NOVEL CHO INTEGRATION SITES AND USES THEREOF

BACKGROUND

Cross Reference to Related Applications

[0001] This application claims the benefit of priority of U.S. Provisional Application No. 62/067,774 filed October 23, 2014, the entire contents of which are incorporated herein by reference.

Incorporation by Reference of Sequence Listing

[0002] The Sequence Listing in an ASCII text file, named 32353_T0045US01_ SequenceListing.txt of 28 KB, created on October 20, 2015, and submitted to the United States Patent and Trademark Office via EFS-Web, is incorporated herein by reference.

Field of the Invention

[0003] The invention provides for stable integration and/or expression of recombinant proteins in eukaryotic cells. In particular, the invention includes methods and compositions for improved expression of proteins in eukaryotic cells, particularly Chinese hamster (*Cricetulus griseus*) cell lines, by employing expression-enhancing nucleotide sequences. The invention includes polynucleotides and modified cells that facilitate recombination-mediated cassette exchange (RMCE). The methods of the invention integrate exogenous nucleic acids at specific chromosomal loci in the Chinese hamster cellular genome in order to facilitate enhanced and stable expression of recombinant proteins by the modified cells.

Description of Related Art

[0004] Cellular expression systems aim to provide a reliable and efficient source for the manufacture of a given protein, whether for research or therapeutic use. Recombinant protein expression in mammalian cells is a preferred method for manufacturing therapeutic proteins due to, for example, the ability of mammalian expression systems to appropriately post-translationally modify recombinant proteins.

[0005] Several cellular systems are available for expression of proteins, each containing various combinations of *cis*- and, in some cases, *trans*-regulatory elements to achieve high levels of recombinant protein with short incubation times. Despite the availability of numerous systems, the challenge of efficient gene transfer and stability of the integrated gene for expression of a recombinant protein still exists. Multiple local genetic factors will

determine not only when the target gene of interest is to be expressed, but whether the cell can functionally drive the transcription of the gene towards a productive output, or whether the expression will even be sustained long-term. Chromosomal integration sites, *e.g.* Chinese hamster ovary cell (CHO) integration sites and locus control regions within or adjacent to specific genes have been characterized in the art (WO2012/138887A1; Li, Q. et al., 2002 *Blood*. 100:3077-3086). As such, targeted regulatory regions are typically identified in a region coding for endogenous proteins. However, for long-term expression of a target transgene, a key consideration is minimal disruption of cellular genes to avoid changes in the phenotype of the cell line.

[0006] Engineering stable cell lines to accommodate additional genes for expression, such as additional antibody chains as in multispecific antibodies, is particularly challenging. Wide variations in expression levels of integrated genes may occur. Integrating additional genes may lead to greater variation in expression and instability due to the local genetic environment (*i.e.,* position effects). Accordingly, there is a need in the art for improved mammalian expression systems.

BRIEF SUMMARY

[0007] In one aspect, the invention provides a cell comprising an exogenous nucleic acid sequence integrated at a specific site within a locus, wherein the locus comprises a nucleotide sequence that is at least 90% identical to SEQ ID NO:1 or SEQ ID NO: 4. In some embodiments, the locus comprises a nucleotide sequence that is at least 90% identical to SEQ ID NO:1. In some embodiments, the locus comprises a nucleotide sequence that is at least 90% identical to SEQ ID NO:4.

[0008] In another aspect, the invention provides a polynucleotide comprising a first nucleic acid sequence integrated into a specific site within a second nucleic acid sequence (*e.g.* a locus of the invention). In one embodiment, the second nucleic acid sequence comprises the nucleotide sequence of SEQ ID NO:1. In another embodiment, the second nucleic acid sequence comprises the nucleotide sequence of SEQ ID NO:4.

[0009] In one embodiment, the second nucleic acid sequence is an expression-enhancing sequence selected from a nucleotide sequence having at least 90% nucleic acid identity to SEQ ID NO:1, or an expression-enhancing fragment thereof. In one embodiment, the second nucleic acid sequence is an expression-enhancing sequence selected from a nucleotide sequence having at least 90% nucleic acid identity to SEQ ID NO:4, or an expression-enhancing fragment thereof. In another embodiment, the expression-enhancing sequence is capable of enhancing expression of a protein encoded by an exogenous nucleic acid sequence. In another embodiment, the expression-

enhancing sequence is capable of enhancing expression of a protein encoded by an exogenous nucleic acid sequence at least about 1.5-fold to at least about 3-fold enhancement in expression compared to expression typically observed by random integration into a genome.

[0010] In another embodiment, the exogenous nucleic acid sequence is integrated into a specific site at any position within SEQ ID NO:1 or SEQ ID NO:4.

[0011] In some embodiments, the specific site at a position within SEQ ID NO:1 or adjacent to a position within SEQ ID NO:1 is selected from the group consisting of nucleotides spanning positions numbered 10-4,000; 100-3,900; 200-3,800; 300-3,700; 400-3,600; 500-3,500; 600-3,400; 700-3,300; 800-3,200; 900-3,100; 1,000-3,000; 1,100-2,900; 1,200-2,800; 1,300-2,700; 1,200-2,600; 1,300-2,500; 1,400-2,400; 1,500-2,300; 1,600-2,200; 1,700-2100; 1,800-2050; 1850-2050, 1,900-2040; 1950-2,025, 1990-2021, 2002-2021 and 2,010-2,015 of SEQ ID NO:1. In certain embodiments, the specific site at a position within SEQ ID NO:1 or adjacent to a position within SEQ ID NO:1 is selected from the group consisting of nucleotides spanning positions numbered 1990-1991, 1991-1992, 1992-1993, 1993-1994, 1995-1996, 1996-1997, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2003-2004, 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2019-2020, and 2020-2021 of SEQ ID NO:1.

[0012] In another embodiment, the specific site at a position within SEQ ID NO:1 or adjacent to a position within SEQ ID NO:1 is selected from the group consisting of nucleotides spanning positions numbered 10-500; 500-1,000; 500-2,100; 1,000-1,500; 1,000-2,100; 1,500-2,000; 1,500-2,500; 2,000-2,500; 2,500-3,000; 2,500-3,500; 3,000-3,500; 3,000-4,000; and 3,500-4,000 of SEQ ID NO:1. In certain embodiments, the exogenous nucleic acid sequence is integrated at, within or near any one or more of the specific sites described above.

[0013] In another embodiment, the exogenous nucleic acid sequence comprises a recognition site positioned within an expression-enhancing sequence as described above, providing that the expression-enhancing sequence comprises a sequence that is at least about 90% identical, at least about 91% identical, at least about 92% identical, at least about 93% identical, at least about 94% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, or at least about 99% identical to the expression-enhancing sequence of SEQ ID NO:1 or SEQ ID NO:4, an expression-enhancing fragment thereof.

[0014] In one embodiment, the exogenous nucleic acid sequence comprises a recombinase recognition site. In some embodiments, the exogenous nucleic acid

sequence further comprises at least one recombinase recognition site comprising a sequence independently selected from a *Lox*P site, a *Lox*511 site, a *Lox*2272 site, *Lox*2372, *Lox*5171, *Lox*m2, *Lox*71, *Lox*66, *Lox*Fas and a frt site. In one embodiment, the recombinase recognition site is integrated within the expression-enhancing sequence. In another embodiment, the recombinase recognition site is immediately adjacent in the 5' direction to the terminal nucleotide of the 5' end of a gene cassette, or immediately adjacent in the 3' direction to the terminal nucleotide of the 3' end of a gene cassette. In some embodiments, the at least one recombinase recognition site and gene cassette are integrated within the expression-enhancing sequence.

[0015] In one embodiment, at least two recombinase recognition sites are present within the expression-enhancing sequence. In another embodiment, two recombinase recognition sites of opposite orientation are integrated within the expression-enhancing sequence. In another embodiment, three recombinase recognition sites are integrated within the expression-enhancing sequence.

[0016] In one aspect, an isolated Chinese hamster ovary (CHO) cell is provided that comprises an engineered expression-enhancing sequence of SEQ ID NO:1 or an expression-enhancing fragment thereof. In one embodiment, the expression-enhancing sequence comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:4, or a stable variant thereof, is engineered to integrate an exogenous nucleic acid sequence as described above. In other embodiments, the invention provides an isolated CHO cell comprising an exogenous nucleic acid sequence inserted into a locus comprising an expression-enhancing sequence of SEQ ID NO:1, or SEQ ID NO:4, or a stable variant thereof.

[0017] In one embodiment, the CHO cell further comprises at least one recombinase recognition sequence within the expression-enhancing sequence. In another embodiment, the at least one recombinase recognition sequence is independently selected from a *Lox*P site, a *Lox*511 site, a *Lox*2272 site, *Lox*2372, *Lox*5171, *Lox*m2, *Lox*71, *Lox*66 *Lox*Fas and a frt site. In another embodiment, the recombinase recognition site is immediately adjacent in the 5' direction to the terminal nucleotide of the 5' end of a gene cassette, or immediately adjacent in the 3' direction to the terminal nucleotide of the 3' end of a gene cassette. In some embodiments, the at least one recombinase recognition site and gene cassette are integrated within the expression-enhancing sequence of the CHO cell genome described herein.

[0018] In another embodiment, the at least one recombination recognition site is positioned as described above, with the caveat that the gene cassette comprises an expression-enhancing sequence comprising at least 90% identity, at least about 91%

identity, at least about 92% identity, at least about 93% identity, at least about 94% identity, at least about 95% identity, at least about 96% identity, at least about 97% identity, at least about 98% identity, or at least about 99% identity to nucleotides 1001 through 2001 of SEQ ID NO:1 (SEQ ID NO:2) or an expression-enhancing fragment thereof. In another embodiment, the at least one recombination recognition site is positioned as described above, with the caveat that the gene cassette comprises an expression-enhancing sequence comprising at least 90% identity, at least about 91% identity, at least about 92% identity, at least about 93% identity, at least about 94% identity, at least about 95% identity, at least about 96% identity, at least about 97% identity, at least about 98% identity, or at least about 99% identity to nucleotides 2022 through 3022 of SEQ ID NO:1 (SEQ ID NO:3) or an expression-enhancing fragment thereof.

[0019] In yet another embodiment, the at least one recombinase recognition site is inserted in the CHO cell genome at or within nucleotides 1990-1991, 1991-1992, 1992-1993, 1993-1994, 1995-1996, 1996-1997, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2003-2004, 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2019-2020, 2020-2021 or 2021-2022 of SEQ ID NO:1.

[0020] In another embodiment, the exogenous nucleic acid is inserted in the CHO genome at or within nucleotides 1990-1991, 1991-1992, 1992-1993, 1993-1994, 1995-1996, 1996-1997, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2003-2004, 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2019-2020, 2020-2021 or 2021-2022 of SEQ ID NO:1.

[0021] In another embodiment, the exogenous nucleic acid is inserted in the CHO genome at or within nucleotides 2001-2022 of SEQ ID NO:1. In some embodiments, the exogenous nucleic acid is inserted at or within nucleotides 2001-2002 or nucleotides 2021-2022 of SEQ ID NO:1 and nucleotides 2002-2021 of SEQ ID NO:1 are deleted, as a result of the insertion. Likewise, the exogenous nucleic acid is inserted in the CHO genome at or within nucleotides 9302-9321 of SEQ ID NO:4. In some embodiments, the exogenous nucleic acid is inserted at or within nucleotides 9301-9302 or nucleotides 9321-9322 of SEQ ID NO:4 and nucleotides 9302-9321 of SEQ ID NO:4 are deleted, as a result of the insertion.

[0022] In some embodiments, the exogenous nucleic acid sequence integrated at a specific site within a locus, such as the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:4, comprises a gene of interest (GOI) (e.g., a nucleotide sequence encoding a protein

of interest or "POI"). In certain embodiments, the exogenous nucleic acid sequence comprises one or more genes of interest. In some embodiments, the one or more genes of interest are selected from the group consisting of a first GOI, a second GOI and a third GOI.

[0023] In some embodiments, the exogenous nucleic acid sequence integrated at a specific site within a locus, such as the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:4, comprises a GOI and at least one recombinase recognition site. In one embodiment, a first GOI is inserted within the expression-enhancing sequence of SEQ ID NO:1 or SEQ ID NO:4, or the expression-enhancing sequence having at least 90% nucleotide identity to SEQ ID NO:1 or SEQ ID NO:4, or the expression-enhancing fragment thereof, as described above, and the first GOI is optionally operably linked to a promoter, wherein the promoter-linked GOI (or the GOI) is flanked 5' by a first recombinase recognition site and 3' by a second recombinase recognition site. In another embodiment, a second GOI is inserted 3' of the second recombinase recognition site, and the second GOI is flanked 3' by a third recombinase recognition site.

[0024] In yet another embodiment, the GOI is operably linked to a promoter capable of driving expression of the GOI, wherein the promoter comprises a eukaryotic promoter that can be regulated by an activator or inhibitor. In other embodiments, the eukaryotic promoter is operably linked to a prokaryotic operator, and the eukaryotic cell optionally further comprises a prokaryotic repressor protein.

[0025] In another embodiment, one or more selectable markers are included between the first and the second and/or the second and the third recombinase recognition sites. In some embodiments, the first and/or the second genes of interest and/or the one or more selectable markers are operably linked to a promoter, wherein the promoter may be the same or different. In another embodiment, the promoter comprises a eukaryotic promoter (such as, for example, a CMV promoter or an SV40 late promoter), optionally controlled by a prokaryotic operator (such as, for example, a tet operator). In other embodiments, the cell further comprises a gene encoding a prokaryotic repressor (such as, for example, a tet repressor).

[0026] In another embodiment, the cell further comprises a gene capable of expressing a recombinase. In some embodiments, the recombinase is a Cre recombinase.

[0027] In one aspect, a CHO host cell is provided, comprising an expression-enhancing sequence selected from SEQ ID NO:1 or SEQ ID NO:4, or an expression-enhancing sequence having at least 90% nucleotide identity to SEQ ID NO:1 or SEQ ID NO:4, or an expression-enhancing fragment thereof, comprising a first recombinase recognition site followed by a first eukaryotic promoter, a first selectable marker gene, a second

eukaryotic promoter, a second selectable marker gene, and a second recombinase recognition site. In more embodiments, the CHO host cell further provides a third eukaryotic promoter, a third marker gene, and a third recombinase recognition site. In one embodiment, the expression-enhancing sequence is within SEQ ID NO:1 or SEQ ID NO:4 as described above.

[0028] In one embodiment, the first, second, and third recombinase recognition sites are different from each other. In some embodiments, the recombinase recognition sites are selected from a *Lox*P site, a *Lox*511 site, a *Lox*2272 site, *Lox*2372, *Lox*5171, *Lox*m2, *Lox*71, *Lox*66, *Lox*Fas and a frt site.

[0029] In one embodiment, the first selectable marker gene is a drug resistance gene. In another embodiment, the drug resistance gene is a neomycin resistance gene or a hygromycin resistance gene. In another embodiment, the second and third selectable marker genes encode two different fluorescent proteins. In one embodiment, the two different fluorescent proteins are selected from the group consisting of *Discosoma* coral (DsRed), green fluorescent protein (GFP), enhanced green fluorescent protein (eGFP), cyano fluorescent protein (CFP), enhanced cyano fluorescent protein (eCFP), yellow fluorescent protein (YFP), enhanced yellow fluorescent protein (eYFP) and far-red fluorescent protein (*e.g.* mKate, mKate2, mPlum, mRaspberry or E2-crimson).

[0030] In one embodiment, the first, second, and third promoters are the same. In another embodiment, the first, second, and third promoters are different from each other. In another embodiment, the first promoter is different from the second and third promoters, and the second and third promoters are the same. In more embodiments, the first promoter is an SV40 late promoter, and the second and third promoters are each a human CMV promoter. In other embodiments, the first and second promoters are operably linked to a prokaryotic operator.

[0031] In one embodiment, the host cell line has an exogenously added gene encoding a recombinase integrated into its genome, operably linked to a promoter. In another embodiment, the recombinase is Cre recombinase. In another embodiment, the host cell has a gene encoding a regulatory protein integrated into its genome, operably linked to a promoter. In more embodiments, the regulatory protein is a tet repressor protein.

[0032] In one embodiment, the first GOI and the second GOI encode a light chain, or fragment thereof, of an antibody or a heavy chain, or fragment thereof, of an antibody. In another embodiment, the first GOI encodes a light chain of an antibody and the second GOI encodes a heavy chain of an antibody.

[0033] In certain embodiments, the first, second and third GOI encode a polypeptide selected from the group consisting of a first light chain, or fragment thereof, a second light

chain, or fragment thereof and a heavy chain, or fragment thereof. In yet another embodiment, the first, second and third GOI encode a polypeptide selected from the group consisting of a light chain, or fragment thereof, a first heavy chain, or fragment thereof and a second heavy chain, or fragment thereof.

[0034] In one aspect, a method is provided for making a protein of interest, comprising (a) introducing into a CHO host cell a gene of interest (GOI), wherein the GOI integrates into a specific locus comprising a nucleotide sequence that is at least 90% identical to SEQ ID NO:1 or SEQ ID NO:4; (b) culturing the cell of (a) under conditions that allow expression of the GOI; and (c) recovering the protein of interest. In one embodiment, the protein of interest is selected from the group consisting of a subunit of an immunoglobulin, or fragment thereof, and a receptor, or ligand-binding fragment thereof. In certain embodiments, the protein of interest is selected from the group consisting of an antibody light chain, or antigen-binding fragment thereof, and an antibody heavy chain, or antigen-binding fragment thereof.

[0035] In some embodiments, the GOI is introduced into the cell employing a targeting vector for recombinase-mediated cassette exchange (RMCE) and the CHO host cell genome comprises at least one exogenous recognition sequence within the specific locus. In other embodiments, the CHO host cell genome comprises at least one exogenous recognition sequence and a selectable marker, optionally linked to a promoter, IRES and/or polyadenylation (polyA) sequence, within the specific locus.

[0036] In certain embodiments, the CHO host cell genome comprises one or more recombinase recognition sites as described above, and the GOI is introduced into the specific locus through the action of a recombinase that recognizes the recombinase recognition site.

[0037] In another embodiment, the GOI is introduced into the cell employing a targeting vector for homologous recombination, and wherein the targeting vector comprises a 5' homology arm homologous to a sequence present in the specific locus, a GOI, and a 3' homology arm homologous to a sequence present in the specific locus. In another embodiment, the targeting vector further comprises two, three, four, or five or more genes of interest. In another embodiment, one or more of the genes of interest are operably linked to a promoter.

[0038] In another aspect, a targeting vector is provided wherein the targeting vector comprises a 5' homology arm homologous to a sequence present in a locus comprising a nucleotide sequence that is at least 90% identical to SEQ ID NO:1 or SEQ ID NO:4, a GOI, and a 3' homology arm homologous to a sequence present in a locus comprising a nucleotide sequence that is at least 90% identical to SEQ ID NO:1 or SEQ ID NO:4. In

another embodiment, the targeting vector further comprises two, three, four, or five or more genes of interest.

[0039] In another aspect, a method is provided for modifying a CHO cell genome to integrate an exogenous nucleic acid sequence, comprising the step of introducing into the cell a vehicle that includes a vector, wherein the vector comprises an exogenous nucleic acid sequence wherein the exogenous nucleic acid integrates within a locus of the genome comprising a nucleotide sequence at least 90% identical to SEQ ID NO: 1 or SEQ ID NO: 4.

[0040] In some embodiments, the vector comprises a 5' homology arm homologous to a sequence present in a locus of the genome comprising a nucleotide sequence at least 90% identical to SEQ ID NO: 1 or SEQ ID NO: 4, an exogenous nucleic acid sequence, and a 3' homology arm homologous to a sequence present in a locus of the genome comprising a nucleotide sequence at least 90% identical to SEQ ID NO: 1 or SEQ ID NO: 4.

[0041] In some embodiments, the exogenous nucleic acid sequence in the vector comprises

one or more recognition sequences. In other embodiments, the exogenous nucleic acid comprises one or more GOIs, such as a selectable marker or a nucleic acid encoding a POI. In still other embodiments, the exogenous nucleic acid comprises one or more GOIs and one or more recognition sequences.

[0042] In one embodiment, the vehicle comprises at least one additional vector or mRNA. In another embodiment, the additional vector is chosen from the group consisting of an adenovirus, a lentivirus, a retrovirus, an adeno-associated virus, an integrating phage vector, a non-viral vector, a transposon and/or transposase, an integrase substrate, and a plasmid. In some embodiments, the additional vector comprises a nucleotide sequence encoding a site-specific nuclease for integrating the exogenous nucleic acid sequence.

[0043] In certain embodiments, the site-specific nuclease comprises a zinc finger nuclease (ZFN), a ZFN dimer, a transcription activator-like effector nuclease (TALEN), a TAL effector domain fusion protein, or an RNA-guided DNA endonuclease.

[0044] In another aspect, a vehicle is provided for modifying a CHO cell genome to integrate an exogenous nucleic acid sequence, wherein the vehicle includes a vector, wherein the vector comprises a 5' homology arm homologous to a sequence present in a locus of the genome comprising a nucleotide sequence at least 90% identical to SEQ ID NO: 1 or SEQ ID NO: 4, an exogenous nucleic acid sequence, and a 3' homology arm homologous to a sequence present in a locus of the genome comprising a nucleotide sequence at least 90% identical to SEQ ID NO: 1 or SEQ ID NO: 4.

[0045] In some embodiments, the exogenous nucleic acid sequence comprises one or more recognition sequences. In other embodiments, the exogenous nucleic acid comprises one or more GOIs, such as a selectable marker or a nucleic acid encoding a POI. In still other embodiments, the exogenous nucleic acid comprises one or more GOIs and one or more recognition sequences.

[0046] In yet another aspect, a method is provided for modifying a CHO cell genome to express a therapeutic agent comprising a vehicle for introducing, into the genome, an exogenous nucleic acid comprising a sequence for expression of the therapeutic agent, wherein the vehicle comprises a 5' homology arm homologous to a sequence present in the nucleotide sequence of SEQ ID NO:1, a nucleic acid encoding the therapeutic agent, and a 3' homology arm homologous to a sequence present in the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:4.

[0047] In one more aspect, the invention provides a modified CHO host cell comprising a modified CHO genome wherein the CHO genome is modified by insertion of an exogenous recognition sequence within a locus of the genome having a nucleotide sequence at least 90% identical to SEQ ID NO: 1.

[0048] In another aspect, the invention provides a modified eukaryotic host cell comprising a modified eukaryotic genome wherein the eukaryotic genome is modified at a target integration site in a non-coding region of the genome to insert an exogenous nucleic acid. In some embodiments, the exogenous nucleic acid is a recognition sequence. In other embodiments, the host cell is a mammalian host cell, such as a CHO cell. In other embodiments, the target integration site comprises an expression-enhancing sequence such as SEQ ID NO:1, provided that the sequence does not code for any endogenous proteins. The invention also provides methods of making such a modified eukaryotic host cell.

[0049] In any of the aspects and embodiments described above, the expression-enhancing sequence can be placed in the indicated orientation as in SEQ ID NO:1, or in the reverse of the orientation of SEQ ID NO:1.

[0050] Any of the aspects and embodiments of the invention can be used in conjunction with any other aspect or embodiment of the invention, unless otherwise specified or apparent from the context.

[0051] Other objects and advantages will become apparent from a review of the ensuing detailed description.

BRIEF DESCRIPTION OF THE FIGURES

[0052] Figures 1A and 1B. Figure 1A: Schematic diagram of an operable construct

utilizing random introduction of a nucleic acid molecule expressing a GOI (for example, a multi-chain antibody) and multiple copies of a selection marker into a cell genome, for example a CHO genome for identifying a target locus. The exemplified construct includes: Heavy chain (HC); First copy selection marker, such as: hygromycin resistance gene (Hyg); First copy Light Chain (LC); Second copy selection marker (*e.g.* Hyg), Second copy Light Chain (LC); Third copy selection marker (*e.g.* Hyg). Figure 1B: Example donor vector for integration via homologous recombination into the native locus identified as SEQ ID NO:1. 5' and 3' homology arms are derived from SEQ ID NO:1.

[0053] Figures 2A through 2C illustrate that the locus of SEQ ID NO:1 (*LOCUS 1*), operably linked to a gene of interest (GOI), exhibits enhanced mRNA expression of the GOI compared to the same GOI that is not operably linked to *LOCUS 1*, instead linked to a Control Locus. Figure 2A: Equivalent number of gene copies exhibited for cells encoding an antibody gene of interest, *i.e.* one heavy chain (HC) and two light chains (LC) operably linked to the Control Locus vs. *LOCUS 1*. Figure 2B: mRNA levels are higher for GOI expressed in *LOCUS 1* compared to Control Locus mRNA. Figure 2C: Protein titer is 3-fold higher for cells expressing the GOI in *LOCUS 1* compared to protein titer produced from the cells expressing the same GOI in the Control Locus.

[0054] Figures 3A and 3B illustrate an example cassette comprising a fluorescent marker and a GOI integrated at *LOCUS 1* (*e.g.* mKate flanked by lox sites to be exchanged with eYFP and a GOI) compared to the same cassette integrated at a Control Locus (exchanged with a different fluorescent marker, *e.g.* dsRed2, flanked by lox sites), wherein such integration employs Cre recombinase and recombinase-mediated cassette exchange (RMCE). Such cassettes were used in experiments to measure recombination efficiency and transcription of the GOI.

[0055] Figure 4 shows a higher mRNA level of a gene of interest (GOI) as measured in a CHO cell pool expressing the GOI in *LOCUS 1* (SEQ ID NO:1) compared to mRNA from a CHO cell pool expressing the same GOI, under the same regulatory conditions, but integrated within the control locus, *i.e.* EESYR.

DETAILED DESCRIPTION

[0056] Before the present methods are described, it is to be understood that this invention is not limited to particular methods, and experimental conditions described, as such methods and conditions may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0057] As used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural references unless the context clearly dictates otherwise. Thus for example, a reference to "a method" includes one or more methods, and/or steps of the type described herein and/or which will become apparent to those persons skilled in the art upon reading this disclosure.

[0058] Unless defined otherwise, or otherwise specified, 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.

[0059] Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, particular methods and materials are now described. All publications mentioned herein are incorporated herein by reference in their entirety.

Definitions

[0060] DNA regions are operably linked when they are functionally related to each other. For example, a promoter is operably linked to a coding sequence if the promoter is capable of participating in the transcription of the sequence; a ribosome-binding site is operably linked to a coding sequence if it is positioned so as to permit translation. Generally, operably linked can include, but does not require, contiguity. In the case of sequences such as secretory leaders, contiguity and proper placement in a reading frame are typical features. An expression-enhancing sequence of the locus of interest is operably linked to a gene of interest (GOI) where it is functionally related to the GOI, for example, where its presence results in enhanced expression and/or stable integration of the GOI.

[0061] The term "enhanced" when used to describe enhanced expression includes an enhancement of at least about 1.5-fold to at least about 3-fold enhancement in expression over what is typically observed by random integration of an exogenous sequence into a genome or by integration at a different locus, for example, as compared to a pool of random integrants of a single copy of the same expression construct. Fold-expression enhancement observed employing the sequences of the invention is in comparison to an expression level of the same gene, measured under substantially the same conditions, in the absence of a sequence of the invention, for example in comparison to integration at another locus into the same species genome. Enhanced recombination efficiency includes an enhancement of the ability of a locus to recombine (for example, employing recombinase-recognition sites). Enhancement refers to an efficiency of recombination over random recombination for example, without employing

recombinase-recognition sites or the like, which is typically 0.1%. A preferred enhanced recombination efficiency is about 10-fold over random, or about 1%. Unless specified, the claimed invention is not limited to a specific recombination efficiency.

[0062] Where the phrase "exogenously added gene" or "exogenously added nucleic acid" is employed with reference to a locus of interest, the phrase refers to any DNA sequence or gene not present within the locus of interest as the locus is found in nature. For example, an "exogenously added gene" within a CHO locus (*e.g.,* a locus comprising a sequence of SEQ ID NO:1), can be a hamster gene not found within the particular CHO locus in nature (*i.e.,* a hamster gene from another locus in the hamster genome), a gene from any other species (*e.g.*, a human gene), a chimeric gene (*e.g.,* human/mouse), or any other gene not found in nature to exist within the CHO locus of interest.

[0063] Percent identity, when describing a locus of interest, such as SEQ ID NO:1 or SEQ ID NO:4, or a fragment thereof, is meant to include homologous sequences that display the recited identity along regions of contiguous homology, but the presence of gaps, deletions, or insertions that have no homolog in the compared sequence are not taken into account in calculating percent identity.

[0064] As used herein, a "percent identity" determination between, *e.g.*, SEQ ID NO:1, or fragment thereof, with a species homolog would not include a comparison of sequences where the species homolog has no homologous sequence to compare in an alignment (*i.e.,* SEQ ID NO:1 or the fragment thereof has an insertion at that point, or the species homolog has a gap or deletion, as the case may be). Thus, "percent identity" does not include penalties for gaps, deletions, and insertions.

[0065] A "homologous sequence" in the context of nucleic acid sequences refers to a sequence that is substantially homologous to a reference nucleic acid sequence. In some embodiments, two sequences are considered to be substantially homologous if at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more of their corresponding nucleotides are identical over a relevant stretch of residues. In some embodiments, the relevant stretch is a complete (i.e., full) sequence.

[0066] "Targeted insertion" refers to gene targeting methods employed to direct insertion or integration of the gene or nucleic acid sequence to a specific location on the genome, *i.e.,* to direct the DNA to a specific site between two nucleotides in a contiguous polynucleotide chain. Targeted insertion may also be done for a particular gene cassette, which includes multiple genes, regulatory elements, and/or nucleic acid sequences. "Insertion" and "integration" are used interchangeably. It is understood that insertion of a gene or nucleic acid sequence (for example a nucleic acid sequence comprising an

expression cassette) may result in (or may be engineered for) the replacement or deletion of one or more nucleic acids depending on the gene editing technique being utilized.

[0067] "Recognition site" or "recognition sequence" is a specific DNA sequence recognized by a nuclease or other enzyme to bind and direct site-specific cleavage of the DNA backbone. Endonucleases cleave DNA within a DNA molecule. Recognition sites are also referred to in the art as recognition target sites.

[0068] "Recombinase recognition site" is the specific DNA sequence recognized by a recombinase, such as Cre recombinase (Cre) or flippase (flp). Site-specific recombinases can perform DNA rearrangements, including deletions, inversions and translocations when one or more of their target recognition sequences are placed strategically into the genome of an organism. In one example, Cre specifically mediates recombination events at its DNA target recognition site *loxP*, which is composed of two 13-bp inverted repeats separated by an 8-bp spacer. More than one recombinase recognition site may be employed, for example, to facilitate a recombination-mediated exchange of DNA. Variants or mutants of recombinase recognition sites, for example lox sites, may also be employed (Araki, N. et al, 2002, *Nucleic Acids Research*, 30:19, e103).

[0069] "Recombinase-mediated cassette exchange" relates to a process for precisely replacing a genomic target cassette with a donor cassette. The molecular compositions typically provided in order to perform this process include 1) a genomic target cassette flanked both 5' and 3' by recognition target sites specific to a particular recombinase, 2) a donor cassette flanked by matching recognition target sites, and 3) the site-specific recombinase. Recombinase proteins are well known in the art (Turan, S. and Bode J., 2011, *FASEB J.*, 25, pp. 4088–4107) and enable precise cleavage of DNA within a specific recognition target site (sequence of DNA) without gain or loss of nucleotides. Common recombinase/site combinations include, but are not limited to, Cre/*lox* and Flp/*frt*.

[0070] A "vehicle" is a composition consisting of any polynucleotide or set of polynucleotides carrying an exogenous nucleic acid for introduction into a cell. A vehicle includes vectors, plasmids and mRNA molecules that are delivered to the cell by well-known transfection methods. In one example, an mRNA introduced into the cells may be transient and does not integrate into the genome, however the mRNA may carry exogenous nucleic acid necessary for the integration process to take place.

General Description

[0071] The invention is based at least in part on the discovery of unique sequences, *i.e.* loci, in a genome that exhibit more efficient recombination, insert stability, and higher

level expression than other regions or sequences in the genome. The invention is also based at least in part on the finding that when such expression-enhancing sequences are identified, a suitable gene or construct can be exogenously added in or near the sequences and that the exogenously added gene can be advantageously expressed or utilized for further genomic modifications. Such sequences, termed expression-enhancing sequences are considered stable and are not located within a coding region of the genome. These expression-enhancing and stability regions can be engineered for future cloning or genome editing events. Thus, a reliable expression system is built into the genomic backbone of the cell.

[0072] The invention is also based on the specific targeting of an exogenous gene to the integration site. The methods of the invention allow efficient "conversion" of the cell genome into a useful cloning cassette, for example, by employing recombinase-mediated cassette exchange (RMCE). To this end, the methods of the invention employ cellular genome recombinase recognition sites for the placement of genes of interest to create highly productive cell lines for recombinant protein production.

[0073] The compositions of the invention can also be included in expression constructs for example, in expression vectors for cloning and engineering new cell lines. Expression vectors comprising the polynucleotides of the invention can be used to express proteins transiently, or can be integrated into a genome by random or targeted recombination such as, for example, homologous recombination or recombination mediated by recombinases that recognize specific recombination sites (*e.g.*, Cre-lox-mediated recombination). Expression vectors comprising the polynucleotides of the invention can also be used to assess efficacy of other DNA sequences, for example, *cis*-acting regulatory sequences.

[0074] Integration sites are typically identified by either random integration or analysis of retroviral integration events. The CHO integration site described in detail herein was identified by random integration of DNA encoding a multi-chain antibody and the expressed protein was found to exhibit enhanced expression.

[0075] The example multi-chain antibody comprising one heavy chain (HC) and two copies of a light chain (LC) were randomly integrated into the genome in an expression cassette containing alternating hygromycin resistance genes (see, *e.g.* three identical Hyg genes as depicted in Figure 1A). One stable and high expression clone resulted from the integration of the expression cassette within the loci identified as SEQ ID NO:1.

[0076] Compared to integration into another region of the CHO genome (control integration site), the example multi-chain antibody exhibits higher expression levels when integrated within the locus of SEQ ID NO:1. Interestingly, gene copy number is comparable for the antibody-expressing polynucleotides integrated within SEQ ID NO:1

versus the control integration site, however protein titers are 3-fold higher for antibody-expressing polynucleotides integrated within SEQ ID NO:1.

[0077] Targeted recombination methods were used to convert the CHO cell genome into a cloning construct containing recombinase recognition sites (see, *e.g.* Figures 3A-B).

[0078] Essentially, following identification of the integration site of SEQ ID NO:1, recombinase recognition sites (*e.g.,* lox sites) were employed in the locus for introducing expression cassettes that comprise an expressible GOI, such as a selectable marker (see, *e.g.* Figures 3A-B), along with any other desirable elements such as, *e.g.*, promoters, enhancers, markers, operators, ribosome binding sites (*e.g.* internal ribosome entry sites), *etc.*

[0079] An illustration of an example donor construct used for targeted integration of lox sites within SEQ ID NO:1, is shown in Figure 1B. The donor construct comprises an expression cassette driven by a neomycin (neo) resistance gene and an internal ribosome entry site (IRES), wherein the cassette comprises a fluorescent marker (mKate) and is flanked on the 5' and 3' ends with recombinase recognition sites and 5' and 3' homology arms (homologous to SEQ ID NO:1). Insertion within the locus of SEQ ID NO:1 is shown, wherein the insertion results in the donor neo/mKate construct replacing the expression cassette comprising the hygromycin resistance marker, wherein the expression cassette within the SEQ ID NO:1 locus is flanked on its 5' and 3' ends by recombinase recognition sites connected to 5' and 3' homology arms (homologous to SEQ ID NO:1) (see Figure 1B).

[0080] Compositions and methods are provided for stably integrating a nucleic acid sequence into a eukaryotic cell, wherein the nucleic acid sequence is capable of enhanced expression by virtue of being integrated in SEQ ID NO:1 or an expression-enhancing fragment thereof. Cells are provided that contain a recombinase recognition sequence within SEQ ID NO:1 convenient for inserting a GOI, in order to achieve expression of a protein of interest from the GOI. Compositions and methods are also provided for targeting the integration sites in connection with expression constructs, for example, expression vectors, and for adding an exogenous nucleic acid(s) into a CHO cell of interest.

Physical and Functional Characterization of a CHO Integration Site

[0081] The nucleic acid sequence of SEQ ID NO:1 (and broader nucleic acid sequence of SEQ ID NO:4) was empirically identified by sequences upstream and downstream of the integration site of a nucleic acid construct (comprising an expression cassette) of a cell line expressing a protein at a high level. The nucleic acid sequences of the invention

provide sequences with a new functionality associated with enhanced expression and stability of a nucleic acid (for example, an exogenous nucleic acid comprising a GOI) and without being bound by any one theory, may function the same or differently from that previously described for *cis*-acting elements such as promoters, enhancers, locus control regions, scaffold attachment regions or matrix attachment regions. SEQ ID NO:1 does not appear to have any open reading frames (ORFs), making it unlikely that the locus encodes novel *trans*-activator proteins. A putative Zinc finger protein has been identified in the genomic locus 3' (downstream) of SEQ ID NO:4.

[0082] Expression-enhancing activity was identified with respect to integration of an expression cassette comprising a first hygromycin (Hyg) gene, a first GOI, a second Hyg gene, a second GOI, a third Hyg gene and a third GOI encoding sequence within a unique site of a non-coding region of CHO genomic DNA. Expression vectors comprising, for example, a 5' isolated 1 kb region and a 3' isolated 1 kb region identified from the non-coding region of CHO genomic DNA with respect to an expression cassette expressing a GOI were able to confer upon CHO cells transfected with them high levels of expression of recombinant proteins.

[0083] The invention encompasses expression vectors comprising reverse orientated SEQ ID NO:1 fragments or SEQ ID NO:4 fragments. Other combinations of the fragments described herein can also be developed. Examples of other combinations of the fragments described herein that can also be developed include sequences that include multiple copies of the expression-enhancing sequences disclosed herein, or sequences derived by combining the disclosed SEQ ID NO:1 fragments or SEQ ID NO:4 fragments with other nucleotide sequences to achieve optimal combinations of regulatory elements. Such combinations can be contiguously linked or arranged to provide optimal spacing of the SEQ ID NO:1 or SEQ ID NO:4 fragments (*e.g.,* by the introduction of spacer nucleotides between the fragments). Regulatory elements can also be arranged to provide optimal spacing of a SEQ ID NO:1 fragment with respect to the regulatory elements.

[0084] SEQ ID NO:1 and SEQ ID NO: 4 disclosed herein were isolated from CHO cells. Other mammalian species (such as, for example, humans or mice), were found to have limited homology to the identified expression-enhancing region, however homologous sequences may be found in cell lines derived from other tissue types of *Cricetulus griseus*, or other homologous species, and can be isolated by techniques that are well-known in the art. For example, one may identify other homologous sequences by cross-species hybridization or PCR-based techniques. In addition, changes can be made in the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO: 4, or fragments thereof, by

site-directed or random mutagenesis techniques that are well known in the art. The resulting sequence variants can then be tested for expression-enhancing activity as described herein. DNAs that are at least about 90% identical in nucleic acid identity to SEQ ID NO:1, SEQ ID NO: 4, or fragments thereof, having expression-enhancing activity are isolatable by routine experimentation, and are expected to exhibit expression-enhancing activity. For fragments of SEQ ID NO:1 or SEQ ID NO: 4, percent identity refers to that portion of the reference native sequence that is found in the SEQ ID NO:1 fragment or SEQ ID NO: 4 fragment. Accordingly, homologs of SEQ ID NO:1, SEQ ID NO: 4, or fragments thereof, and variants thereof, are also encompassed by embodiments of the invention.

[0085] In certain embodiments, the fragment of SEQ ID NO:1 is selected from the group consisting of nucleotides spanning positions numbered 10-4,000; 100-3,900; 200-3,800; 300-3,700; 400-3,600; 500-3,500; 600-3,400; 700-3,300; 800-3,200; 900-3,100; 1,000-3,000; 1,100-2,900; 1,200-2,800; 1,300-2,700; 1,200-2,600; 1,300-2,500; 1,400-2,400; 1,500-2,300; 1,600-2,200; 1,700-2100; 1,800-2050; 1850-2050, 1,900-2040; 1950-2,025, 1990-2021, 2002-2021 and 2,010-2,015 of SEQ ID NO:1. In another embodiment, the fragment of SEQ ID NO:1 is selected from the group consisting of nucleotides spanning positions numbered 10-500; 500-1,000; 500-2,100; 1,000-1,500; 1,000-2,100; 1,500-2,000; 1,500-2,500; 2,000-2,500; 2,500-3,000; 2,500-3,500; 3,000-3,500; 3,000-4,000; and 3,500-4,000 of SEQ ID NO:1. In certain embodiments, the exogenous nucleic acid sequence integrates at or near specific sites within the fragment described above.

[0086] In another embodiment, the exogenous nucleic acid sequence is positioned within SEQ ID NO:1 or fragments thereof as described above, or within a sequence that is at least about 90% identical, at least about 91% identical, at least about 92% identical, at least about 93% identical, at least about 94% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, or at least about 99% identical to the expression-enhancing sequence of SEQ ID NO:1 or an expression-enhancing fragment thereof.

[0087] Cell populations expressing enhanced levels of a protein of interest can be developed using the methods provided herein. The absolute level of expression will vary with the specific protein, depending on how efficiently the protein is processed by the cell. Cell pools developed with exogenous sequence(s) integrated within the expression-enhancing sequences of the invention are stable over time, and can be treated as stable cell lines for most purposes. Recombination steps can also be delayed until later in the process of development of the cell lines of the invention.

CHO Expression-Enhancing Locus and Fragments Thereof

[0088] The invention encompasses an expression-enhancing fragment of a nucleotide sequence that is at least about 90% identical, at least about 91% identical, at least about 92% identical, at least about 93% identical, at least about 94% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, or at least about 99% identical to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO: 4. The invention includes vectors comprising a fragment, including for transient or stable transfection, spanning positions numbered 10-4,000; 100-3,900; 200-3,800; 300-3,700; 400-3,600; 500-3,500; 600-3,400; 700-3,300; 800-3,200; 900-3,100; 1,000-3,000; 1,100-2,900; 1,200-2,800; 1,300-2,700; 1,200-2,600; 1,300-2,500; 1,400-2,400; 1,500-2,300; 1,600-2,200; 1,700-2100; 1,800-2050; 1850-2050, 1,900-2040; 1950-2,025, 1990-2021, 2002-2021 and 2,010-2,015 of SEQ ID NO:1. The invention also includes a eukaryotic cell comprising such a fragment wherein the fragment is exogenous to the cell and is integrated into the cell genome, and cells comprising such a fragment having at least one recombinase recognition site that is within, immediately 5', or immediately 3' to the fragment.

[0089] In one embodiment, the expression-enhancing fragment of SEQ ID NO:1 is located at a position within SEQ ID NO:1 spanning positions numbered 10-500; 500-1,000; 500-2,100; 1,000-1,500; 1,000-2,100; 1,500-2,000; 1,500-2,500; 2,000-2,500; 2,500-3,000; 2,500-3,500; 3,000-3,500; 3,000-4,000; or 3,500-4,000 of SEQ ID NO:1.

[0090] Where stable integration and/or enhanced transcription of an integrated polynucleotide is supported, the exact location of the locus insertion (*i.e.* integration) site with respect to the exemplified sites is not essential. Rather, the integration site can be at any position that is within or adjacent to SEQ ID NO:1 or a fragment of SEQ ID NO:1, or SEQ ID NO: 4 or a fragment of SEQ ID NO: 4, as described herein. Whether a specific chromosomal location within or adjacent to the locus of interest supports stable integration and efficient transcription of an integrated exogenous gene can be determined in accordance with standard procedures well known in the art or methods exemplified herein.

[0091] The integration sites considered herein are located within a locus comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:4, or within close proximity to the locus of interest, *e.g.,* less than about 1 kb, 500 base pairs (bp), 250 bp, 100 bp, 50 bp, 25 bp, 10 bp, or less than about 5 bp upstream (5') or downstream (3') with respect to the location of SEQ ID NO:1 on the chromosomal DNA. In still some other embodiments, the employed integration site is located at about 1000, 2500, 5000 or more base pairs upstream (5') or downstream (3') with respect to the location of SEQ ID NO:1 or SEQ ID

NO:4 on the chromosomal DNA.

[0092] It is understood in the art that large genomic regions, such as scaffold/matrix attachment regions, are employed for efficient replication and transcription of chromosomal DNA. A scaffold/matrix attachment region (S/MAR), also known as called scaffold-attachment region (SAR), or matrix-associated or matrix attachment region (MAR), is a eukaryotic genomic DNA region where the nuclear matrix attaches. Without being bound by any one theory, S/MARs typically map to non-coding regions, separate a given transcriptional region (*e.g.* chromatin domain) from its neighbors, and also provide platforms for the machinery and/or binding of factors that enable transcription, such as recognition sites for DNAses or polymerases. Some S/MARs have been characterized at about 14-20 kb in length (Klar, et al. 2005, *Gene* 364:79-89). As such, integration of genes at *LOCUS 1* (within or near SEQ ID NO:1 or SEQ ID NO:4) is expected to confer enhanced expression.

[0093] Those in the art will recognize that several elements may be optimized for high transcriptional activity at the subject locus, resulting in high expression of an inserted gene encoding a protein of interest. Elements to consider include a strong promoter to drive transcription, adequate transcriptional machinery, and DNA having an open and accessible configuration. Insertion at the subject locus may be optimized within the skill of the person in the art by targeting an integration site selected within SEQ ID NO:1 or SEQ ID NO:4.

[0094] In one embodiment, the expression-enhancing sequence of SEQ ID NO:1 is employed to enhance the expression of a GOI. Figure 2A shows results of a GOI operably linked to SEQ ID NO:1 (*LOCUS 1*) compared to the same GOI integrated in a different locus in the CHO cell genome (Control Locus), The gene copy number measured for each cell line is equivalent, yet experiments show that the mRNA level and the protein titer of cells expressing the GOI are 3-fold higher for GOI operably linked to *LOCUS 1*.

[0095] In various embodiments, expression of a GOI can be enhanced by placing the GOI within SEQ ID NO:1 or SEQ ID NO: 4. In various embodiments, enhancement in expression is at least about 1.5-fold to about 3-fold or more.

Genetically Modifying the Target Locus

[0096] Methods for genetically engineering a cell genome in a particular location (*i.e.* target locus) may be achieved in several ways. Genetic editing techniques were used to stably integrate a nucleic acid sequence into a eukaryotic cell, wherein the nucleic acid sequence is an exogenous sequence not normally found in such cells. Clonal expansion

is necessary to ensure that the cell progeny will share the identical genotypic and phenotypic characteristics of the engineered cell line. In some examples, native cells are modified by a homologous recombination technique to integrate an exogenous nucleic acid sequence within SEQ ID NO:1 or SEQ ID NO: 4. In other examples, cells are provided that contain at least one recombinase recognition sequence within SEQ ID NO:1 or SEQ ID NO: 4 convenient for integrating an exogenous nucleic acid sequence or a gene of interest.

[0097] In some examples, cells are provided that contain a first recombinase recognition sequence and a second recombinase recognition sequence wherein each of the first and the second recombinase recognition sequences is selected from the group comprising *Lox*P, *Lox*511, *Lox*5171, *Lox*2272, *Lox*2372, *Lox*m2, *Lox*-FAS, *Lox*71, *Lox*66 and the mutants thereof. In this case, where recombinase-mediated cassette exchange (RMCE) is desired, the site specific recombinase is Cre recombinase or its derivative. In other examples, each of the first and the second recombinase recognition sequences is selected from the group comprising FRT, F3, F5, FRT mutant-10, FRT mutant+10 and the mutants thereof, and in this scenario, where RCME is desired, the site-specific recombinase is Flp recombinase or its derivative. In yet another example, each of said the first and the second recombinase recognition sequences is selected from the group comprising attB, attP and the mutants thereof, and in this case where RMCE is desired, the site-specific recombinase is phiC31 integrase or its derivative.

[0098] In one aspect, methods and compositions for stably integrating a nucleic acid sequence within SEQ ID NO:1 or SEQ ID NO: 4, or an expression-enhancing fragment thereof, are via homologous recombination. A nucleic acid molecule, *.i.e.* gene or polynucleotide of interest, can be inserted into the targeted locus (*i.e.* SEQ ID NO:1) by homologous recombination or by using site-specific nuclease methods that specifically target sequences at the integration sites. For homologous recombination, homologous polynucleotide molecules (*i.e.* homologous arms) line up and exchange a stretch of their sequences. A transgene can be introduced during this exchange if the transgene is flanked by homologous genomic sequences. In one example, a recombinase recognition site can be introduced into the host cell genome at the integration sites.

[0099] Homologous recombination in eukaryotic cells can be facilitated by introducing a break in the chromosomal DNA at the integration site. Model systems have demonstrated that the frequency of homologous recombination during gene targeting increases if a double-strand break is introduced within the chromosomal target sequence. This may be accomplished by targeting certain nucleases to the specific site of integration. DNA-binding proteins that recognize DNA sequences at the target locus are known in the art.

Gene targeting vectors are also employed to facilitate homologous recombination. In the absence of a gene targeting vector for homology directed repair, the cells frequently close the double-strand break by non-homologous end-joining (NHEJ) which may lead to deletion or insertion of multiple nucleotides at the cleavage site. Should insertions or deletions (InDels) occur, as such, a small number of nucleotides are either inserted or deleted at random at the site of the break and these InDels may shift or disrupt any open reading frame (ORF) of a gene within the target locus. It is understood that the locus identified as SEQ ID NO:1 (or SEQ ID NO:4) is not a gene coding region. Hence, no disruption of endogenous gene transcription is envisioned by insertion and/or deletion at this locus.

[0100] Homology directed repair (or homology directed recombination) (HDR) is particularly useful for inserting or integrating genes at the subject locus. A donor construct comprises homologous arms derived from SEQ ID NO:1 or SEQ ID NO:4 as described herein.

[0101] Gene targeting vector construction and nuclease selection are within the skill of the artisan to whom this invention pertains.

[0102] In some examples, zinc finger nucleases (ZFNs), which have a modular structure and contain individual zinc finger domains, recognize a particular 3-nucleotide sequence in the target sequence (*e.g.* site of targeted integration). Some embodiments can utilize ZFNs with a combination of individual zinc finger domains targeting multiple target sequences.

[0103] Transcription activator-like (TAL) effector nucleases (TALENs) may also be employed for site-specific genome editing. TAL effector protein DNA-binding domain is typically utilized in combination with a non-specific cleavage domain of a restriction nuclease, such as FokI. In some embodiments, a fusion protein comprising a TAL effector protein DNA-binding domain and a restriction nuclease cleavage domain is employed to recognize and cleave DNA at a target sequence within the locus of the invention (Boch J et al., 2009 *Science* 326:1509-1512).

[0104] RNA-guided endonucleases (RGENs) are programmable genome engineering tools that were developed from bacterial adaptive immune machinery. In this system—the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) immune response—the protein Cas9 forms a sequence-specific endonuclease when complexed with two RNAs, one of which guides target selection. RGENs consist of components (Cas9 and tracrRNA) and a target-specific CRISPR RNA (crRNA). Both the efficiency of DNA target cleavage and the location of the cleavage sites vary based on the position of a protospacer adjacent motif (PAM), an additional requirement for target

recognition (Chen, H. et al, *J. Biol. Chem.* published online March 14, 2014, as Manuscript M113.539726).

[0105] Strategies for identifying sequences unique for the specific targeting locus of SEQ ID NO:1 are known in the art, however, alignment of many of these sequences to the CHO genome reveals potential off-target sites with 16-17 base pair match. One example 20 bp Guide RNA encoded by the sequence set forth in SEQ ID NO:5 (corresponding to nucleotides 1990-2001 of SEQ ID NO: 1) is useful for RNA-guided CRISPR/Cas gene editing of SEQ ID NO:1 or SEQ ID NO:4. A plasmid comprising a promoter that drives expression of the small guided RNA and a tracrRNA (for ex. SEQ ID NO:6), as well as carrying a suitable Cas9 enzyme under control of a promoter may be co-transfected with a donor vector (carrying the gene of interest flanked by 5' and 3' homology arms) to employ targeted integration by this method. Various modifications and variants of the RNA molecules in addition to those described hereinabove are apparent to those skilled in the art and are intended to fall within the scope of the invention.

[0106] In some embodiments, the vehicle for introducing, into the genome, an exogenous nucleic acid comprising a sequence coding for the gene of interest or recognition sequence or gene cassette, as the case may be, comprises a vector carrying the exogenous nucleic acid and one or more additional vectors or mRNA. In one embodiment, the one or more additional vectors or mRNA comprise a nucleotide sequence encoding a site-specific nuclease, including but not limited to a zinc finger nuclease (ZFN), a ZFN dimer, a transcription activator-like effector nuclease (TALEN), a TAL effector domain fusion protein, and an RNA-guided DNA endonuclease. In certain embodiments, the one or more vectors or mRNA comprise a first vector comprising a guide RNA, a tracrRNA and a nucleotide sequence encoding a Cas enzyme, and a second vector comprising a donor (exogenous) nucleotide sequence. Such donor sequence comprises a nucleotide sequence coding for the gene of interest, or the recognition sequence, or the gene cassette comprising any one of these exogenous elements intended for targeted insertion. Where mRNA is used, the mRNA can be transfected into the cell by means of common transfection methods known to the skilled person and may encode an enzyme, for example a transposase or endonuclease. Although an mRNA introduced into the cells may be transient and does not integrate into the genome, the mRNA may carry an exogenous nucleic acid necessary or beneficial for the integration to take place. In some instances, mRNA is chosen in order to eliminate any risk of long-lasting side effects of an accessory polynucleotide, where only short-term expression is required to achieve the desired integration of a GOI.

[0107] Still other methods of homologous recombination are available to the skilled

artisan, such as BuD-derived nucleases (BuDNs) with precise DNA-binding specificities (Stella, S. et al. *Acta Cryst.* 2014, D70, 2042-2052). Precise genome modification methods are chosen based on the tools available compatible with unique target sequences within SEQ ID NO:1 so that disruption of the cell phenotype is avoided.

Gene Targeting Constructs

[0108] The polynucleotide sequence to be integrated into the host genome may be any industrially useful DNA sequence, such as a recognition sequence, for the generation of cellular expression systems. The polynucleotide sequence to be integrated into the host genome may encode any therapeutically or industrially useful protein or proteins as described herein. Identifying the target sequence within the target locus to integrate the exogenous nucleic acid sequence depends on a number of factors. Depending on the method of homologous recombination utilized, it is well within the skill of the artisan to select sequences homologous to SEQ ID NO:1 or SEQ ID NO: 4. Site-specific nuclease vectors, when employed, require additional components (sequence compositions) that recognize the specific site intended for DNA cleavage.

[0109] As such, a gene targeting construct typically incorporates such nucleotide sequences that facilitate the targeted integration of an exogenous nucleic acid sequence into the locus of interest. In some embodiments, the construct comprises a first homologous arm and a second homologous arm. In other embodiments, the construct (*e.g.* a gene cassette) comprises homologous arms derived from SEQ ID NO:1 or SEQ ID NO:4. In some embodiments, the homology arms comprise a nucleotide sequence homologous to a nucleotide sequence present in SEQ ID NO:1 or SEQ ID NO:4. In specific embodiments, the construct comprises a 5' homology arm having the nucleotide sequence of SEQ ID NO: 2 (corresponding to nucleotides 1001-2001 of SEQ ID NO: 1), and a 3' homology arm having the nucleotide sequence of SEQ ID NO:3 (corresponding to nucleotides 2022-2001 of SEQ ID NO: 1). Homologous arms, for example a first homologous arm (also called 5' homology arm) and a second homologous arm (also called 3' homology arm) are homologous to a targeted sequence within the locus. The homologous arms from 5' to 3' may expand a region or targeted sequence within the locus that comprises at least 1 kb, or at least about 2 kb, or at least about 3 kb, or at least about 4 kb, or at least 5 kb, or at least about 10 kb. In other embodiments, the total number of nucleotides of a targeted sequence selected for a first and second homologous arm comprises at least 1 kb, or at least about 2 kb, or at least about 3 kb, or at least about 4 kb, or at least 5 kb, or at least about 10 kb. In some instances, the distance between the 5' homology arm and the 3' homology arm (homologous to the targeted

sequence) comprises at least 5 bp, 10 bp, 20 bp, 30 bp, 40 bp, 50 bp, 60 bp, 70 bp, 80 bp, 90 bp, 100 bp, 200 bp, 300 bp, 400 bp, 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, or at least 1 kb, or at least about 2 kb, or at least about 3 kb, or at least about 4 kb, or at least 5 kb, or at least about 10 kb. In instances where SEQ ID NO: 2 and SEQ ID NO: 3 are chosen as 5' and 3' homology arms, the distance between the two homology arms can be 20 nucleotides (corresponding to nucleotides 2002-2021 of SEQ ID NO: 1); and such homology arms can mediate integration of an exogenous nucleic acid sequence within a locus comprising SEQ ID NO: 1, e.g., within nucleotides 1990-2021 or 2002-2021 of SEQ ID NO: 1, and a simultaneous deletion of nucleotides 2002-2021 of SEQ ID NO: 1.

[0110] In other embodiments, the construct comprises a first homologous arm and a second homologous arm, wherein the first and second homologous arms combined comprise a targeted sequence which replaces an endogenous sequence within the locus. In yet other embodiments, the first and second homologous arms comprise a targeted sequence which integrates or inserts within an endogenous sequence within the locus.

[0111] Modified cell lines were created by integrating one or more recombinase recognition sites at a location within SEQ ID NO:1. These modified cell lines may also include additional exogenous genes for negative or positive selection of the expressed gene of interest.

[0112] The invention provides methods for modifying a CHO cell genome comprising introducing one or more vehicles into the cell, wherein the one or more vehicles comprise an exogenous nucleic acid comprising a sequence for integration, a 5' homology arm homologous to a sequence present in the nucleotide sequence of SEQ ID NO:1, and a 3' homology arm homologous to a sequence present in the nucleotide sequence of SEQ ID NO:1. In some embodiments, the methods further provide one or more vehicles comprising a nuclease and compositions for site-specific DNA cleavage at the integration site.

[0113] The modified cell lines may be utilized as convenient and stable expression systems for recombinase-mediated cassette exchange (RMCE). A nucleic acid sequence encoding a protein of interest can be conveniently integrated into the modified cell comprising SEQ ID NO:1 or an expression-enhancing fragment thereof, having at least one recombinase recognition site, for example, through an RMCE process.

[0114] Recombinant expression vectors can comprise synthetic or cDNA-derived DNA fragments encoding a protein, operably linked to a suitable transcriptional and/or translational regulatory element derived from mammalian, viral or insect genes. Such regulatory elements include transcriptional promoters, enhancers, sequences encoding suitable mRNA ribosomal binding sites, and sequences that control the termination of

transcription and translation, as described in detail below. Mammalian expression vectors can also comprise nontranscribed elements such as an origin of replication, other 5' or 3' flanking nontranscribed sequences, and 5' or 3' nontranslated sequences such as splice donor and acceptor sites. A selectable marker gene to facilitate recognition of transfectants may also be incorporated.

[0115] Fluorescent markers are suitable selectable marker genes for the recognition of gene cassettes that have or have not been successfully inserted and/or replaced, as the case may be. Examples of fluorescent markers are well-known in the art, including, but not limited to *Discosoma* coral (DsRed), green fluorescent protein (GFP), enhanced green fluorescent protein (eGFP), cyano fluorescent protein (CFP), enhanced cyano fluorescent protein (eCFP), yellow fluorescent protein (YFP), enhanced yellow fluorescent protein (eYFP) and far-red fluorescent protein (*e.g.* mKate, mKate2, mPlum, mRaspberry or E2-crimson. See also, e.g., Nagai, T., et al. 2002 *Nature Biotechnology* 20:87-90; Heim, R. et al. 23 February 1995 *Nature* 373:663-664; and Strack, R.L. et al. 2009 *Biochemistry* 48:8279-81.

[0116] Transcriptional and translational control sequences in expression vectors useful for transfecting vertebrate cells may be provided by viral sources. For example, commonly used promoters and enhancers are derived from viruses such as polyoma, adenovirus 2, simian virus 40 (SV40), and human cytomegalovirus (CMV). Viral genomic promoters, control and/or signal sequences may be utilized to drive expression, provided such control sequences are compatible with the host cell chosen. Non-viral cellular promoters can also be used (*e.g.,* the β-globin and the EF-1α promoters), depending on the cell type in which the recombinant protein is to be expressed.

[0117] DNA sequences derived from the SV40 viral genome, for example, the SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites may be used to provide other genetic elements useful for expression of a heterologous DNA sequence. Early and late promoters are particularly useful because both are obtained easily from the SV40 virus as a fragment that also comprises the SV40 viral origin of replication (Fiers *et al.,* Nature 273:113, 1978). Smaller or larger SV40 fragments may also be used. Typically, the approximately 250 bp sequence extending from the Hind III site toward the BglI site located in the SV40 origin of replication is included.

[0118] Bicistronic expression vectors used for the expression of multiple transcripts have been described previously (Kim S. K. and Wold B. J., Cell 42:129, 1985) and can be used in combination with an expression-enhancing sequence of the invention, e.g. SEQ ID NO:1, or a fragment thereof. Other types of expression vectors will also be useful, for example, those described in U.S. Pat. No. 4,634,665 (Axel *et al.*) and U.S. Pat. No.

4,656,134 (Ringold *et al.*).

Proteins of Interest

[0119] Any protein of interest suitable for expression in eukaryotic cells can be used. For example, the protein of interest includes, but is not limited to, an antibody or antigen-binding fragment thereof, a chimeric antibody or antigen-binding fragment thereof, an ScFv or fragment thereof, an Fc-fusion protein or fragment thereof, a growth factor or a fragment thereof, a cytokine or a fragment thereof, or an extracellular domain of a cell surface receptor or a fragment thereof. Proteins of interest may be simple polypeptides consisting of a single subunit, or complex multisubunit proteins comprising two or more subunits.

Host Cells and Transfection

[0120] The host cells used in the methods of the invention are mammalian host cells including, for example, Chinese hamster ovary (CHO) cells and mouse cells. In a preferred embodiment, the invention provides a nucleic acid sequence fragment of SEQ ID NO:1 that encodes an expression-enhancing sequence in a CHO cell. An integration site can be found within SEQ ID NO:1, or any fragment of SEQ ID NO:1. An integration site, for example, may be a recombinase recognition site placed within SEQ ID NO:1, or any fragment of SEQ ID NO:1. One example of a suitable integration site is a *Lox*P site. Another example of a suitable integration site is two recombinase recognition sites, for example, selected from the group consisting of a *Lox*P site, a *Lox*511 site, a *Lox*2272 site, a *Lox*2372 site, a *Lox*m2 site, a *Lox*71 site, a *Lox*66 site and a *Lox*5171 site. In other embodiments, the integration site is located at a position within a sequence or adjacent to a position within a sequence selected from the group consisting of nucleotides spanning positions numbered 10-4,000; 100-3,900; 200-3,800; 300-3,700; 400-3,600; 500-3,500; 600-3,400; 700-3,300; 800-3,200; 900-3,100; 1,000-3,000; 1,100-2,900; 1,200-2,800; 1,300-2,700; 1,200-2,600; 1,300-2,500; 1,400-2,400; 1,500-2,300; 1,600-2,200; 1,700-2100; 1,800-2050; 1850-2050, 1,900-2040; 1950-2,025, 1990-2021, 2002-2021 and 2,010-2,015 of SEQ ID NO:1. In certain embodiments, the integration site at a position within SEQ ID NO:1 or adjacent to a position within SEQ ID NO:1 is selected from the group consisting of nucleotides spanning positions numbered 1990-1991, 1991-1992, 1992-1993, 1993-1994, 1995-1996, 1996-1997, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2003-2004, 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2019-2020, and 2020-2021of SEQ ID NO:1.

[0121] The invention includes a mammalian host cell transfected with an expression vector or an mRNA of the invention. While any mammalian cell may be used, in one particular embodiment the host cell is a CHO cell.

[0122] Transfected host cells include cells that have been transfected with expression vectors or mRNA molecules that comprise a sequence encoding a protein or polypeptide. Expressed proteins may be secreted into the culture medium, depending on the nucleic acid sequence selected, but may be retained in the cell or deposited in the cell membrane. Various mammalian cell culture systems can be employed to express recombinant proteins. Other cell lines developed for specific selection or amplification schemes will also be useful with the methods and compositions provided herein, provided that a target locus having at least 80% homology to SEQ ID NO:1 has been identified. An embodied cell line is the CHO cell line designated K1. To achieve high volume production of recombinant proteins, the host cell line may be pre-adapted to bioreactor medium in the appropriate case.

[0123] Several transfection protocols are known in the art, and are reviewed in Kaufman (1988) Meth. Enzymology 185:537. The transfection protocol chosen will depend on the host cell type and the nature of the GOI, and can be chosen based upon routine experimentation. The basic requirements of any such protocol are first to introduce DNA encoding the protein of interest into a suitable host cell, and then to identify and isolate host cells which have incorporated the heterologous DNA in a relatively stable, expressible manner. mRNA molecules encoding proteins useful for integration into the host cell genome or other function may be transient and therefore time-limited.

[0124] 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. mRNA delivery includes methods using TransMessenger™ and TransIT® (Bire et al. *BMC Biotechnology* 2013, 13:75).

[0125] One commonly used method of introducing heterologous DNA into a cell is calcium phosphate precipitation, for example, as described by Wigler *et al.* (Proc. Natl.

Acad. Sci. USA 77:3567, 1980). DNA introduced into a host cell by this method frequently undergoes rearrangement, making this procedure useful for cotransfection of independent genes.

[0126] Polyethylene-induced fusion of bacterial protoplasts with mammalian cells (Schaffner *et al.,* (1980) Proc. Natl. Acad. Sci. USA 77:2163) is another useful method of introducing heterologous DNA. Protoplast fusion protocols frequently yield multiple copies of the plasmid DNA integrated into the mammalian host cell genome, and this technique requires the selection and amplification marker to be on the same plasmid as the GOI.

[0127] Electroporation can also be used to introduce DNA directly into the cytoplasm of a host cell, for example, as described by Potter *et al.* (Proc. Natl. Acad. Sci. USA 81:7161, 1988) or Shigekawa *et al.* (BioTechniques 6:742, 1988). Unlike protoplast fusion, electroporation does not require the selection marker and the GOI to be on the same plasmid.

[0128] Other reagents useful for introducing heterologous DNA into a mammalian cell have been described, such as Lipofectin™ Reagent and Lipofectamine™Reagent (Gibco BRL, Gaithersburg, Md.). Both of these commercially available reagents are used to form lipid-nucleic acid complexes (or liposomes) which, when applied to cultured cells, facilitate uptake of the nucleic acid into the cells.

[0129] 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™.

[0130] A method for amplifying the GOI is also desirable for expression of the recombinant protein, and typically involves the use of a selection marker (reviewed in Kaufman *supra*). Resistance to cytotoxic drugs is the characteristic most frequently used as a selection marker, and can be the result of either a dominant trait (*e.g.,* can be used independent of host cell type) or a recessive trait (*e.g.,* useful in particular host cell types that are deficient in whatever activity is being selected for). Several amplifiable markers are suitable for use in the expression vectors of the invention (*e.g.*, as described in Sambrook, Molecular Biology: A Laboratory Manual, Cold Spring Harbor Laboratory, NY, 1989; pgs 16.9-16.14).

[0131] Useful selectable markers for gene amplification in drug-resistant mammalian cells are shown in Table 1 of Kaufman, R. J., *supra*, and include DHFR-MTX resistance, P-glycoprotein and multiple drug resistance (MDR)-various lipophilic cytotoxic agents (*e.g.,* adriamycin, colchicine, vincristine), and adenosine deaminase (ADA)-Xyl-A or adenosine

and 2'-deoxycoformycin.

[0132] Other dominant selectable markers include microbially derived antibiotic resistance genes, for example neomycin, kanamycin or hygromycin resistance. However, these selection markers have not been shown to be amplifiable (Kaufman, R. J., *supra*,). Several suitable selection systems exist for mammalian hosts (Sambrook supra, pgs 16.9-16.15). Co-transfection protocols employing two dominant selectable markers have also been described (Okayama and Berg, Mol. Cell Biol 5:1136, 1985).

[0133] Useful regulatory elements, described previously or known in the art, can also be included in the nucleic acid constructs used to transfect mammalian cells. The transfection protocol chosen and the elements selected for use therein will depend on the type of host cell used. Those of skill in the art are aware of numerous different protocols and host cells, and can select an appropriate system for expression of a desired protein, based on the requirements of the cell culture system used.

[0134] Other features of the invention will become apparent in the course of the following descriptions of exemplary embodiments which are given for illustration of the invention and are not intended to be limiting thereof.

EXAMPLES

[0135] The following examples are put forth so as to provide those of ordinary skill in the art how to make and use the methods and compositions described herein, and are not intended to limit the scope of the invention. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.,* amount, temperature, *etc.*) but some experimental error and deviation should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

Example 1. Identification of Locus of Interest and Characterization of Integration Sites

[0136] CHO K1 cells were transfected with two plasmids containing antibody sequences and selectable antibiotic resistance genes as selectable markers. Selection of stable transfectants was performed by expanding cells in the presence of antibiotics. Individual cell clones expressing high levels of antibodies were isolated with FASTR® sorting technology (see US Patent No. 8673589B2). Several clones exhibiting the highest antibody expression levels were identified.

[0137] The genomic DNA from these clones was fragmented with Covaris Adaptive Focused Acoustics (AFA)™ technology (Fisher, S. et al. 2011, *Genome Biology* 12:R1).

DNA libraries were generated (Agilent SureSelectXT #G9612A) and incubated with custom-made biotinylated RNA baits (Agilent SureSelectXT #5190-4811) designed against the entire plasmid sequences that were introduced into CHO cells. Genomic DNA fragments that contain plasmid sequences were enriched with magnetic Streptavidin beads and subjected to Illumina MiSeq sequencing to identify the plasmid integration sites. Fusion sequences that contain both plasmid sequence and CHO genome sequence were analyzed and aligned to the CHO genome. A single integration site was confirmed by Southern blot analysis and PCR followed by sequencing. The integration site having the nucleotide sequence of SEQ ID NO:1 was identified as an expression hotspot (see also GenBank Locus ID No. AFTD01150902.1, nt35529:39558). The integration sites were analyzed to determine their suitability for further generation of cell lines. It was desirable that the integration sites are located in a non-coding region that does not disrupt the cell's normal genomic machinery, *e.g.* translation of proteins, or alter the cell's phenotype.

[0138] From Blat search (Kent WJ., BLAT - the BLAST-like alignment tool. *Genome Res.* 2002 Apr;12(4):656-64) alignment, SEQ ID NO:1 shares very low homology to mouse and human genome sequences. Sequence blast of SEQ ID NO:1 against CHO-1[ATCC]_refseq_transcript (www.chogenome.org) revealed that the identified locus sequence does not contain any coding regions for any known genes. The broader sequence of SED ID NO:4, which encompasses SEQ ID NO:1, was also identified as a locus suitable for targeted integration.

[0139] The integration site sequences were determined to be located in non-coding regions of the CHO and mouse genomes, and further utilized in the below described experiments.

Example 2. Exogenous DNA Efficiently Incorporated into Host Cell Integration Sites

[0140] Targeted insertion of exogenous genes into the specific locus of the CHO genome identified as SEQ ID NO:1 was done by employing a TALE nuclease (TALEN). The construct containing antibody heavy and light chain sequences randomly integrated into the cell genome, as in Example 1, was targeted by TALEN. TALEN was targeted to locations within the three identical Hyg genes of the antibody expression construct (see Figure 1A). The TALEN target cleavage site for the Hyg sequence was based on ZiFit.partners.org (ZiFit Targeter Version 4.2). TALENs were designed based on known methods (Boch J et al., 2009 Science 326:1509-1512).

[0141] A donor mKate vector (see Figure 1B) and TALEN-encoding vector were

transfected into the CHO host cells using standard Lipofectin protocol (LIPOFECTAMINE, Life Technologies, Gaithersburg, Md.). Cells were cultured and stable clones with desirable features were isolated and sorted by FACS. Single integration in the desired locus was confirmed by Southern blot and PCR.

Example 3. Targeted Recombination of the Engineered Cells at the Locus of Interest by RMCE

[0142] A CHO cell line expressing high levels of a flourescent gene, *e.g.* mKate, wherein the gene is flanked by lox sites within the locus of interest, was selected for isolation. A second CHO cell line expressing a second flourescent gene, *dsRed*, wherein the gene is flanked by lox sites is located within a control locus, *i.e.* EESYR (US Patent No. 8389239B2, issued March 5, 2013).

[0143] Transfected CHO cells were adapted to grow in suspension in a serum-free production medium. The cells were then transfected in a ten centimeter plate with a donor expression vector and a plasmid encoding Cre recombinase. The donor expression vector contains a gene of interest encoding an Fc fusion protein flanked by Lox sites (see Figures 3A or 3B). Cells were cultured in culture medium with 400 µg/ml hygromycin for two weeks after transfection, and cells expressing eYFP but not mKate (or dsRed in the case of EESYR locus integration) were isolated using flow cytometry. Cells expressing eYFP were expanded in suspension cultures in serum-free production medium, and mRNA levels were determined by qRT-PCR using standard procedures for each cell pool encoding the Fc fusion protein (see Figure 4).

[0144] Recombination exchange efficiency (percent population of surviving cells expressing from the donor cassette marker, *i.e.* eYFP, as exchanged with the red marker, *i.e.* mKate or dsRed) was compared between cell pools (Table 1). High recombination exchange efficiency was observed at each locus.

TABLE 1: Recombination Efficiency

	Red marker	Exchange efficiency (%) (Red marker+/eYFP-)	Random Integration (%) (Red marker+/eYFP+)
LOCUS 1 (SEQ ID NO:1)	mKate	72	27
Control Locus (EESYR)	dsRed	92	7

[0145] Transcription was observed at a higher rate (1.5-fold higher) in the cell pool having an engineered LOCUS1 compared to the Control Locus (Figure 4).

[0146] The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and the accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

We claim:

1. An isolated cell comprising an exogenous nucleic acid sequence integrated within a locus of the genome of the cell, wherein the locus comprises a nucleotide sequence that is:

(a) at least 90% identical to SEQ ID NO:1 or SEQ ID NO: 4, and

(b) capable of enhancing expression of a protein encoded by an exogenous nucleic acid sequence by at least 1.5-fold compared to expression typically observed by random integration into the genome.

2. The cell of claim 1, wherein the cell is a CHO cell.

3. The cell of claim 1 or 2, wherein the exogenous nucleic acid sequence comprises one or more recombination recognition sequences,

optionally wherein the exogenous nucleic acid sequence comprises at least two recombination recognition sequences and a selectable marker placed between the two recombination recognition sequences,

optionally wherein the one or more recombination recognition sequences are selected from the group consisting of a *Lox*P site, a *Lox511* site, a *Lox2272* site, a *Lox2372* site, a *Lox5171* site, a *Lox*m2 site, a *Lox*71 site, a *Lox*66 site, a *Lox*Fas site, and a frt site.

4. The cell of any one of claims 1 to 3, wherein the exogenous nucleic acid sequence comprises a first exogenous gene of interest (GOI) and a first exogenous promoter, wherein the first exogenous GOI is operably linked to the first exogenous promoter.

5. The cell of claim 4, wherein the exogenous nucleic acid sequence further comprises:

(a) a second exogenous GOI and a second exogenous promoter, wherein the second exogenous GOI is positioned 3' of the first GOI and is operably linked to the second exogenous promoter; and

(b) a first recombination recognition sequence 5' of the first exogenous GOI, a second recombination recognition sequence 3' of the first exogenous GOI; and a third recombinase recognition sequence 3' to the second exogenous GOI,

optionally wherein the first and second recombination recognition sequences are different, and are selected from the group consisting of a *Lox*P site, a *Lox511* site, a *Lox2272* site, a *Lox2372* site, a *Lox5171* site, a *Lox*m2 site, a *Lox*71 site, a *Lox*66 site, a *Lox*Fas site,

and a frt site.

6. The cell of claim 5, wherein the first exogenous GOI encodes a light chain of an antibody or fragment thereof, and the second exogenous GOI encodes a heavy chain of an antibody or fragment thereof.

7. The cell of claim 5 or claim 6, wherein the exogenous nucleic acid sequence further comprises a third exogenous GOI-,

optionally wherein the third exogenous GOI is positioned 3' of the second exogenous GOI,

optionally wherein the first, second and third GOI encode a polypeptide selected from the group consisting of (i) a first light chain of an antibody or antigen-binding fragment thereof, a second light chain of an antibody or antigen-binding fragment thereof, and a heavy chain of an antibody or antigen-binding fragment thereof, or (ii) a light chain of an antibody or antigen-binding fragment thereof, a first heavy chain of an antibody or antigen-binding fragment thereof, and a second heavy chain of an antibody or antigen-binding fragment thereof.

8. A method of modifying a CHO cell genome comprising introducing into a CHO cell an exogenous nucleic acid, wherein the exogenous nucleic acid integrates into a locus of the genome of the cell, the locus comprising a nucleotide sequence that is:

(a) at least 90% identical to SEQ ID NO:1 or SEQ ID NO:4, and

(b) capable of enhancing expression of a protein encoded by an exogenous nucleic acid sequence by at least 1.5-fold compared to expression typically observed by random integration into the genome.

9. The method of claim 8, wherein the exogenous nucleic acid comprises one or more recombination recognition sequences,

optionally wherein the exogenous nucleic acid comprises at least two recombination recognition sequences and a selectable marker placed between the two recombination recognition sequences,

optionally wherein the one or more recombination recognition sequences are selected from the group consisting of a *Lox*P site, a *Lox511* site, a *Lox2272* site, a *Lox2372* site, a *Lox5171* site, a *Lox*m2 site, a *Lox*71 site, a *Lox*66 site, a *Lox*Fas site, and a frt site.

10. The method of claim 9, wherein the exogenous nucleic acid comprises a first exogenous gene of interest (GOI),

optionally wherein the exogenous nucleic acid further comprises a second exogenous GOI and is placed 3' of the first exogenous GOI,

optionally wherein the first exogenous GOI encodes a light chain of an antibody or an antigen-binding fragment thereof, and the second exogenous GOI encodes a heavy chain of an antibody or an antigen-binding fragment thereof.

11. The method of any one of claims 8 to 10, wherein the exogenous nucleic acid is introduced using a vehicle comprising a vector, the vector comprising:

a) a 5' homology arm homologous to a sequence present in the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO: 4,
b) the exogenous nucleic acid, and
c) a 3' homology arm homologous to a sequence present in the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO: 4.

12. The method of claim 11, wherein the exogenous nucleic acid is introduced using a vehicle comprising at least an additional vector or an mRNA molecule,

optionally wherein the additional vector is selected from the group consisting of an adenovirus, a lentivirus, a retrovirus, an adeno-associated virus, an integrating phage vector, a non-viral vector, a transposon and/or transposase, an integrase substrate, and a plasmid, and/or the additional vector comprises a nucleic acid encoding a site-specific nuclease, the site-specific nuclease being optionally selected from the group consisting of a zinc finger nuclease (ZFN), a ZFN dimer, a transcription activator-like effector nuclease (TALEN), a TAL effector domain fusion protein, and an RNA-guided DNA endonuclease.

13. The method of any one of claims 8 to 12, wherein prior to introducing the exogenous nucleic acid, the CHO cell comprises one or more recombination recognition sequences integrated in said locus, and wherein the exogenous nucleic acid integrates into the one or more recombination recognition sequences in said locus,

optionally wherein prior to introducing the exogenous nucleic acid, the CHO cell comprises at least two recombination recognition sequences integrated in said locus, and a selectable marker placed between the two recombination recognition sequences, wherein optionally the at least two recombination recognition sequences are different, and are selected

from the group consisting of a *Lox*P site, a *Lox511* site, a *Lox2272* site, a *Lox2372* site, a *Lox5171* site, a *Lox*m2 site, a *Lox*71 site, a *Lox*66 site, a *Lox*Fas site, and a frt site.

14. The method of any one of claims 8 to 13, wherein the exogenous nucleic acid comprises a first exogenous gene of interest (GOI) operably linked to a first exogenous promoter, a first recombination recognition sequence placed at 5' of the first exogenous promoter, and a second recombination recognition sequence placed at 3' of the first exogenous GOI, and wherein the first and second recombination recognition sequences are identical, respectively, to the at least two recombination recognition sequences present in the locus of the genome prior to introducing the exogenous nucleic acid,

optionally wherein the first and second recombination recognition sequences are different, and are selected from the group consisting of a *Lox*P site, a *Lox511* site, a *Lox2272* site, a *Lox2372* site, a *Lox5171* site, a *Lox*m2 site, a *Lox*71 site, a *Lox*66 site, a *Lox*Fas site, and a frt site.

15. The method of claim 14, wherein the exogenous nucleic acid further comprises a second exogenous GOI operably linked to a second exogenous promoter, wherein the second exogenous GOI is placed 3’ of the first exogenous GOI and 5' of the second recombination recognition sequence,

optionally wherein the first exogenous GOI encodes a light chain of an antibody or an antigen-binding fragment thereof, and the second exogenous GOI encodes a heavy chain of an antibody or an antigen-binding fragment thereof.

16. A method of expressing a gene of interest (GOI) in a cell comprising:

a) providing a cell comprising an exogenous nucleic acid sequence integrated within a locus of the genome of the cell, wherein the exogenous nucleic acid sequence comprises a first exogenous (GOI) operably linked to a first promoter, wherein the locus comprises a nucleotide sequence that is: (i) at least 90% identical to SEQ ID NO:1 or SEQ ID NO: 4, and (ii) capable of enhancing expression of a protein encoded by an exogenous nucleic acid sequence by at least 1.5-fold compared to expression typically observed by random integration into the genome; and wherein the first exogenous GOI encodes a first protein of interest (POI); and

b) culturing the cell of (a) under conditions that allow expression of the first exogenous GOI; and
c) recovering the first POI.

17. The method of claim 16, wherein the cell is a CHO cell.

18. The method of claim 16 or 17, wherein the exogenous nucleic acid sequence further comprises a second exogenous GOI operably linked to a second promoter, and wherein the conditions allow expression of the first exogenous GOI and the second exogenous GOI,

optionally wherein the first exogenous GOI encodes a light chain of an antibody, and the second exogenous GOI encodes a heavy chain of an antibody.

19. A vehicle for modifying a CHO cell genome to (i) express a protein of interest or (ii) integrate a recognition sequence, comprising a vector that comprises:

a. a 5' homology arm homologous to a sequence present in the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO: 4,
b. an exogenous nucleic acid sequence which (i) encodes the protein of interest or (ii) comprises the recognition sequence, respectively, and
c. a 3' homology arm homologous to a sequence present in the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO: 4

20. The vehicle of claim 19, wherein the exogenous nucleic acid sequence comprises one or more recombination recognition sequences that flank a selectable marker, and/or comprises a first exogenous gene of interest (GOI),

optionally wherein the vehicle comprises at least an additional vector or an mRNA molecule, wherein the additional vector optionally comprises a nucleic acid encoding a site-specific nuclease for integrating the recognition sequence, and wherein the site-specific nuclease is optionally selected from the group consisting of a zinc finger nuclease (ZFN), a ZFN dimer, a transcription activator-like effector nuclease (TALEN), a TAL effector domain fusion protein, or an RNA-guided DNA endonuclease.