Rag2* LTVEC having a hygromycin resistance gene (see Figure 48) was used alone to target the rat *Rag2* locus or was used in combination with a CRISPR/Cas9 nuclease to target the rat *Rag2* locus. In each experiment, electroporated cells were plated at a high density and subjected to drug selection to find colonies that were drug-resistant. Drug-resistant colonies were picked and screened for the targeted modification using the modification of allele (MOA) assay as described herein. Specifically, 4 x 10^6 cells were electroporated with 2 ug of *Rag2* LTVEC at a voltage of 400V, a capacitance of 100 uF, and a resistance of 0. In the latter experiment, 6 ug of Cas9 expression plasmid and 3 ug of *Rag2* gRNA1 or 3 ug of *Rag2* gRNA4 were also electroporated. Selection was done using 75 ug/mL of G418. *Rag2* gRNA1 has a sequence of

CCAGCTACTTGCTCGTACAA (SEQ ID NO: 89) and targets a region 219 bp 3' of the rat *Rag2* start codon (ATG). *Rag2* gRNA4 has a sequence of CCCCTCAGATTCACGTGCGT (SEQ ID NO: 90) and targets a region 12 bp 3' of the rat *Rag2* stop codon (TAG) (see Figure 48). As shown in Table 27, when Cas9 and either of the gRNAs were introduced into the cells together with the *Rag2* LTVEC, targeting efficiency increased (from 0 to 10% or 38%). Biallelic targeting was observed in one colony.

[00883] Table 27. Comparison of Rag2 LTVEC Targeting with and without CRISPR/Cas9

Vector	Cas9	gRNA	Colonies Screened	Targeted Clones	Biallelic Targeted	Targeting Efficiency
Rag2 LTVEC	NO	NO	36	0	0	0
Rag2 LTVEC	Yes	Rag2 gRNA1	23	5	1	22%
Rag2 LTVEC	Yes	Rag2 gRNA4	16	1	0	6%

3.4.(b)(i): Targeting the Rag1 and the Rag 2 Locus in Rats

[00884] Figure 27 provides the genomic structure of the rat *Rag1*/*Rag2* locus. CDS denotes the coding sequence and grey boxes represent exons. *Rag2* is on the "plus" strand with transcription to the right. *Rag1* is on the "minus" strand with transcription to the left. Mbp = million base pairs.

[00885] Table 28 provides a summary of the genomic organization of the rat *Rag2* and *Rag1* locus and the positions shown were taken from build 5.0 of the Reference Sequence of the rat genome (ENSMBL). *Rag1* is on chromosome 3 on the (-) strand.

[00886] Table 28. Genomic organization summary of the rat *Rag1* locus.

Feature	Start	End	Length	Notes
Exon 1	97,877,145	97,877,066	80	
Exon 2	97,872,503	97,866,047	6,457	contains entire coding sequence
ATG	97,872,489	97,872,487	3	start codon
TAA	97,869,369	97,869,367	3	stop codon
Rag1-2 deletion	97,856,289	97,872,486	16,198	

[00887] Figure 28 provides a schematic of the rat *Rag2* and *Rag1* locus and a large targeting vector (LTVEC). The LTVEC is about 70 kb and targets an approximately 16.6 kb rat genomic locus comprising the *Rag1* and *Rag2* loci for deletion. The upper schematic of Figure 28 shows the genomic organization of the *Rag1* and *Rag2* loci and the genomic regions corresponding to the 5' and 3' homology arms (48 kb and 15 kb, respectively; dark grey boxes). *Rag2* and *Rag* 1 each comprises a single exon denoted by the stippled grey shading. The lower schematic in Figure 28 is the LTVEC. The 5' and 3' homology arms (48 kb and 15 kb, respectively) are denoted by the dark grey boxes. The LTVEC comprises a reporter gene (lacZ) and a self-deleting cassette flanked by loxP sites (open arrows). The self-deleting cassette comprises a rat *Prm1* promoter operably linked to the Crei gene and a drug selection cassette comprising a human ubiquitin promoter operably linked to a neomycin resistance gene. Another version of the LTVEC was generated in which the neomycin resistance gene was replaced with a hygromycin resistance gene to enable retargeting of *Il2rg*-targeted rat ES cells. The Crei comprises two exons encoding the Cre recombinase are separated by an intron (Crei) to prevent its expression in a prokaryotic cell. See, for example, U.S. Patent 8,697,851 and U.S. Application Publication 2013-0312129, which describe the self-deleting cassette in detail and is hereby incorporated by reference in their entirety. By employing a rat *Prm1* promoter that drives expression of Crei specifically in male germ cells, the self-deleting cassette can be deleted from the male germ cells of F0 rats.

[00888] The LTVEC was electroporated into the rat ES cells obtained in Example 1 and the cells were plated on 15 cm 2x dense neoR MEFs in 2i + 10 uM ROCKi. The transformed rat ES cells were cultured and maintained as described in Example 1.

[00889] Colonies are screened as described elsewhere herein and targeted clones are obtained. The targeted clones are then injected into a host embryo as described elsewhere herein to produce an F0 rat.

3.4.(b)(ii): Retargeting the Rag1 and the Rag2 Locus in Rats ES Cells in which the Il2rg Locus Has Already Been Targeted

[00890] An LTVEC as in Figure 50 was prepared to target the *Rag1* and *Rag2* loci for deletion. The total length of the LTVEC was 72 kb. The LTVEC was electroporated into rat ES cells that had already been targeted for deletion of the *Il2rg* locus as in Example 3.3. Specifically, the rat ES cells were from clone *Il2rg*-CG12, for which germline transmission was confirmed in Example 3.3(a). The transformed rat ES cells were cultured and maintained as described in Example 1. Double targeted clones were screened as described elsewhere herein, and targeted clones were obtained. *Il2rg*-CG12 cells were retargeted at an efficiency of 85%, and *Il2rg* mutations were still present in the targeted clones. Electroporation was carried out as described elsewhere herein, and antibiotic selection was carried out using 1.5 ug/ml of puromycin. The targeted clones will then be injected into a host embryo as described elsewhere herein to produce an F0 rat. Retargeting is advantageous because it is faster than interbreeding *Rag1/Rag2*-targeted rats with *Il2rg*-targeted rats.

Example 4. Humanization

4.1. Humanization of Rat Genomic Loci

[00891] Humanization of rat genomic loci is carried out employing the rat ES cells described herein, which are capable of sustaining their pluripotency following one or more electroporations *in vitro*, and are capable of transmitting the targeted genetic modifications to subsequent generations. In addition, in order to circumvent the limitations of plasmids in accommodating a large genomic DNA fragment, and to overcome the low efficiency of introducing a targeted genetic modification into an endogenous locus in rat ES cells, one or more targeted genetic modifications are carried out in bacteria, *e.g.*, *E. coli*, by utilizing bacterial homologous recombination (BHR) and employing a large targeting vector (LTVEC). The LTVEC described herein, for example, includes a large fragment of an endogenous rat genomic sequence with one or more modifications or comprises an exogenous nucleic acid (*e.g.*, a homologous or orthologous human nucleic acid) flanked with rat homology arms complementary to specific genomic regions.

4.2. Humanization of Rat Immunoglobulin Loci

[00892] Humanization of an endogenous rat immunoglobulin heavy chain locus is carried out by removing one or more endogenous rat immunoglobulin heavy chain nucleic acid sequences (*e.g.*, one or more endogenous V_H gene segments, one or more human D gene segments, and one or more human J_H gene segments); and introducing into the modified immunoglobulin locus a targeting vector, e.g., a large targeting vector (LTVEC) comprising: (i) one or more unrearranged human variable region nucleic acid sequences (*e.g.*, one or more human V_H gene segments, one or more human D gene segments, and one or more human J_H gene segments), or one or more rearranged human variable region nucleic acid sequences (*e.g.*, one or more human rearranged V-D-J gene segments); (ii) a selection cassette (*e.g.,* neomycin resistance gene flanked with loxP sites); and (iii) 5' and 3' rat homology arms.

[00893] Briefly, one or more endogenous rat immunoglobulin heavy chain variable region gene segments (*i.e.*, one or more V_H gene segments, one or more human D gene segments, and one or more human J_H gene segments) in a rat BAC clone are removed or inactivated by targeting the endogenous rat immunoglobulin heavy chain locus with a selection cassette flanked by rat homology arms. More specifically, a targeting vector is constructed to contain a selection cassette (*e.g.*, a neomycin resistance gene flanked with loxP sites) flanked with 5' and 3' rat homology arms that are complementary to target rat genomic sequences (e.g., upstream and downstream rat genomic DNA sequences encompassing one or more rat V_H gene segments, one or more human D gene segments, and one or more human J_H gene segments).

[00894] Next, bacterial cells containing a large rat genomic DNA fragment encompassing a rat immunoglobulin heavy chain locus are selected and introduced with a plasmid (e.g., pABG) encoding a recombinase operably linked to a transiently inducible promoter. The targeting vector constructed above is then introduced into the recombination-competent bacterial cells. Following electroporation, the bacterial cells are treated with an inducer (e.g., arabinoside) to initiate homologous recombination between the targeting vector and the target rat genomic sequence in the BAC clone. Transformed cells are plated at a high density and subjected to drug selection to find colonies that are drug-resistant. Drug-resistant colonies are picked and screened for the targeted modification.

[00895] In order to facilitate identification of the targeted genetic modification, a high-throughput quantitative assay, namely, modification of allele (MOA) assay, is employed, which 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 can comprise a fluorescent probe that recognizes the amplified sequence. Alternatively, the quantitative assay can 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).

[00896] The bacterial cells comprising the modified rat BAC clone, i.e., a BAC clone containing a rat genomic DNA sequence wherein one or more endogenous heavy chain variable region gene segments (V_H, D, and/or J_H gene segments) have been deleted or inactivated, are then electroporated with a large targeting vector (LTVEC) comprising: (i) one or more unrearranged human variable region nucleic acid sequences (e.g., one or more unrearranged human V_H gene segments, one or more human D gene segments, and one or more human J_H gene segments), or one or more rearranged human variable region nucleic acid sequences (e.g., one or more rearranged human V-D-J gene segments).

[00897] Initiation of homologous recombination in the bacterial cells and the selection of positive clones are performed as described above. The unrearranged or rearranged human immunoglobulin heavy chain variable region nucleic acid sequences, when targeted into the endogenous immunoglobulin heavy chain locus, become operably linked to an endogenous rat immunoglobulin heavy chain constant region nucleic acid sequence. Alternatively, endogenous rat heavy chain constant region locus can be inactivated, for example, by deleting one or more rat heavy chain constant region gene segments (CH) from the endogenous heavy chain constant region locus, and can be replaced with a human heavy chain constant region nucleic acid sequence.

[00898] Likewise, humanization of an endogenous rat immunoglobulin κ or λ light chain locus is carried out by removing one or more endogenous rat immunoglobulin κ and/or λ light chain variable region nucleic acid sequences (*e.g.*, one or more endogenous rat V_κ gene segments and one or more endogenous rat J_κ gene segments); and targeting the modified immunoglobulin light chain locus with a targeting vector, *e.g.*, a large targeting vector (LTVEC), comprising: (i) one or more unrearranged human immunoglobulin light chain variable region nucleic acid sequences (*e.g.*, one or more human V_κ gene segments and one or more human J_κ gene segments), or one or more rearranged human variable region nucleic acid sequences (*e.g.*, one or more human rearranged V_κ-J_κ gene segments); (ii) a selection cassette (*e.g.,* neomycin resistance gene flanked with loxP sites); and (iii) 5' and 3' rat homology arms.

[00899] The unrearranged or rearranged human immunoglobulin light chain variable region nucleic acid sequences, when targeted into the endogenous immunoglobulin light chain locus, become operably linked to the endogenous rat immunoglobulin light chain constant region nucleic acid sequence.

[00900] The LTVEC so produced in the bacterial cells comprises, for example, an insert nucleic acid that contains a humanized rat immunoglobulin heavy chain or light chain locus in which one or more endogenous rat heavy or light chain variable region gene segments have been replaced with one or more human heavy or light chain variable region gene segments; and rat homologous arms (*e.g.*, ranging from 5 kb to 150 kb) complementary to specific genomic target sequences. The LTVEC comprising the genetic modification described above is then linearized and electroporated into the rat ES cells. Electroporated rat ES cells are plated at a high density to select drug-resistant ES cells comprising the targeting vector. The drug selection process removes the majority of the plated cells (~99%), leaving behind individual colonies, each of which is a clone derived from a single cell. Of the remaining cells, most cells (~ 80-100%) contain the targeting vector integrated at a random location in the genome. Therefore, the colonies are picked and genotyped individually in order to identify rat ES cells comprising the targeting vector at the correct genomic location (e.g., using the modification of allele (MOA) assay described above).

[00901] In order to increase the efficiency of the targeted genetic modification, the rat ES cells are electroporated with expression vectors (or mRNA) that express ZFNs 1 and 2 (or TALENs 1 and 2) together with the LTVEC. The targeting vector's homology arms lie outside the ZFN target site, therefore, the targeting vector is not cleaved by the ZFNs. The double strand break produced by the ZFNs stimulates homology-directed repair (HDR), which otherwise accounts for a very small percentage of repairs occurred normally in mammalian cells (compared to non-homologous end-joining; NHEJ).

[00902] Alternatively, expression vectors containing a type II CRISPR-associated nuclease (e.g., Cas9), a guide RNA (including CRISPR-RNA (cr-RNA) and trans-activating CRISPR RNA (tracrRNA)), as described herein, can be introduced into the bacterial cells together with the LTVEC to increase the efficiency of homologous recombination at the target genomic locus. Electroporated cells are plated at a high density and subjected to drug selection to find colonies that are drug-resistant. Drug-resistant colonies are picked and screened for the targeted modification using the modification of allele (MOA) assay as described herein. Following these procedures, improvement in the targeting efficiency can be achieved. For example, the amount of improvement can be small (e.g., improve from 10% to 15%) or large (e.g., improve from 10% to 80%).

[00903] The selected rat ES cells comprising the targeted genetic modification are then introduced into a host rat embryo, for example, a pre-morula stage or blastocyst stage rat embryo, and implanted in the uterus of a surrogate mother to generate a founder rat (F0 rat). Subsequently, the founder rat is bred to a wild-type rat to create F1 progeny heterozygous for the genetic modification. Mating of the heterozygous F1 rat can produce progeny homozygous for the genetic modification.

4.3(a). Replacing Rat IL2rg with Human IL2 Receptor Gamma

[00904] Table 29 provides a summary of the genomic organization of the rat Interleukin 2 receptor gamma locus and the positions shown were taken from build 5.0 of the Reference Sequence of the rat genome (ENSMBL). *Il2rg* is on chromosome X on the (-) strand.

[00905] Table 29. Summary of the genomic organization of the rat *Il2rg* locus

Feature	Start	End	length	Notes
Exon 1	72,021,388	72,021,516	129	contains ATG
ATG	72,017,500	72,017,502	3	start codon
Exon2	72,021,007	72,021,160	154	
ZFN1a binding site	72,021,014	72,021,028	15	CAGGCCCTGAACCGC (SEQ ID NO: 17)
ZFN1 cutting site	72,021,008	72,021,013	6	TTCTGG (SEQ ID NO: 18)
ZFN1b binding site	72,020,993	72,021,007	15	GATTACCTGCGCTGGG (SEQ ID NO: 20)
Exon3	72,020,606	72,020,790	185	
Exon4	72,020,274	72,020,413	140	
Exon5	72,019,662	72,019,824	163	
Exon6	72,019,101	72,019,197	97	
Exon7	72,018,844	72,018,910	67	
Exon8	72,017,856	72,018,506	651	contains TGA
TGA	72,018,321	72,018,323	3	stop codon
Il2rg deletion	72,018,323	72,021,502	3,180	

[00906] The lower schematic in Figure 25 is the targeting vector for the *Il2rg* 3.2 kb deletion. The targeting vector comprises a reporter gene (eGFP) operably linked to the endogenous promoter and a self-deleting cassette flanked by loxP sites (open arrows). The self-deleting cassette comprises the Crei gene operably linked to a mouse *Prm1* promoter and a selection cassette comprising a neomycin resistance gene operably linked to a human ubiquitin promoter.

[00907] The Crei gene comprises two exons encoding a Cre recombinase, which are separated by an intron (Crei) to prevent its expression in a prokaryotic cell. See, for example, U.S. Patent 8,697,851 and U.S. Application Publication 2013-0312129, which describe the self-deleting cassette in detail and are hereby incorporated by reference in their entirety. By employing the mouse *Prm1* promoter the Cre expression cassette and the drug selection cassette can be deleted specifically in male germ cells of F0 rats. The targeting vector was electroporated into the rat ES cells obtained in Example 1 and the cells were plated on 15 cm 2x dense neomycin-resistant MEFs in 2i + 10 uM ROCKi. The transformed rat ES cells were cultured, selected, and maintained as described in Example 1.

[00908] A plasmid targeting vector was constructed to replace the full-length rat interleukin 2 receptor gamma coding region with the full-length human interleukin 2

receptor gamma coding region as shown in Figure 31. The targeting vector was electroporated into the rat ES cells obtained in Example 1, and the cells were plated on 15 cm 2x dense neomycin-resistant MEFs in 2i + 10 uM ROCKi. Specifically, 4 x 10^6 cells were electroporated with 2 ug of *Il2rg* full-length humanization vector at a voltage of 400V, a capacitance of 100 uF, and a resistance of 0. Selection was done using 75 ug/mL of G418. The transformed rat ES cells were cultured, selected, and maintained as described in Example 1.

[00909] As shown in Table 44, 168 colonies were screened and 6 targeted clones were obtained. The targeting efficiency was 3.57%. One clone was injected into blastocysts as described in Example 1, and one clone producing chimeras was obtained.

[00910] Clones were injected into blastocysts as described herein in Example 1. Clones producing F0 chimeric rats were obtained. The blastocysts were transferred to pseudopregnant recipient females using standard techniques, and chimeric F0 rats were obtained. F0 rats that transmit the targeted modification through the germline are obtained.

4.3(b)(i). Replacing Rat IL2rg Ecto-Domain with Human IL2rg Ecto-Domain

[00911] The full-length humanization of IL 2 receptor gamma is useful because rats having this modified locus will produce human Il2rg; and this would allow for the detection of human Il2rg in rats with antibodies specific to human Il2rg.

[00912] The ecto-humanization (i.e., replacing the rat ecto-domain of Il2rg with the human ecto-domain of Il2rg) will result in an Il2rg polypeptide that will bind the human ligands for Il2rg, but because the cytoplasmic domain is still rat, it ecto-humanized form of Il2rg will also interact with the rat signaling machinery. Figure 33 provides a sequence alignment of the human IL-2rg protein (SEQ ID NO: 20; NP_000197.1); the rat IL-2rg protein (SEQ ID NO: 21; NP_543165.1); and the chimeric IL-2rg protein (SEQ ID NO: 22) comprising the human ecto-domain of IL-2rg fused to the remainder of the rat IL-2rg protein. The junction between the human and rat IL-2rg is noted by the vertical line.

[00913] Table 30 provides a summary of the genomic organization of the rat Interleukin 2 receptor gamma locus and the positions shown were taken from build 5.0 of

the Reference Sequence of the rat genome (ENSMBL). *Il2rg* is on chromosome X on the (-) strand. Further noted is the position of the ecto-domain of *Il2rg*.

[00914] Table 30. Summary of the genomic organization of the rat *Il2rg* locus

Feature	Start	End	Length	Notes
Exon 1	71,111,444	71,111,543	100	contains ATG
ATG	71,111,537	71,111,539	3	start codon
Exon2	71,110,897	71,111,050	154	
Exon3	71,110,504	71,110,688	185	
Exon4	71,110,156	71,110,295	140	
Exon5	71,109,228	71,109,390	163	
Exon6	71,108,599	71,108,645	47	contains transmembrane domain
Exon7	71,108,277	71,108,346	70	
Exon8	71,107,404	71,107,921	518	contains TGA
TGA	71,108,736	71,108,738	3	stop codon
full-length humanization:	71,107,404	71,111,539	4,136	(ATG to TGA plus 3' poly-A)
ecto-humanization	71,108,679	71,111,539	2,861	(ATG to beginning of transmembrane domain)

[00915] A plasmid targeting vector was constructed to replace the rat ecto-domain of the interleukin 2 receptor gamma coding region with the human ecto domain as shown in Figure 32. The targeting vector was electroporated into the rat ES cells obtained in Example 1 and the cells were plated on 15 cm 2x dense neomycin-resistant MEFs in 2i + 10 uM ROCKi. The transformed rat ES cells were cultured, selected, and maintained as described in Example 1.

[00916] As shown in Table 44, 192 colonies were screened and 13 targeted clones were obtained. The targeting efficiency was 6.77%.

[00917] Two clones were injected into blastocysts as described herein in Example 1, and two clones producing chimeras were obtained. Clones producing F0 rats were obtained. F0 rats that transmit the targeted modification through the germline are obtained.

4.3(b)(ii). Replacing Rat IL2rg Ecto-Domain with Human IL2rg Ecto-Domain Using Plasmid in Combination with CRISPR/Cas9

[00918] Table 31 shows a comparison of the results of experiments in which a version of the *Il2rg* ecto-domain humanization vector shown in Figure 32 was used alone to target the rat *Il2rg* locus or was used in combination with a CRISPR/Cas9 nuclease to target the rat *Il2rg* locus. In each experiment, electroporated cells were plated at a high density and subjected to drug selection to find colonies that were drug-resistant. Drug-resistant colonies were picked and screened for the targeted modification using the modification of allele (MOA) assay as described herein. Specifically, 4 x 10^6 cells were electroporated with 2 ug of *Il2rg* ecto-domain humanization vector at a voltage of 400V, a capacitance of 100 uF, and a resistance of 0. In the latter experiment, 6 ug of Cas9 expression plasmid and 3 ug of *Il2rg* gRNA2 or 3 ug of *Il2rg* gRNA4 were also electroporated. Selection was done using 75 ug/mL of G418. *Il2rg* gRNA2 has a sequence of GAAGCTCTTTCTATACAATCTGG (SEQ ID NO: 91) and targets a region 190 bp 3' of the rat *Il2rg* exon 1. *Il2rg* gRNA4 has a sequence of CCCCCGAAAGGAGGAGCCCTAGG (SEQ ID NO: 92) and targets a region 80 bp 5' of the rat *Il2rg* stop codon (TGA) (see Figure 49).

[00919] Table 31. Comparison of *Il2rg* Ecto-Domain Humanization Vector Targeting with and without CRISPR/Cas9

Vector	Cas9	gRNA	Colonies Screened	Targeted Clones	Targeting Efficiency
Il2rg plasmid vector	NO	NO	77	46	60%
Il2rg plasmid vector	Yes	Il2rg gRNA2	84	54	64%
Il2rg plasmid vector	Yes	Il2rg gRNA4	88	50	57%

4.4(a). Enhanced Targeting by CRISPR/Cas9 Endonucleases of Large Non-Human Animal Gene Deletions with Simultaneous Human Gene Replacements

[00920] Newly developed drugs for human disease conditions, such as fully human antibodies, are often highly specific for their targets in human cells and tissues and do not recognize the homologous targets in rodents. This high level of selectivity makes it

impossible to test the efficacy and mechanism of action of the drugs in rodents prior to their first use in humans.

[00921] A very effective solution to this problem is to create a genetically modified mouse or rat in which the human gene encoding the drug target replaces the rodent homolog. One way to create such a humanized allele in a rodent is to first delete the rodent gene in an embryonic stem (ES) cell and then, in a second gene modification event, to insert the human gene precisely at the deleted locus. The ES cells are then injected into a rodent embryo and implanted in the uterus of a surrogate mother rodent, which subsequently gives birth to genetically modified pups that carry the humanized allele.

[00922] A more efficient method of creating the humanized gene modification is to use a large targeting vector (LTVEC) that directs the simultaneous deletion of the rodent gene and replacement with its human counterpart. By employing VELOCIGENE® genetic engineering methods, such single-step humanizations can be achieved with relatively high efficiency when the rodent gene deletion and human gene insertion are smaller than about 20 kilobase pairs (kb). Larger single-step humanizations entailing deletions and replacements of greater than 100 kb are possible with LTVECs and genetic engineering methods such as VELOCIGENE® genetic engineering methods, but because of reduced targeting efficiencies sometimes encountered with very large modifications, success often requires the screening or hundreds to thousands of ES cell clones to find one that carries the desired gene modification.

[00923] To improve the efficiency of large humanizations we have developed methods that combine LTVEC gene targeting with clustered regularly interspaced short palindromic repeat RNA-guided Cas9 endonucleases (CRISPR/Cas9). CRISPR/Cas9 nucleases are ribonucleoprotein enzymes comprised of a bacterial Cas9 DNA endonuclease bound to a CRISPR RNA that guides Cas9 to cleave at a specific DNA sequence by Watson-Crick base pairing between the guide RNA and one strand of the target DNA. Because of the simplicity of the targeting mechanism, it is easy to design CRISPR/Cas9 endonucleases that direct a double strand break at nearly any genomic locus. Double strand breaks induce cellular genomic repair by the non-homologous end joining (NHEJ) pathways, which are error prone and often result in deletions or insertions

at the site of the double strand break. An alternative mechanism of repairing the double strand break is homology-directed repair (HDR) in which an endogenous or exogenous piece of DNA that shares sequence identity or similarity with the broken site seamlessly repairs the broken ends by the action of the cellular homologous recombination machinery. HDR can result in a perfect repair that restores the original sequence at the broken site, or it can be used to direct a designed modification, such as a deletion, insertion, or replacement of the sequence at the site of the double strand break. CRISPR/Cas9 nucleases can greatly enhance the rate of engineered HDR events by directing precise double strand cleavages at the sites of the intended gene modifications.

[00924] To effect a precise, single-step deletion of all or part of a rodent gene and simultaneous replacement with all or part of its human homolog, we introduced by electroporation into rodent ES cells three nucleic acid molecules: (1) an LTVEC; (2) a plasmid or mRNA encoding a Cas9 endonuclease; and (3) a plasmid encoding a CRISPR single guide RNA (sgRNA) or the sgRNA itself. The LTVEC comprised all or part of a human gene that encodes the gene product (protein or RNA) flanked by homology arms of rodent DNA designed to direct an HR event that deletes the rodent gene and inserts the human gene. The humanizing LTVEC also carried a drug selection cassette that directs the expression of an enzyme (e.g., neomycin phosphotransferase) that imparts resistance to an antibiotic drug (for example, G418). ES cells that took up the LTVEC and incorporated it into their genomes were able to grow and form colonies on a Petri dish in a growth medium containing the antibiotic drug. Because we introduced 500 to 1,000 times more CRISPR/Cas9-encoding nucleic molecules than LTVEC molecules, most of the LTVEC-containing drug resistant colonies also contained, at least transiently, the CRISPR/Cas9 components. We picked drug resistant colonies and screened them by the loss-of-allele method (Valenzuela, D. *et al.* (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, Nature Biotech. 21:652-660; Frendewey, D. *et al.* (2010) The loss-of-allele assay for ES cell screening and mouse genotyping, Methods Enzymol. 476:295-307; incorporated herein by reference in their entireties) to identify clones that had the correctly targeted humanized allele.

[00925] In one particular experiment the LTVEC was designed to create a 68 kb deletion of the mouse *Lrp5* (low-density lipoprotein receptor-related protein 5) gene and a simultaneous replacement with a 91 kb fragment of the homologous human *LRP5* gene (Figure 34). The LTVEC comprised the 91-kb fragment of the human *LRP5* gene flanked by homology arms containing 7 kb and 33 kb of genomic DNA derived from parts of the mouse *Lrp5* locus that flank the 68 kb sequence of the mouse *Lrp5* gene intended for deletion. In separate experiments, we combined the *Lrp5* humanizing LTVEC with a plasmid encoding Cas9 and a second plasmid encoding one of eight sgRNAs (gA, gB, gB2, gC, gD, gE2, gE, gF) designed to create double strand breaks within the region of the mouse *Lrp5* gene that was targeted for deletion. The sgRNAs were designed to avoid recognition of any sequence in the inserted portion of the human *LRP5* gene.

[00926] The results of the CRISPR/Cas9-assisted humanization of the *Lrp5* gene are shown in Table 32. When the LTVEC alone was introduced into ES cells, we found that 1.0% of the screened drug resistant clones carried a correctly targeted mono-allelic heterozygous humanized allele. In contrast, combining the LTVEC with Cas9 endonucleases guided by seven of the eight tested sgRNAs (sgRNA-5′A, sgRNA-5′B, sgRNA-5′B2, sgRNA-C, sgRNA-D, sgRNA-3′E2, and sgRNA-3′F; sequences provided in Table 33) produced correctly targeted monoallelic heterozygous mutations at efficiencies that ranged from 2.1 to 7.3%, representing a 2- to 9-fold enhancement of single-step humanized gene targeting compared with the unaided LTVEC. For Cas9-guided cleavage by sgRNA-5′B2, in addition to monoallelic targeting, we detected biallelic homozygous humanization at a frequency of 1%. The homozygous *Lrp5* humanized ES cells can be converted by the VELOCIMOUSE® genetic engineering method (Poueymirou, W. T. *et al.* (2007) F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses, Nature Biotech. 25:91-99, incorporated herein by reference in its entirety) directly into completely ES cell-derived mice ready for phenotypic and drug efficacy studies.

[00927] Table 32. Screening Results for CRISPR/Cas9-Assisted Humanization of the *Lrp5* Gene.

Experiment	Clones Screened	CRISPR Activity	Monoallelic Heterozygous	Biallelic Compound	Biallelic Homozygous

		(%)	Mutation Frequency (%)	Heterozygous Mutation Frequency (%)	Mutation Frequency (%)
LTVEC alone	96	N/A	1.0 (1/96)	0	0
LTVEC + Cas9 + sgRNA-5′A	96	75.6	7.3 (7/96)	0	0
LTVEC + Cas9 + sgRNA-5′B	96	79.5	4.2 (4/96)	0	0
LTVEC + Cas9 + sgRNA-5′B2	96	60.5	6.2 (6/96)	0	1.0 (1/96)
LTVEC + Cas9 + sgRNA-C	96	no assay	4.2 (4/96)	0	0
LTVEC + Cas9 + sgRNA-D	96	no assay	7.3 (7/96)	0	0
LTVEC + Cas9 + sgRNA-3′E2	96	84.5	2.1 (2/96)	0	0
LTVEC + Cas9 + sgRNA-3′E	96	52.4	0	0	0
LTVEC + Cas9 + sgRNA-3′F	96	79.8	6.2 (6/96)	0	0

[00928] Table 33. Sequences of the Guide Portions of Six sgRNAs Targeting the Mouse *Lrp5* Gene.

sgRNA	Approximate Distance from Deletion Endpoint (bp)	Guide Sequence (5′to 3′)
sgRNA-5′A	50	GGGAACCCACAGCATACTCC (SEQ ID NO: 24)
sgRNA-5′B	500	GAATCATGCACGGCTACCCC (SEQ ID NO: 25)
sgRNA-5′B2	1000	TGCTCCTATGGGGAGGCGCG (SEQ ID NO: 26)
sgRNA-C	29900/ 38430	ACTGAGATCAATGACCCCGA (SEQ ID NO: 85)
sgRNA-D	29950/ 38380	GGGTCGCCCGGAACCTCTAC (SEQ ID NO: 86)
sgRNA-3′E2	1000	CTTGGATAACATTGATACCC (SEQ ID NO: 27)
sgRNA-3′E	500	GGGGCAGAGCCCTTATATCA (SEQ ID NO: 28)
sgRNA-3′F	50	TCGCTCACATTAATCCCTAG (SEQ ID NO: 29)

[00929] The enhanced targeting of the large *Lrp5* humanization by CRISPR/Cas9 endonucleases is remarkable when compared with equivalent experiments performed with zinc finger nucleases (ZFNs). We obtained four ZFNs designed to make double strand breaks at sites within the region of the mouse *Lrp5* gene targeted for deletion (Figure 34). One ZFN targeted a sequence near the 5′end of the deletion (a), one targeted a sequence in the middle of the deletion (b), and two targeted sequences near the 3′end of the deletion (c, d). In separate experiments, we combined the *Lrp5* humanizing LTVEC

with a plasmid encoding one of the four ZFNs (a-d) designed to create double strand breaks within the region of the mouse *Lrp5* gene that were targeted for deletion. We determined that all of the ZFNs were active and able to induce NHEJ mutations in the *Lrp5* gene (data not shown), but when combined with the LTVEC, none enhanced HDR-mediated gene targeting compared with the LTVEC alone.

[00930] The enhanced targeting efficiency of the large *Lrp5* humanization by CRISPR/Cas9 endonucleases is also remarkable when compared with a series of ZFN-assisted humanization experiments. In these experiments, a series of ZFN-assisted humanizations were performed in which the mouse target gene deletions and the human gene insertions were generally of increasing size (Table 34; Figure 35). Figure 35A depicts the percent targeting efficiency of LTVECs targeting genes of increasing size for deletion. The LTVECs were used alone (gray squares) or in combination with ZFNs (black squares). Figure 35B depicts the percent targeting efficiency of LTVECs with human gene insertions of increasing size. Again, the LTVECs were used alone (gray triangles) or in combination with ZFNs (black triangles). As shown in Table 34 and Figure 35, the ability of ZFN-mediated DNA cleavage to enhance LTVEC targeting efficiency disappeared when the size of the mouse target gene deletion was greater than 24.7 kb and when the size of the human gene insertion was greater than 22.2 kb (Table 34; Figure 35A). In contrast, CRISPR/Cas9 was capable of enhancing LTVEC targeting efficiency of the *Lrp5* gene, which involved a mouse gene deletion of 68.3 kb and a human gene insertion of 91.0 kb (Table 32; Figure 34). This indicates that CRISPR/Cas9 endonucleases are able to enhance LTVEC targeting efficiency in situations where other nucleases (e.g., zinc finger nucleases) cannot.

[00931] Table 34. Summary of ZFN-Assisted Humanizations.

						Targeting Efficiency (%)		
Target Gene	**Mouse Gene Deletion (kb)**	**Human Gene Insertion (kb)**	**5′Homology Arm (kb)**	**3′Homology Arm (kb)**	**ZFN Cleavage Efficiency (%)**	**LTVEC Alone**	**LTVEC + ZFN**	**Fold Enhancement**
Fcer1a	4.1	4.2	10.9	76.8	22.9	5.20	32.81	6.3
Tlr4	7.1	5.7	67.6	85.5	12.5	5.20	22.39	4.3
Prlr	8.7	18.0	49.6	112.9	30.7	1.56	24.48	15.7
Notch4	18.7	22.2	50.1	34.9	27.1	10.41	12.50	1.2
Accn2	24.7	18.9	57.8	60.1	20.8	4.17	8.33	2.0
Adamts5	37.6	39.6	83.3	61.5	4.2	0.00	0.00	n.a.
Trpa1	45.3	53.9	41.3	57.8	8.8	0.52	0.00	0
Folh1	55.1	61.3	18.4	114.7	8.8	1.04	0.00	0

Lrp5	68.3	91.0	6.9	33.4	35.9	2.08	1.04	(0.5)
Erbb4	101.6	126.7	47.8	26.0	n.d.	0.00	0.00	n.a.

n.d. = not determined
n.a. = not applicable
( ) = targeting efficiency lower with ZFN than without

[00932] Comparable experiments were performed for humanization of other mouse genes. In one experiment, the LTVEC was designed to create a 45 kb deletion of the mouse *Trpa1* (transient receptor potential cation channel subfamily A member 1) gene and a simultaneous replacement with a 55 kb fragment of the homologous human *TRPA1* gene (Figure 36). The LTVEC comprised the 55 kb fragment of the human *TRPA1* gene flanked by homology arms containing 41 kb and 58 kb of genomic DNA derived from parts of the mouse *Trpa1* locus that flank the 45 kb sequence of the mouse *Trpa1* gene intended for deletion. In separate experiments, we combined the *Trpa1* humanizing LTVEC with a plasmid encoding Cas9 and a second plasmid encoding one of eight sgRNAs (gA, gA2, gB, gC, gD, gE, gE2, and gF) designed to create double strand breaks within the region of the mouse *Trpa1* gene that was targeted for deletion. The sgRNAs were designed to avoid recognition of any sequence in the inserted portion of the human *TRPA1* gene.

[00933] The results of the CRISPR/Cas9-assisted humanization of the *Trpa1* gene are shown in Table 35. When the LTVEC alone was introduced into ES cells, we found that 1.0% of the screened drug resistant clones carried a correctly targeted monoallelic heterozygous humanized allele. In contrast, combining the LTVEC with Cas9 endonuclease guided by six of eight tested sgRNAs (A, A2, B, C, D, and F; sequences provided in Table 43) produced correctly targeted monoallelic heterozygous mutations or biallelic compound heterozygous or homozygous mutations at efficiencies that ranged from 1.0 to 3.1%. For Cas9-guided cleavage by gRNA A and gRNA F, we detected compound heterozygous mutations at a frequency of 1.0%.

[00934] Table 35. Screening Results for CRISPR/Cas9-Assisted Humanization of the *Trpa1* Gene.

sgRNA Position	**Approximate Distance from Deletion Endpoint (bp)**	**gRNA**	**CRISPR Activity (%)**	**Clones Screened**	**Heterozygous Targeted**	**Compound Heterozygous**	**Homozygous Targeted**
5′	100	gRNA A	30.9	96	0	1	0

sgRNA Position	Approximate Distance from Deletion Endpoint (bp)	gRNA	CRISPR Activity (%)	Clones Screened	Heterozygous Targeted	Compound Heterozygous	Homozygous Targeted
5′	500	gRNA A2	no assay	96	2	0	0
5'	1000	gRNA B	42.8	96	3	0	0
middle	25600 / 19740	gRNA C	no assay	96	1	0	0
middle	26970 / 18370	gRNA D	no assay	96	2	0	0
3'	1000	gRNA E2	no assay	96	0	0	0
3′	500	gRNA E	22.6	96	0	0	0
3'	100	gRNA F	28.6	96	1	1	0
N/A	N/A	none	N/A	96	1	0	0

[00935] In another experiment, the LTVEC was designed to create a 55 kb deletion of the mouse *Folh1* (glutamate carboxypeptidase 2) gene and a simultaneous replacement with a 61 kb fragment of the homologous human *FOLH1* gene (Figure 37). The LTVEC comprised the 61 kb fragment of the human *FOLH1* gene flanked by homology arms containing 22 kb and 46 kb of genomic DNA derived from parts of the mouse *Folh1* locus that flank the 55 kb sequence of the mouse *Folh1* gene intended for deletion. In separate experiments, we combined the *Folh1* humanizing LTVEC with a plasmid encoding Cas9 and a second plasmid encoding one of six sgRNAs (gA, gA2, gC, gD, gE, and gE2) designed to create double strand breaks within the region of the mouse *Folh1* gene that was targeted for deletion. The sgRNAs were designed to avoid recognition of any sequence in the inserted portion of the human *FOLH1* gene.

[00936] The results of the CRISPR/Cas9-assisted humanization of the *Folh1* gene are shown in Table 36. When the LTVEC alone was introduced into ES cells, we found that none of the 96 screened drug resistant clones carried a correctly targeted monoallelic heterozygous humanized allele. In contrast, combining the LTVEC with Cas9 endonuclease guided by three of six tested sgRNAs (A, D, and E2; sequences provided in Table 43) produced correctly targeted monoallelic heterozygous mutations at efficiencies that ranged from 1.0 to 3.1%.

[00937] Table 36. Screening Results for CRISPR/Cas9-Assisted Humanization of the *Folh1* Gene.

sgRNA	Approximate	gRNA	CRISPR	Clones	Heterozygous	Compound	Homozygous

Position	Distance from Deletion Endpoint (bp)		Activity (%)	Screened	Targeted	Heterozygous	Targeted
5′	100	gRNA A	45.2	96	2	0	0
5'	500	gRNA A2	61.9	96	0	0	0
middle	30300 / 24800	gRNA C	7.1	96	0	0	0
middle	31290 / 23810	gRNA D	39.2	96	1	0	0
3'	500	gRNA E2	no assay	96	1	0	0
3'	100	gRNA E	1.2	96	0	0	0
N/A	N/A	none	N/A	96	0	0	0

[00938] In another experiment, the LTVEC was designed to create a 76 kb deletion of the mouse gene for complement component 5 (*C5* or *Hc*) and a simultaneous replacement with a 97 kb fragment of the homologous human *C5* gene (Figure 38). The LTVEC comprised the 97 kb fragment of the human *C5* gene flanked by homology arms containing 34.1 kb and 31.2 kb of genomic DNA derived from parts of the mouse *C5* (*Hc*) locus that flank the 76 kb sequence of the mouse *C5* (*Hc*) gene intended for deletion. In separate experiments, we combined the *C5* (*Hc*) humanizing LTVEC with a plasmid encoding Cas9 and a second plasmid encoding one of six sgRNAs (gA, gB, gC, gD, gE, and gE2) designed to create double strand breaks within the region of the mouse *C5* (*Hc*) gene that was targeted for deletion. The sgRNAs were designed to avoid recognition of any sequence in the inserted portion of the human *C5* gene.

[00939] The results of the CRISPR/Cas9-assisted humanization of the *C5* (*Hc*) gene are shown in Table 37. When the LTVEC alone was introduced into ES cells, we found that 1.0% of the screened drug resistant clones carried a correctly targeted monoallelic heterozygous humanized allele. In contrast, combining the LTVEC with Cas9 endonuclease guided by all six tested sgRNAs (A, B, C, D, E, and E2; sequences provided in Table 43) produced correctly targeted monoallelic heterozygous mutations or biallelic compound heterozygous or homozygous mutations at efficiencies that ranged from 4.2 to 16.7%. For Cas9-guided cleavage by gRNAs A and E, we detected compound heterozygous mutations at frequencies of 5.2% and 4.2%, respectively.

[00940] Table 37. Screening Results for CRISPR/Cas9-Assisted Humanization of the *C5* (*Hc*) Gene.

sgRNA Position	Approximate Distance from Deletion Endpoint	gRNA	CRISPR Activity (%)	Clones Screened	Heterozygous Targeted	Compound Heterozygous	Homozygous Targeted

	(bp)						
5′	100	gRNA A	64.3	96	11	5	0
5'	500	gRNA B	72.6	96	14	0	0
middle	38200 / 37500	gRNA C	47.6	96	11	0	0
middle	43500 / 32200	gRNA D	47.6	96	7	0	0
3'	500	gRNA E	25.0	96	0	4	0
3'	100	gRNA E2	27.4	96	6	0	0
N/A	N/A	none	N/A	96	1	0	0

[00941] In another experiment, the LTVEC was designed to create a 38 kb deletion of the mouse *Adamts5* (a disintegrin and metalloproteinase with thrombospondin motifs 5) gene and a simultaneous replacement with a 43 kb fragment of the homologous human *ADAMTS5* gene (Figure 39). The LTVEC comprised the 43 kb fragment of the human *ADAMTS5* gene flanked by homology arms containing 22 kb and 46 kb of genomic DNA derived from parts of the mouse *Adamts5* locus that flank the 38 kb sequence of the mouse *Adamts5* gene intended for deletion. In separate experiments, we combined the *Adamts5* humanizing LTVEC with a plasmid encoding Cas9 and a second plasmid encoding one of eight sgRNAs (gA, gA2, gB, gC, gD, gE, gE2, and gF) designed to create double strand breaks within the region of the mouse *Adamts5* gene that was targeted for deletion. The sgRNAs were designed to avoid recognition of any sequence in the inserted portion of the human *ADAMTS5* gene.

[00942] The results of the CRISPR/Cas9-assisted humanization of the *Adamts5* gene are shown in Table 38. When the LTVEC alone was introduced into ES cells, we found that none of the 96 screened drug resistant clones carried a correctly targeted monoallelic heterozygous humanized allele. In contrast, combining the LTVEC with Cas9 endonuclease guided by two of eight tested sgRNAs (B and F; sequences provided in Table 43) produced correctly targeted monoallelic heterozygous mutations or biallelic compound heterozygous mutations at an efficiency of 1.0%. For Cas9-guided cleavage by gRNA E2, we detected compound heterozygous mutations at a frequency of 1.0%.

[00943] Table 38. Screening Results for CRISPR/Cas9-Assisted Humanization of the *Adamts5* Gene.

sgRNA Position	Approximate Distance from Deletion Endpoint (bp)	gRNA	CRISPR Activity (%)	Clones Screened	Heterozygous Targeted	Compound Heterozygous	Homozygous Targeted
5′	100	gRNA A	85.7	96	0	0	0
5'	500	gRNA	54.8	96	0	0	0

		A2					
5′	1000	gRNA B	66.7	96	1	0	0
middle	18700 / 18950	gRNA C	9.5	96	0	0	0
middle	18800 / 18850	gRNA D	4.8	96	0	0	0
3′	1000	gRNA F	36.9	96	0	1	0
3'	500	gRNA E	54.8	96	0	0	0
3'	100	gRNA E2	54.8	96	0	0	0
N/A	N/A	none	N/A	96	0	0	0

[00944] In another experiment, the LTVEC was designed to create a 102 kb deletion of the mouse *Erbb4* (receptor tyrosine-protein kinase erbB-4) gene and a simultaneous replacement with a 127 kb fragment of the homologous human *ERBB4* gene (Figure 40). The LTVEC comprised the 127 kb fragment of the human *ERBB4* gene flanked by homology arms containing 48 kb and 26 kb of genomic DNA derived from parts of the mouse *Erbb4* locus that flank the 102 kb sequence of the mouse *Erbb4* gene intended for deletion. In separate experiments, we combined the *Erbb4* humanizing LTVEC with a plasmid encoding Cas9 and a second plasmid encoding one of eight sgRNAs (gA, gB, gB2, gC, gD, gE, gE2, and gF) designed to create double strand breaks within the region of the mouse *Erbb4* gene that was targeted for deletion. The sgRNAs were designed to avoid recognition of any sequence in the inserted portion of the human *ERBB4* gene.

[00945] The results of the CRISPR/Cas9-assisted humanization of the *Erbb4* gene are shown in Table 39. When the LTVEC alone was introduced into ES cells, we found that none of the 96 screened drug resistant clones carried a correctly targeted monoallelic heterozygous humanized allele. In contrast, combining the LTVEC with Cas9 endonuclease guided by one of eight tested sgRNAs (D; sequence provided in Table 43) produced correctly targeted monoallelic heterozygous mutations or biallelic compound heterozygous mutations at an efficiency of 1.0%. For Cas9-guided cleavage by gRNA D, we detected compound heterozygous mutations at a frequency of 1%.

[00946] Table 39. Screening Results for CRISPR/Cas9-Assisted Humanization of the *Erbb4* Gene.

sgRNA Position	Approximate Distance from	gRNA	CRISPR Activity	Clones Screened	Heterozygous Targeted	Compound Heterozygous	Homozygous Targeted

	Deletion Endpoint (bp)		(%)				
5′	100	gRNA A	25.0	96	0	0	0
5'	500	gRNA B	no assay	96	0	0	0
5′	1000	gRNA B2	47.6	96	0	0	0
middle	50200 / 51350	gRNA C	20.2	96	0	0	0
middle	50230 / 51320	gRNA D	42.8	96	0	1	0
3′	1000	gRNA F	15.5	96	0	0	0
3'	500	gRNA E	89.2	96	0	0	0
3'	100	gRNA E2	14.3	96	0	0	0
N/A	N/A	none	N/A	96	0	0	0

[00947] In another experiment, the LTVEC was designed to create a 110 kb deletion of the mouse *Ror1* (tyrosine-protein kinase transmembrane receptor ROR1) gene and a simultaneous replacement with a 134 kb fragment of the homologous human *ROR1* gene (Figure 41). The LTVEC comprised the 134 kb fragment of the human *ROR1* gene flanked by homology arms containing 41.8 kb and 96.4 kb of genomic DNA derived from parts of the mouse *Ror1* locus that flank the 110 kb sequence of the mouse *Ror1* gene intended for deletion. In separate experiments, we combined the *Ror1* humanizing LTVEC with a plasmid encoding Cas9 and a second plasmid encoding one of six sgRNAs (gA, gB, gC, gD, gE, and gF) designed to create double strand breaks within the region of the mouse *Ror1* gene that was targeted for deletion. The sgRNAs were designed to avoid recognition of any sequence in the inserted portion of the human *ROR1* gene.

[00948] The results of the CRISPR/Cas9-assisted humanization of the *Ror1* gene are shown in Table 40. When the LTVEC alone was introduced into ES cells, we found that none of the 96 screened drug resistant clones carried a correctly targeted monoallelic heterozygous humanized allele. In contrast, combining the LTVEC with Cas9 endonuclease guided by two of six tested sgRNAs (D and F; sequences provided in Table 43) produced correctly targeted monoallelic heterozygous or biallelic mutations at efficiencies of 1.0%. For Cas9-guided cleavage by gRNA F, we also detected compound heterozygous mutations at a frequency of 1%.

[00949] Table 40. Screening Results for CRISPR/Cas9-Assisted Humanization of the *Ror1* Gene.

sgRNA Position	Approximate Distance from Deletion Endpoint (bp)	gRNA	CRISPR Activity (%)	Clones Screened	Heterozygous Targeted	Compound Heterozygous	Homozygous Targeted
5′	200	gRNA A	no assay	96	0	0	0
5'	1000	gRNA B	no assay	96	0	0	0
middle	54300 / 55500	gRNA D	no assay	96	1	0	0
middle	54500 / 55300	gRNA C	no assay	96	0	0	0
3′	1000	gRNA E	no assay	96	0	0	0
3'	200	gRNA F	no assay	96	0	1	0
N/A	N/A	none	N/A	96	0	0	0

[00950] In another experiment, the LTVEC was designed to create a 79 kb deletion of the mouse *Dpp4* (dipeptidyl peptidase 4) gene and a simultaneous replacement with an 82 kb fragment of the homologous human *DPP4* gene (Figure 42). The LTVEC comprised the 82 kb fragment of the human *DPP4* gene flanked by 5' and 3' homology arms, each containing 46 kb of genomic DNA derived from parts of the mouse *Dpp4* locus that flank the 79 kb sequence of the mouse *Dpp4* gene intended for deletion. In separate experiments, we combined the *Dpp4* humanizing LTVEC with a plasmid encoding Cas9 and a second plasmid encoding one of eight sgRNAs (gA, gB, gB2, gC, gD, gE, gE2, and gF) designed to create double strand breaks within the region of the mouse *Dpp4* gene that was targeted for deletion. The sgRNAs were designed to avoid recognition of any sequence in the inserted portion of the human *DPP4* gene.

[00951] The results of the CRISPR/Cas9-assisted humanization of the *Dpp4* gene are shown in Table 41. When the LTVEC alone was introduced into ES cells, we found that 2.1% of the screened drug resistant clones carried a correctly targeted monoallelic heterozygous humanized allele. In contrast, combining the LTVEC with Cas9 endonuclease guided by any one of eight tested sgRNAs (A, B, B2, C, D, E, E2, and F; sequences provided in Table 43) produced correctly targeted monoallelic heterozygous mutations at efficiencies that ranged from 2.1 to 7.3%.

[00952] Table 41. Screening Results for CRISPR/Cas9-Assisted Humanization of the *Dpp4* Gene.

sgRNA Position	Approximate Distance from Deletion Endpoint (bp)	gRNA	CRISPR Activity (%)	Clones Screened	Heterozygous Targeted	Compound Heterozygous	Homozygous Targeted
5′	50	gRNA A	no assay	96	7	0	0
5'	400	gRNA B	no assay	96	2	0	0
5′	900	gRNA B2	no assay	96	5	0	0
middle	38800 / 40200	gRNA C	no assay	96	3	0	0
middle	40800 / 38100	gRNA D	no assay	96	3	0	0
3′	900	gRNA E2	no assay	96	2	0	0
3'	500	gRNA E	no assay	96	6	0	0
3'	200	gRNA F	no assay	96	5	0	0
N/A	N/A	none	N/A	96	2	0	0

[00953] A table summarizing the results for CRISPR/Cas9-assisted humanization of the various mouse genes is provided in Table 42. The first row indicates the gene locus being targeted. The second row indicates the deletion size (Del) of the endogenous mouse locus and the insertion size (Ins) of the corresponding human locus. The remaining rows show the number of colonies (out of 96) for each condition that had correctly targeted monoallelic heterozygous mutations, biallelic compound heterozygous mutations, or biallelic homozygous mutations. "No gRNA" represents LTVEC alone, whereas the other rows represent LTVEC plus corresponding gRNAs (indicated by relative position within the deletion locus).

[00954] Table 42. Summary of CRISPR/Cas9-Assisted Humanization of Mouse Genes.

	Lrp5	*Trpa1*	*Folh1*	*C5 (Hc)*	*Adamts5*	*Erbb4*	*Ror1*	*Dpp4*
Del/Ins (kb)	68/91	45/55	55/61	76/97	38/43	102/127	110/134	79/82
Most 5'	7	1	2	16	0	0	0	7
5'	4	2	0	14	0	0	0	2
5'	7	3	N/A	N/A	1	0	N/A	5
Middle	4	1	0	11	0	0	1	3
Middle	7	2	1	7	0	1	0	3
3'	2	0	N/A	N/A	1	0	N/A	2

3'	0	0	1	4	0	0	0	6
Most 3'	6	2	0	6	0	0	1	5
No gRNA	1	1	0	1	0	0	0	2

[00955] Table 43. Guide RNA Sequences Used for CRISPR/Cas9-Assisted Humanization of Mouse Genes.

gRNA	Guide Sequence (5′to 3′)	SEQ ID NO
Trpa1 gRNA A	GTACTGGGGAATCGGTGGTC	30
Trpa1 gRNA A2	CACGCACTCCAAATTTATCC	31
Trpa1 gRNA B	CTAAGTGTGTATCAGTACAT	32
Trpa1 gRNA C	TGCCCTGCACAATAAGCGCA	33
Trpa1 gRNA D	ACTCATTGAAACGTTATGGC	34
Trpa1 gRNA E2	AGTAAGGGTGGATTAAATTC	35
Trpa1 gRNA E	GCCATCTAGATTCATGTAAC	36
Trpa1 gRNA F	GACTAGAAATGTTCTGCACC	37
Folh1 gRNA A	TGAACCAATTGTGTAGCCTT	38
Folh1 gRNA A2	AATAGTGGTAAAGCACCATG	39
Folh1 gRNA B	GTGTGCTAAGGATCGAAGTC	40
Folh1 gRNA C	CACCGAGATGCTTGGGTATT	41
Folh1 gRNA D	TGTAACCGCCCTGAATGACC	42
Folh1 gRNA E	AAAAGGGCATCATAAATCCC	43
Folh1 gRNA E2	TCAAAAATAGTCATACACCT	44
Folh1 gRNA F	GGTCTCTAGTACATTGTAGA	45
C5 (*Hc*) gRNA A	ATCACAAACCAGTTAACCGG	46
C5 (*Hc*) gRNA B	TTTCAGACGAGCCGACCCGG	47
C5 (*Hc*) gRNA B2	CTGTCAACAGTGCCGCGTTT	48
C5 (*Hc*) gRNA C	TGTGTGTCATAGCGATGTCG	49
C5 (*Hc*) gRNA D	AACAGGTACCCTATCCTCAC	50
C5 (*Hc*) gRNA E2	TCGTGGTTGCATGCGCACTG	51
C5 (*Hc*) gRNA E	GGCCCGGACCTAGTCTCTCT	52
C5 (*Hc*) gRNA F	AGTCTGTAAAGTTAGCAGTC	53
Adamts5 gRNA A	GGTGGTGGTGCTGACGGACA	54
Adamts5 gRNA A2	TATGAGATCAACACTCGCTA	55
Adamts5 gRNA B	CCAAGGACTTCCCCACGTTA	56
Adamts5 gRNA C	TGCTTCCCTTATGCAAGATT	57
Adamts5 gRNA D	TTAGGTACCCTATTTGAATA	58
Adamts5 gRNA E2	TGCAGTGGGTGACAGGTCCA	59
Adamts5 gRNA E	AGGGTTATACTGACGTTGTG	60
Adamts5 gRNA F	TGTCTTTCAAGGAGGGCTAC	61

Erbb4 gRNA A	TGATGTGCAGTCAGACAAAG	62
Erbb4 gRNA B	TGCACTATGGTTGACTATGA	63
Erbb4 gRNA B2	GGAATATTCTAATAGGAAGT	64
Erbb4 gRNA C	AAGTGCTGTACCATTCTAGC	65
Erbb4 gRNA D	TAATCAATAGACAACCTCGT	66
Erbb4 gRNA E2	TCATTCCTAATGGTATTATA	67
Erbb4 gRNA E	AGGGTACATAGATGGCATCG	68
Erbb4 gRNA F	CTCTTTAACAATTACCACTT	69
Ror1 gRNA A	TGTGGGCCTTTGCTGATCAC	70
Ror1 gRNA B	AATCTATGATCCTATGGCCT	71
Ror1 gRNA D	TGCCAATAGCAGTGACTTGA	72
Ror1 gRNA C	GGGAAGAATGGGCTATTGTC	73
Ror1 gRNA E	GGTTGTTTGTGCTGATGACG	74
Ror1 gRNA F	CCGTCCTAGGCCTTCTACGT	75
Dpp4 gRNA A	ACTAGTAGACCTGAGGGGTT	76
Dpp4 gRNA B	GCTCCAGTGTTTAGGCCTTG	77
Dpp4 gRNA B2	GGCAAGCTGAAAACGCATGC	78
Dpp4 gRNA C	GTAGATCGCTTTCCACTACC	79
Dpp4 gRNA D	GAACTCCACTGCTCGTGAGC	80
Dpp4 gRNA E2	ATAGGTGGGCACTATTGAAG	81
Dpp4 gRNA E	ATGGGAAGGTTTATACCAGC	82
Dpp4 gRNA F	CGGTGTAAAAACAACGGGAA	83

Example 5. Summary of Targeted Modification of Rat Genomic Loci

[00956] Table 44. Summary of rat targeting with various vector types and nuclease agents discussed in Examples 3 and 4.

[00957] Table 44. Rat Targeting Summary

Example #	Locus	Vector	Colonies screened	Targeted Clones	Targeting efficiency	Biallelic targeted	Biallelic efficiency	Clones Injected	Clones producing chimeras	Clones transmitting through germline	Notes
3.2(a)(ii)	ApoE	plasmid	384	23	5.99%	0	0	3	3	1	
3.2(a)(iii)	ApoE + ZFN	plasmid	384	290	75.52%	8	2.08%	2	2	1	These 2 clones are biallelic targeted
3.3(a)	Il2rg	plasmid	232	5	2.16%	N/A	N/A	6	5	1	
3.2(b)(ii)	ApoE LTVEC	LTVEC	288	8	2.78%	1	0.35%	3	1	0	
3.2(b)(iii)	ApoE LTVEC + ZFN	LTVEC	288	16	5.56%	1	0.35%	1	N/A	0	This clone is biallelic targeted
3.2(b)(iv)	ApoE LTVEC	LTVEC	75	32	42.67%	0	0				
3.2(b)(iv)	ApoE LTVEC + CRISPR/Cas9	LTVEC	169	84	50%	5	3%	0	0	0	
4.3(a)	Il2rg Humanization 1	plasmid	168	6	3.57%	N/A	N/A	1	1	0	replaces entire rat Il2rg with human Il2rg
4.3(b)(i)	Il2rg Humanization 2	plasmid	192	13	6.77%	N/A	N/A	2	2	0	replaces rat Il2rg ecto-domain with human Il2rg ecto-domain
4.3(b)(ii)	Il2rg Humanization 2	plasmid	77	46	59.74%						
4.3(b)(ii)	Il2rg Humanization 2 + CRISPR/Cas9	plasmid	172	104	60.47%	N/A	N/A	0	0	0	replaces rat Il2rg ecto-domain with human Il2rg ecto-domain
3.4(a)(i)	Rag2	LTVEC	270	0	0	0	0	0	0	0	Predicted 5.7 KB deletion
3.4(a)(ii)	Rag2 LTVEC	LTVEC	36	0	0	0	0				
3.4(a)(ii)	Rag2 LTVEC + CRISPR/Cas9	LTVEC	39	6	15.38%	1	2.5%	1	1	0	
3.4(b)(i)	Rag1-2	LTVEC	256	1	0.39%	0	0	1	1	0	Predicted 16.2 kb deletion
3.4(b)(ii)	Rag1-2	LTVEC	94	80	85%	0	0	0	0	0	Rag1-2 LTVEC was electroporated into Il2rg-CG12 mutant clone (retargeting)

[00958] Table 45 shows a summary of the targeting of rat ES cells with either plasmids or LTVECs in combination with CRISPR/Cas9. Two gRNAs were tested separately for each targeted locus: *Rag2*, *ApoE*, and *Il2rg*. The cleavage efficiency of CRISPR/Cas9 was > 20% at all three loci. Increased targeting efficiency and increased biallelic targeting was observed when CRISPR/Cas9 was used in combination with the targeting plasmids and LTVECs.

[00959] Table 45. Summary of Rat ES Cell Targeting with Plasmids or LTVECs in Combination with CRISPR/Cas9

Condition	**Targeting efficiency**	**Biallelic Targeting**
Rag2 (LTVEC)	0	0
Rag2 (LTVEC+ CRISPR)	6 - 22%	0 - 4%
ApoE (LTVEC)	43%	0
ApoE (LTVEC + CRISPR)	47 - 53%	1 - 4%
Il2rg Humanization (plasmid vector)	60%	N/A (X-linked)
Il2rg Humanization (plasmid + CRISPR)	57 – 64%	N/A (X-linked)

[00960] Table 46 shows a summary of germline transmission data for targeted modification of rat genomic loci. Germline transmission was confirmed for *ApoE*-targeted rats and *Il2rg*-targeted rats. The rat ES cells were XY (male) and were heterozygous targeted. Therefore, when the targeted ES cells contribute to the germline, approximately 50% of the sperm derived from the ES cells will carry the mutant allele and will produce heterozygous F1 pups.

[00961] Table 46. Germline Transmission Data for Targeted Modification of Rat Genomic Loci

Targeted Gene	Clones Microinjected	Clones Producing Chimeras	Clones Achieving Germline Transmission	Germline Pups/ Total Pups	Heterozygous F1 Pups*	Heterozygous Pups in Which Neo Cassette Was Deleted
ApoE	3	3	1	7/79 (9%)	4	4
Il2rg	5	5	1	11/257 (5%)	5	5

Example 6. Generating, Maintaining, and Targeting Human Induced Pluripotent Stem Cells

6.1. Generation of Human iPS Cells

[00962] This example describes the generation of human iPS cells from non-pluripotent human cells. PiggyBac (System Biosciences) vectors (PB-600A_CAGGS Bst XI (0.64 μg/μL) and PB-200 (0.99 μg/μL) comprising the genes that encode four reprogramming factors (hOct4, hSox2, hKLF-4, hMYC) operably linked to a CM7 promoter were introduced into neonatal human foreskin fibroblasts using RED and BLUE GeneIn™ transfection reagents (GlobalStem). The transfected cells were incubated on NuFF1 feeder cells in E7 medium to allow for incorporation of the vectors and expression of the reprogramming factors. E7 medium comprised DMEM/F-12, $NaHCO_3$, L-ascorbic acid, insulin, transferrin, selenium, and FGF-2.

[00963] Puromycin selection began 10 days after transfection using 2 μg/mL puromycin in E7 medium. At day 21, colonies were selected and cultured in mTeSR™ medium, which comprised DMEM/F-12, $NaHCO_3$, L-ascorbic acid, insulin, transferrin, selenium, FGF-2, TGF-β1, glutathione, L-glutamine, defined lipids, thiamine, trace elements B and C, β-mercaptoethanol, bovine serum albumin, pipecolic acid, lithium chloride, and GABA. At days 29 to 57, cells were propagated and passaged in mTeSR™ medium until reaching ~50% confluent in 6 well plates. At days 65 to 73, propagation and passage continued using mTeSR™ medium and Gentle Cell Dissociation Reagent (Stem Cell Technologies). At day 76, medium was changed to low osmolality VG2i medium for further propagation, passage, and maintenance of the cells comprising naïve or naïve-looking hiPSCs.

6.2. LTVEC Targeting in Human iPS Cells

[00964] This example describes the use of LTVEC targeting in human iPS cells. As shown in Figure 51, we introduced by electroporation into human iPS cells propagated in VG2i medium the following nucleic acid molecules: (1) an LTVEC (0.67 μg); (2) a plasmid encoding a Cas9 endonuclease (5 μg); and (3) a plasmid encoding a CRISPR single guide RNA (gRNA) (10 μg). In one set of samples, the Cas9 and gRNA were excluded. Specifically, 3 x 10^6 cells were electroporated at a voltage of 700V, a capacitance of 25 uF, and a resistance of 400 ohms. The LTVEC comprised a 16.7 kb

nucleic acid comprising mouse *Adam6a* and *Adam6b* genes flanked by homology arms containing 34 kb and 105 kb of genomic DNA derived from genomic regions that flank the 4.1 kb sequence of the human *ADAM6* locus intended for deletion. The LTVEC also carried a drug selection cassette that directs the expression of an enzyme that imparts resistance to an antibiotic drug (hygromycin). The human ADAM6 gRNA used had the following sequence: GTATAGCCCTGTTACACATT (SEQ ID NO: 94).

[00965] Cells that took up the LTVEC and incorporated it into their genomes were able to grow and form colonies on a GELTREX™-coated tissue culture dish in a growth medium containing the antibiotic drug. Because we introduced 500 to 1,000 times more CRISPR/Cas9-encoding nucleic molecules than LTVEC molecules, most of the LTVEC-containing drug resistant colonies also contained, at least transiently, the CRISPR/Cas9 components. We picked drug resistant colonies and screened them by the loss-of-allele method (Valenzuela *et al.* (2003) *Nat. Biotech.* 21:652-660; Frendewey *et al.* (2010) *Methods Enzymol.* 476:295-307; incorporated herein by reference in their entireties) to identify clones that had the correctly targeted allele.

[00966] The results of the CRISPR/Cas9-assisted LTVEC targeting of the ADAM6 locus are shown in Table 47.

Table 47. CRISPR/Cas9-assisted LTVEC targeting

	Targeting Efficiency
LTVEC Only	3.1%
LTVEC + CRISPR	7.3%

[00967] When the LTVEC alone was introduced into human iPS cells, a targeting efficiency of 3.1% was observed. In contrast, combining the LTVEC with Cas9 guided by the ADAM6 gRNA resulted in a targeting efficiency of 7.3%.

6.3. Effect of Low Osmolality Medium on Human iPS Cell Morphology

[00968] This example describes the effect of salt concentration, ionic strength, and/or osmolality on the pluripotency state of human iPS cells in culture. Human iPS cells were cultured on a MATRIGEL™ or GELTREX™ substrate in a medium described in Table 48 or in mTeSR™ -hLIF medium.

Table 48. Medium for iPS cell culture.

Component	Amount (v/v)
Base Medium	24.75
F-12 Medium	24.75
N2® Supplement	0.5
Neurobasal® Medium	49
B-27® Supplement	1
Penicillin/Streptomycin	1
L-Glutamine (200 mM)	1
2-Mercaptoethanol (55 mM)	0.1836
hLIF (1 x 10^4 units/mL)	0.001
CHIR99021 (10 mM)	0.03
PD0325901 (10 mM)	0.005

[00969] When the base medium used was DMEM, this medium was referred to as 2i medium. When the base medium used was VG-DMEM, this low osmolality medium was referred to as VG2i medium. The osmolality of VG2i medium (233 mOsm/kg) is lower than the osmolality of traditional 2i medium (261 mOsm/kg).

[00970] As shown in Figure 52, human iPS cells cultured on MATRIGEL™ in 2i medium for a period of 8 days (Figure 52A) or 12 days (Figure 52B) displayed a morphology characteristic of iPS cells in a primed state, particularly growth in an epithelial monolayer and the appearance of apico-basal polarity.

[00971] mTeSR-hLIF medium and VG2i medium were further evaluated for their effects on the morphology and pluripotency state of human iPS cells. In this study, human iPS cells were cultured on MATRIGEL™ or NuFF feeder cells in mTeSR™ -hLIF medium (Figures 53A and 53C) or in VG2i medium (Figures 53B and 53D) for a period of 6 days. When cultured in mTeSR™ -hLIF medium on MATRIGEL™ or NuFF feeder cells, human iPS cells displayed a morphology characteristic of a primed pluripotency state, particularly growth in an epithelial monolayer and the appearance of apico-basal polarity. Some cells cultured in mTeSR™ -hLIF medium began to display a morphology characterized by three-dimensional clumping. By contrast, when cultured in VG2i medium on MATRIGEL™ or NuFF feeder cells, the human iPS cells displayed a morphology characteristic of a naïve pluripotency state, particularly growth in round, dome-shaped colonies and a lack of apico-basal polarity.

[00972]

6.4. Effect of Low Osmolality Medium on the Expression of Pluripotency Markers in Human iPS Cells

[00973] This example describes the effect of salt concentration, ionic strength, and/or osmolality on the expression of pluripotency markers in human iPS cells that have been reprogrammed from a primed state to a naïve state. Following 24 days of culture in VG2i medium on a MATRIGEL™ substrate, reprogrammed naïve human iPS cells were stained for the expression of alkaline phosphatase or NANOG. It was observed that the reprogrammed cells strongly expressed both alkaline phosphatase (Figure 54A) and NANOG (Figures 54B and 54C), which are indicative of a naïve pluripotency state.

6.5. Effect of Low Osmolality Medium on Enzymatic Dissociation and Subculture of Human iPS Cells

[00974] In this example, human iPS cells that were reprogrammed to a naïve state using low osmolality VG2i medium were enzymatically dissociated using trypsin to create a single cell suspension (Figure 55A). The cell suspension was passaged onto new GELTREX™-coated plates for subculture in VG2i medium. It was observed after 1 day (Figure 55B) and 4 days (Figure 55C) that the subcultured cells continued to display a morphology characteristic of cells in a naïve pluripotency state. Particularly, the cells grew as rounded dome-shaped colonies and did not exhibit an apico-basal polarity. It was notable that enzymatic dissociation could be performed in the absence of a ROCK inhibitor, which is typically necessary to prevent activation of pro-apoptotic pathways. This suggests that pro-apoptotic pathways are not as strongly activated during enzymatic dissociation and subculture in naïve human iPS cells cultured under the conditions identified herein.

[00975] All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference. Unless otherwise apparent from the context of any embodiment, aspect, step or feature of the invention can be used in

combination with any other. Reference to a range includes any integers within the range, any subrange within the range. Reference to multiple ranges includes composites of such ranges.

CLAIMS

1. An *in vitro* method for modifying a genome at a genomic locus of interest in a non-human mammalian embryonic stem (ES) cell that is not a mouse ES cell, comprising:

contacting the non-human mammalian ES cell with a Cas9 protein, a CRISPR RNA that hybridizes to a CRISPR target sequence at the genomic locus of interest, a tracrRNA, and a large targeting vector (LTVEC) that is at least 10 kb in size and comprises an insert nucleic acid flanked by:

(i) a 5' homology arm that is homologous to a 5' target sequence at the genomic locus of interest; and

(ii) a 3' homology arm that is homologous to a 3' target sequence at the genomic locus of interest,

wherein following contacting the non-human mammalian ES cell with the Cas9 protein, the CRISPR RNA, and the tracrRNA in the presence of the LTVEC, the genome of the non-human mammalian ES cell is modified to comprise a targeted genetic modification comprising deletion of a region of the genomic locus of interest wherein the deletion is at least 30 kb and/or insertion of the insert nucleic acid at the genomic locus of interest wherein the insertion is at least 30 kb.

2. The method of claim 1, wherein the CRISPR RNA and the tracrRNA are introduced as a single nucleic acid molecule comprising the CRISPR RNA and the tracrRNA.

3. The method of claim 2, wherein the single nucleic acid molecule comprises the CRISPR RNA and the tracrRNA fused together in the form of a single guide RNA (sgRNA).

4. The method of claim 1, wherein the CRISPR RNA and the tracrRNA are introduced separately.

5. The method of any preceding claim, wherein:

(a) the Cas9 protein is introduced in the form of a protein, a messenger RNA (mRNA) encoding the Cas9 protein, or a DNA encoding the Cas9 protein;

(b) the CRISPR RNA is introduced in the form of an RNA or a DNA

encoding the CRISPR RNA; and

(c) the tracrRNA is introduced in the form of an RNA or a DNA encoding the tracrRNA.

6. The method of claim 5, wherein the Cas9 protein, the CRISPR RNA, and the tracrRNA are introduced as a protein-RNA complex.

7. The method of claim 5, wherein:

(a) the DNA encoding the Cas9 protein is in the form of a first expression construct comprising a first promoter operably linked to a first nucleic acid encoding the Cas9 protein;

(b) the DNA encoding the CRISPR RNA is in the form of a second expression construct comprising a second promoter operably linked to a second nucleic acid encoding the CRISPR RNA; and

(c) the DNA encoding the tracrRNA is in the form of a third expression construct comprising a third promoter operably linked to a third nucleic acid encoding the tracrRNA;

wherein the first, second, and third promoters are active in the non-human mammalian ES cell, and

wherein the first, second, and third expression constructs are on a single nucleic acid molecule or on multiple nucleic acid molecules.

8. The method of claim 5, wherein:

(a) the DNA encoding the Cas9 protein is in the form of a first expression construct comprising a first promoter operably linked to a first nucleic acid encoding the Cas9 protein; and

(b) the DNA encoding the CRISPR RNA and the DNA encoding the tracrRNA are in the form of a second expression construct comprising a second promoter operably linked to a second nucleic acid encoding a gRNA comprising the CRISPR RNA and the tracrRNA;

wherein the first and second promoters are active in the non-human mammalian ES cell, and

wherein the first and second expression constructs are on a single nucleic acid molecule or on separate nucleic acid molecules.

9. The method of any preceding claim, wherein the targeted genetic modification comprises simultaneous deletion of an endogenous nucleic acid sequence at the genomic locus of interest and insertion of the insert nucleic acid at the genomic locus of interest.

10. The method of claim 9, wherein the deleted endogenous nucleic acid sequence is from 30 kb to about 110 kb, and the insert nucleic acid is from about 40 kb to about 140 kb.

11. The method of any preceding claim, wherein the targeted genetic modification is a biallelic genetic modification.

12. The method of claim 11, wherein the biallelic genetic modification comprises deletion of an endogenous nucleic acid sequence and insertion of the insert nucleic acid at the genomic locus of interest in two homologous chromosomes.

13. The method of claim 11, wherein the modified non-human mammalian ES cell is compound heterozygous or hemizygous at the genomic locus of interest.

14. The method of claim 13, wherein the targeted genetic modification at the genomic locus of interest in one chromosome comprises deletion of an endogenous nucleic acid sequence and insertion of the insert nucleic acid.

15. The method of claim 14, wherein the targeted genetic modification comprises: (1) deletion of an endogenous nucleic acid sequence at the genomic locus of interest in first and second homologous chromosomes; and (2) insertion of the insert nucleic acid into the genomic locus of interest in the first homologous chromosome and disruption of the genomic locus of interest in the second homologous chromosome.

16. The method of any preceding claim, wherein the LTVEC is at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb, or at least 200 kb; or

wherein the targeted genetic modification comprises deletion of a region of the

genomic locus of interest wherein the deletion is at least 30 kb, and the LTVEC is at least 15 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb, or at least 200 kb.

17. The method of any preceding claim, wherein the targeted genetic modification comprises insertion of the insert nucleic acid, wherein the insert nucleic acid is at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb, at least 200 kb, at least 250 kb, or at least 300 kb; or

wherein the targeted genetic modification further comprises deletion of a region of the genomic locus of interest wherein the deletion is at least 30 kb, and the insert nucleic acid is at least 10 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb, at least 200 kb, at least 250 kb, or at least 300kb.

18. The method of any preceding claim, wherein the insert nucleic acid is from about 40 kb to about 140 kb.

19. The method of any preceding claim, wherein the CRISPR target sequence is immediately flanked by a Protospacer Adjacent Motif (PAM) sequence.

20. The method of any preceding claim, wherein the sum total of the 5' and the 3' homology arms of the LTVEC is from 10 kb to about 20 kb, from about 20 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 120 kb, or from about 120 kb to 150 kb.

21. The method of any preceding claim, wherein the targeted genetic modification comprises:

(a) replacement of an endogenous nucleic acid sequence with a homologous or an orthologous nucleic acid sequence;

(b) deletion of an endogenous nucleic acid sequence;

(c) deletion of an endogenous nucleic acid sequence,

wherein the deletion ranges from at least 30 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about

100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 300 kb, from about 300 kb to about 400 kb, from about 400 kb to about 500 kb, from about 500 kb to about 1 Mb, from about 1 Mb to about 1.5 Mb, from about 1.5 Mb to about 2 Mb, from about 2 Mb to about 2.5 Mb, or from about 2.5 Mb to about 3 Mb;

(d) insertion of an exogenous nucleic acid sequence;

(e) insertion of an exogenous nucleic acid sequence ranging from at least 30 kb to about 40 kb, from about 40 kb to about 60 kb, from about 60 kb to about 80 kb, from about 80 kb to about 100 kb, from about 100 kb to about 150 kb, from about 150 kb to about 200 kb, from about 200 kb to about 250 kb, from about 250 kb to about 300 kb, from about 300 kb to about 350 kb, or from about 350 kb to about 400 kb;

(f) insertion of an exogenous nucleic acid sequence comprising a homologous or an orthologous nucleic acid sequence;

(g) insertion of a chimeric nucleic acid sequence comprising a human and a non-human nucleic acid sequence;

(h) insertion of a conditional allele flanked by site-specific recombinase target sequences;

(i) insertion of a selectable marker or a reporter gene operably linked to a promoter active in the non-human mammalian ES cell; or

(j) a combination thereof.

22. The method of any preceding claim, wherein the targeted genetic modification comprises deletion of a region of the genomic locus of interest,

wherein the deletion is at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least about 80 kb, at least 90 kb, at least 100 kb, at least 150 kb, at least 200 kb, at least 250 kb, or at least 300 kb; or

wherein the targeted genetic modification further comprises insertion of the insert nucleic acid at the genomic locus of interest wherein the insertion is at least 30 kb, and the deletion is at least 10 kb, at least 20 kb, at least 30 kb, at least 40 kb, at least 50 kb, at least 60 kb, at least 70 kb, at least 80 kb, at least 90 kb, at least 100 kb, at least 150 kb, at least 200 kb, at least 250 kb, or at least 300 kb.

23. The method of any preceding claim, wherein the region of the genomic

locus of interest deleted is from about 30 kb to about 110 kb.

24. The method of any preceding claim, wherein the targeted genetic modification comprises deletion of a region of the genomic locus of interest wherein the deletion is at least 30 kb and insertion of the insert nucleic acid at the genomic locus of interest wherein the insertion is at least 30 kb.

25. The method of any preceding claim, wherein the genomic locus of interest is endogenous to the non-human mammalian ES cell.

26. The method of any one of claims 1-24, wherein the genomic locus of interest comprises a heterologous or exogenous segment of DNA that was integrated into the genome of the non-human mammalian ES cell.

27. The method of any preceding claim, wherein the genomic locus of interest is an immunoglobulin locus.

28. The method of claim 27, wherein the immunoglobulin locus encodes a mammalian immunoglobulin heavy chain variable region amino acid sequence.

29. The method of claim 27, wherein the immunoglobulin locus encodes a mammalian immunoglobulin light chain variable region amino acid sequence.

30. The method of claim 29, wherein the immunoglobulin locus comprises an unrearranged mammalian λ and/or κ light chain variable region nucleic acid sequence.

31. The method of claim 29, wherein the immunoglobulin locus comprises a rearranged mammalian λ and/or κ light chain variable region nucleic acid sequence.

32. The method of any one of claims 1-26, wherein the genomic locus of interest is a T cell receptor locus.

33. The method of claim 32, wherein the T cell receptor locus is a T cell receptor alpha locus.

34. The method of any one of claims 1-26, wherein the genomic locus of

interest comprises an Interleukin-2 receptor gamma locus, an *ApoE* locus, a *Rag1* locus, a *Rag2* locus, both of the *Rag1* and the *Rag2* loci, an *Adamts5* locus, a *Trpa1* locus, a *Folh1* locus, an *Erbb4* locus, a *Lrp5* locus, a *C5* (*Hc*) locus, a *Ror1* locus, or a *Dpp4* locus.

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

36. The method of claim 35, wherein the insert nucleic acid comprises one or more functional human V_H gene segments comprising V_H1-2, V_H1-3, V_H1-8, V_H1-18, V_H1-24, V_H1-45, V_H1-46, V_H1-58, V_H1-69, V_H2-5, V_H2-26, V_H2-70, V_H3-7, V_H3-9, V_H3-11, V_H3-13, V_H3-15, V_H3-16, V_H3-20, V_H3-21, V_H3-23, V_H3-30, V_H3-30-3, V_H3-30-5, V_H3-33, V_H3-35, V_H3-38, V_H3-43, V_H3-48, V_H3-49, V_H3-53, V_H3-64, V_H3-66, V_H3-72, V_H3-73, V_H3-74, V_H4-4, V_H4-28, V_H4-30-1, V_H4-30-2, V_H4-30-4, V_H4-31, V_H4-34, V_H4-39, V_H4-59, V_H4-61, V_H5-51, V_H6-1, V_H7-4-1, V_H7-81, or a combination thereof.

37. The method of claim 35 or 36, wherein the insert nucleic acid comprises one or more functional human D gene segments comprising D1-1, D1-7, D1-14, D1-20, D1-26, D2-2, D2-8, D2-15, D2-21, D3-3, D3-9, D3-10, D3-16, D3-22, D4-4, D4-11, D4-17, D4-23, D5-12, D5-5, D5-18, D5-24, D6-6, D6-13, D6-19, D6-25, D7-27, or a combination thereof.

38. The method of any one of claims 35-37, wherein the insert nucleic acid comprises one or more functional J_H gene segments comprising J_H1, J_H2, J_H3, J_H4, J_H5, J_H6, or a combination thereof.

39. The method of any one of claims 1-34, wherein the insert nucleic acid comprises a genomic nucleic acid sequence that encodes a human immunoglobulin light chain variable region amino acid sequence.

40. The method of claim 39, wherein the insert nucleic acid comprises one or more human Vκ gene segments comprising Vκ4-1, Vκ5-2, Vκ 7-3, Vκ 2-4, Vκ1-5, Vκ1-6, Vκ3-7, Vκ1-8, Vκ1-9, Vκ2-10, Vκ3-11, Vκ1-12, Vκ1-13, Vκ2-14, Vκ3-15, Vκ1-16, Vκ1-17, Vκ2-18, Vκ2-19, Vκ3-20, Vκ6-21, Vκ1-22, Vκ1-23, Vκ2-24, Vκ3-25, Vκ2-26, Vκ1-27, Vκ2-28, Vκ2-29,

Vκ2-30, Vκ3-31, Vκ1-32, Vκ1-33, Vκ3-34, Vκ1-35, Vκ2-36, Vκ1-37, Vκ2-38, Vκ1-39, Vκ2-40, or a combination thereof.

41. The method of claim 39 or 40, wherein the insert nucleic acid comprises one or more human Vλ gene segments comprising Vλ3-1, Vλ4-3, Vλ2-8, Vλ3-9, Vλ3-10, Vλ2-11, Vλ3-12, Vλ2-14, Vλ3-16, Vλ2-18, Vλ3-19, Vλ3-21, Vλ3-22, Vλ2-23, Vλ3-25, Vλ3-27, or a combination thereof.

42. The method of any one of claims 39-41, wherein the insert nucleic acid comprises one or more human Jκ gene segments comprising Jκ1, Jκ2, Jκ3, Jκ4, Jκ5, or a combination thereof.

43. The method of any one of claims 1-34, wherein the insert nucleic acid comprises a polynucleotide encoding at least a region of a human T cell receptor.

44. The method of claim 43, wherein the T cell receptor is a T cell receptor alpha.

45. The method of any one of claims 1-34, wherein the insert nucleic acid comprises at least one disease allele.

46. The method of claim 45, wherein the insert nucleic acid comprises at least one human disease allele of a human gene.

47. The method of any preceding claim, wherein the targeted genetic modification results in a humanized genomic locus comprising: (a) an insertion of a homologous or orthologous human nucleic acid sequence; (b) a replacement of an endogenous nucleic acid sequence with a homologous or orthologous nucleic acid sequence; or (c) a combination thereof.

48. The method of claim 47, wherein the targeted genetic modification comprises a replacement of an endogenous non-human mammalian nucleic acid sequence with a homologous or orthologous human nucleic acid sequence.

49. The method of any preceding claim, wherein the LTVEC is from 20 kb to 30 kb, from 30 kb to 40 kb, from 40 kb to 50 kb, from 50 kb to 75 kb, from 75 kb to 100 kb,

from 100 kb to 125 kb, from 125 kb to 150 kb, from 150 kb to 175 kb, from 175 kb to 200 kb, from 200 kb to 225 kb, or from 225 kb to 250 kb in length,

wherein the sum total of the 5' and 3' homology arms is from 10 kb to 20 kb, from 20 kb to 30 kb, from 30 kb to 40 kb, from 40 kb to 50 kb, from 50 kb to 60 kb, from 60 kb to 70 kb, from 70 kb to 80 kb, from 80 kb to 90 kb, from 90 kb to 100 kb, from 100 kb to 110 kb, from 110 kb to 120 kb, from 120 kb to 130 kb, from 130 kb to 140 kb, from 140 kb to 150 kb, from 150 kb to 160 kb, from 160 kb to 170 kb, from 170 kb to 180 kb, from 180 kb to 190 kb, or from 190 kb to 200 kb in length, and

wherein the targeted genetic modification comprises deletion of a region of the genomic locus of interest wherein the deletion is from 30 kb to 110 kb and insertion of the insert nucleic acid at the genomic locus of interest wherein the insertion is from 40 kb to 140 kb.

50. The method of any preceding claim, further comprising:

(b) identifying the modified non-human mammalian ES cell comprising the targeted genetic modification at the genomic locus of interest;

(c) introducing the modified non-human mammalian ES cell into a non-human mammalian host embryo; and

(d) gestating the non-human mammalian host embryo in a surrogate mother,

wherein the surrogate mother produces an F0 generation non-human mammal comprising the targeted genetic modification at the genomic locus of interest.

51. The method of any preceding claim, wherein the non-human mammal is a rodent.

52. The method of claim 51, wherein the rodent is a rat.