



(43) International Publication Date
6 December 2012 (06.12.2012)

- (51) International Patent Classification:
C12Q 1/68 (2006.01) C12N 15/85 (2006.01)
- (21) International Application Number:
PCT/US2012/040599
- (22) International Filing Date:
1 June 2012 (01.06.2012)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/492,174 1 June 2011 (01.06.2011) US
- (71) Applicant (for all designated States except US): **PRECISION BIOSCIENCES, INC.** [US/US]; 302 East Pettigrew Street, Dibrell Building, Suite A-100, Durham, NC 27701 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **JANTZ, Derek** [US/US]; 215 Hillview Drive, Durham, NC 27703 (US).

SMITH, James, Jefferson [US/US]; 1808 Carnation Drive, Durham, NC 27703 (US). **NICHOLSON, Michael, G.** [US/US]; 1200 Willow Drive, Chapel Hill, NC 27517 (US).

(74) Agents: **BAIK, Peter, W.** et al.; Wilmer Cutler Pickering Hale And Dorr LLP, 60 State Street, Boston, MA 02109 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[Continued on next page]

(54) Title: METHODS AND PRODUCTS FOR PRODUCING ENGINEERED MAMMALIAN CELL LINES WITH AMPLIFIED TRANSGENES

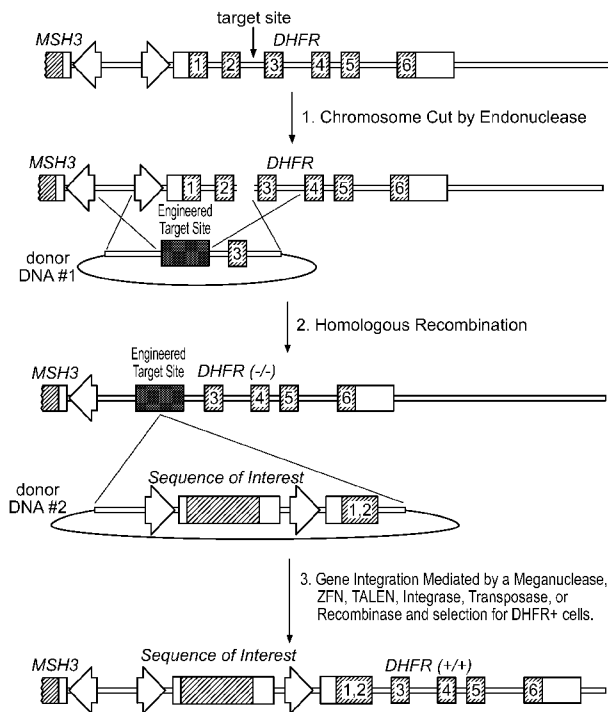


FIG. 4

(57) Abstract: Methods of inserting genes into defined locations in the chromosomal DNA of cultured mammalian cell lines which are subject to gene amplification are disclosed. In particular, sequences of interest (e.g., genes encoding bio-therapeutic proteins) are inserted proximal to selectable genes in amplifiable loci, and the transformed cells are subjected to selection to induce co-amplification of the selectable gene and the sequence of interest. The invention also relates to meganucleases, vectors and engineered cell lines necessary for performing the methods, to cell lines resulting from the application of the methods, and use of the cell lines to produce protein products of interest.

WO 2012/167192 A2



(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
- *with sequence listing part of description (Rule 5.2(a))*

METHODS AND PRODUCTS FOR PRODUCING ENGINEERED MAMMALIAN CELL LINES WITH AMPLIFIED TRANSGENES

FIELD OF THE INVENTION

5 The invention relates to the field of molecular biology and recombinant nucleic acid technology. In particular, the invention relates to methods of inserting genes into defined locations in the chromosomal DNA of cultured mammalian cell lines which are subject to gene amplification. The invention also relates to meganucleases, vectors and engineered cell lines necessary for performing the methods, cell lines resulting from the
10 application of the methods, and use of the cell lines to produce protein products of interest.

BACKGROUND OF THE INVENTION

Therapeutic proteins are the primary growth driver in the global pharmaceutical market (Kresse, *Eur J Pharm Biopharm* **72**, 479 (2009)). In 2001, biopharmaceuticals
15 accounted for \$24.3 billion in sales. By 2007, this number had more than doubled to \$54.5 billion. The market is currently estimated to reach \$78 billion by 2012 (Pickering, *Spectrum Pharmaceutical Industry Dynamics Report, Decision Resources, Inc.*, 5 (2008)). This includes sales of “blockbuster” drugs such as erythropoietin, tissue plasminogen activator, and interferon, as well as numerous “niche” drugs such as enzyme replacement therapies for
20 lysosomal storage disorders. The unparalleled growth in market size, however, is driven primarily by skyrocketing demand for fully human and humanized monoclonal antibodies (Reichert, *Curr Pharm Biotechnol* **9**, 423 (2008)). Because they have the ability to confer a virtually unlimited spectrum of biological activities, monoclonal antibodies are quickly becoming the most powerful class of therapeutics available to physicians. Not surprisingly,
25 more than 25% of the molecules currently undergoing clinical trials in the United States and Europe are monoclonal antibodies (Reichert, *Curr Pharm Biotechnol* **9**, 423 (2008)).

Unlike more traditional pharmaceuticals, therapeutic proteins are produced in living cells. This greatly complicates the manufacturing process and introduces significant heterogeneity into product formulations (Field, *Recombinant Human IgG Production from*
30 *Myeloma and Chinese Hamster Ovary Cells*, in Cell Culture and Upstream Processing, Butler, ed., (Taylor and Francis Group, New York, 2007)). In addition, protein drugs are typically required at unusually high doses, which necessitates highly scalable manufacturing processes and makes manufacturing input costs a major price determinant. For these reasons, treatment with a typical therapeutic antibody (*e.g.*, the anti-HER2-neu monoclonal

Herceptin®) costs \$60,000-\$80,000 for a full course of treatment (Fleck, *Hastings Center Report* **36**, 12 (2006)). Further complicating the economics of biopharmaceutical production is the fact that many of the early blockbuster biopharmaceuticals are off-patent (or will be off-patent soon) and the US and EU governments are expected to greatly streamline the regulatory approval process for “biogeneric” and “biosimilar” therapeutics (Kresse, *Eur J Pharm Biopharm* **72**, 479 (2009)). These factors should lead to a significant increase in competition for sales of many prominent biopharmaceuticals (Pickering, *Spectrum Pharmaceutical Industry Dynamics Report, Decision Resources, Inc.*, 5 (2008)). Therefore, there is enormous interest in technologies which reduce manufacturing costs of protein therapeutics (Seth *et al.*, *Curr Opin Biotechnol* **18**, 557 (2007)).

Many of the protein pharmaceuticals on the market are glycoproteins that cannot readily be produced in easy-to-manipulate biological systems such as bacteria or yeast. For this reason, recombinant therapeutic proteins are produced almost exclusively in mammalian cell lines, primarily Chinese hamster ovary (*e.g.*, CHO-K1), mouse myeloma (*e.g.*, NS0), baby hamster kidney (BHK), murine C127, human embryonic kidney (*e.g.*, HEK-293), or human retina-derived (*e.g.*, PER-C6) cells (Andersen and Krummen, *Curr Opin Biotechnol* **13**, 117 (2002)). Of these, CHO cells are, by far, the most common platform for bioproduction because they offer the best combination of high protein expression levels, short doubling time, tolerance to a wide range of media conditions, established transfection and amplification protocols, an inability to propagate most human pathogens, a paucity of blocking intellectual property, and the longest track record of FDA approval (Field, *Recombinant Human IgG Production from Myeloma and Chinese Hamster Ovary Cells, in Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)).

Large-market biopharmaceuticals are typically produced in enormous stirred-tank bioreactors containing hundreds of liters of CHO cells stably expressing the protein product of interest (Chu and Robinson, *Curr Opin Biotechnol* **12**, 180 (2001), Coco-Martin and Harmsen, *Bioprocess International* **6**, 28 (2008)). Under optimized industrial conditions, such manufacturing processes can yield in excess of 5g of protein per liter of cells per day (Coco-Martin and Harmsen, *Bioprocess International* **6**, 28 (2008)). Because of the large number of cells involved (~50 billion cells per liter), the level of protein expression per cell has a very dramatic effect on yield. For this reason, all of the cells involved in the production of a particular biopharmaceutical must be derived from a single “high-producer” clone, the production of which constitutes one of the most time- and resource-intensive steps in the

manufacturing process (Clarke and Compton, *Bioprocess International* **6**, 24 (2008)).

The first step in the large-scale manufacture of a biopharmaceutical is the transfection of mammalian cells with plasmid DNA encoding the protein product of interest under the control of a strong constitutive promoter. Stable transfectants are selected by using
5 a selectable marker gene also carried on the plasmid. Most frequently, this marker is a dihydrofolate reductase (DHFR) gene which, when transfected into a DHFR deficient cell line such as DG44, allows for the selection of stable transfectants using media deficient in hypoxanthine. The primary reason for using DHFR as a selectable marker is that it enables a process called “gene amplification”. By growing stable transfectants in gradually increasing
10 concentrations of methotrexate (MTX), a DHFR inhibitor, it is possible to amplify the number of copies of the DHFR gene present in the genome. Because the gene encoding the protein product of interest is physically coupled to the DHFR gene, this results in amplification of both genes with a concomitant increase in the expression level of the therapeutic protein (Butler, *Cell Line Development for Culture Strategies: Future Prospects to Improve Yields*, in Cell Culture and Upstream Processing, Butler, ed., (Taylor and Francis
15 Group, New York, 2007)). Related systems for the creation of stable bioproduction lines use the glutamine synthetase (GS) or hypoxanthine phosphoribosyltransferase (HPRT) genes as selectable markers and require the use of GS- or HPRT- deficient cell lines as hosts for transfection (Clarke and Compton, *Bioprocess International* **6**, 24 (2008)). In the case of the
20 GS system, gene amplification is accomplished by growing cells in the presence of methionine sulphoximine (MSX) (Clarke and Compton, *Bioprocess International* **6**, 24 (2008)). In the case of the HPRT system, gene amplification is accomplished by growing cells in HAT medium, which contains aminopterin, hypoxanthine, and thymidine (Kellems, ed. *Gene amplification in mammalian cells: a comprehensive guide*, Marcel Dekker, New
25 York, 1993).

In all of these systems, the initial plasmid DNA comprising a biotherapeutic gene expression cassette and a selectable marker integrates into a random location in the genome, resulting in extreme variability in therapeutic protein expression from one stable transfectant to another (Collingwood and Urnov, *Targeted Gene Insertion to Enhance Protein
30 Production from Cell Lines*, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). For this reason, it is necessary to screen hundreds to thousands of initial transfectants to identify cells which express acceptably high levels of gene product both before and after gene amplification (Butler, *Cell Line Development for Culture Strategies: Future Prospects to Improve Yields*, in Cell Culture and Upstream

Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). A second and more problematic consequence of random gene integration is the phenomenon of transgene silencing, in which recombinant protein expression slows or ceases entirely over time (Collingwood and Urnov, *Targeted Gene Insertion to Enhance Protein Production from Cell Lines*, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). Because these effects often do not manifest themselves for weeks to months following the initial transfection and screening process, it is generally necessary to carry and expand dozens of independent clonal lines to identify one that expresses the protein of interest consistently over time (Butler, *Cell Line Development for Culture Strategies: Future Prospects to Improve Yields*, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)).

This large number of screening and expansion steps results in a very lengthy and expensive process to simply generate the cell line that will, ultimately, produce the therapeutic of interest. Indeed, using conventional methods, a minimum of 10 months (with an average of 18 months) and an upfront investment of tens of millions of dollars in labor and material is required to produce an initial pool of protein-expressing cells suitable for industrial manufacturing (Butler, *Cell Line Development for Culture Strategies: Future Prospects to Improve Yields*, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). If one takes into account lost time on market for a blockbuster protein therapeutic, inefficiencies in cell line production can cost biopharmaceutical manufacturers hundreds of millions of dollars (Seth *et al.*, *Curr Opin Biotechnol* **18**, 557 (2007)).

Much of the time and expense of bioproduction cell line creation can be attributed to random genomic integration of the bioproduct gene resulting in clone-to-clone variability in genotype and, hence, variability in gene expression. One way to overcome this is to target gene integration to a defined location that is known to support a high level of gene expression. To this end, a number of systems have been described which use the Cre, Flp, or Φ C31 recombinases to target the insertion of a bioproduct gene (reviewed in Collingwood and Urnov, *Targeted Gene Insertion to Enhance Protein Production from Cell Lines*, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). Recent embodiments of these systems, most notably the Flp-In[®] system marketed by Invitrogen Corp. (Carlsbad, CA), couple bioproduct gene integration with the reconstitution of a split selectable marker so that cells with correctly targeted genes can be selected. As expected, these systems result in greatly reduced heterogeneity in gene expression and, in

some cases, individual stable transfectants can be pooled, obviating the time and expense associated with expanding a single clone.

The principal drawback to recombinase-based gene targeting systems is that the recombinase recognition sites (loxP, FRT, or attB/attP sites) do not naturally occur in mammalian genomes. Therefore, cells must be pre-engineered to incorporate a recognition site for the recombinase before that site can be subsequently targeted for gene insertion. Because the recombinase site itself integrates randomly into the genome, it is still necessary to undertake extensive screening and evaluation to identify clones which carry the site at a location that is suitable for high level, long-term gene expression (Collingwood and Urnov, *Targeted Gene Insertion to Enhance Protein Production from Cell Lines*, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). In addition, the biomanufacturing industry is notoriously hesitant to adopt “new” cell lines, such as those that have been engineered to carry a recombinase site, that do not have a track record of FDA approval. For these reasons, recombinase-based cell engineering systems may not readily be adopted by the industry and an approach that allows biomanufacturers to utilize their existing cell lines is preferable.

SUMMARY OF THE INVENTION

The present invention depends, in part, upon the development of mammalian cell lines in which sequences of interest (*e.g.*, exogenous, actively transcribed transgenes) are inserted proximal to an endogenous selectable gene in an amplifiable locus, and the discovery that (a) the insertion of such exogenous sequences of interest does not inhibit amplification of the endogenous selectable gene, (b) the exogenous sequence of interest can be co-amplified with the endogenous selectable gene, and (c) the resultant cell lines, with an amplified region comprising multiple copies of the endogenous selectable gene and the exogenous sequence of interest, are stable for extended periods even in the absence of the selection regime which was employed to induce amplification. Thus, in one aspect, the invention provides a method for producing cell lines which can be used for biomanufacturing of a protein product of interest by specifically targeting the insertion of an exogenous sequence of interest capable of actively expressing the protein product of interest proximal to an endogenous selectable gene. In another aspect, the invention provides engineered cell lines that can be used to produce protein products of interest (*e.g.*, therapeutic proteins such as monoclonal antibodies) at high levels.

It is understood that any of the embodiments described below can be combined in any desired way, and any embodiment or combination of embodiments can be applied to each of the aspects described below, unless the context indicates otherwise.

In one aspect, the invention provides a recombinant mammalian cell comprising an engineered target site stably integrated within selectable gene within an amplifiable locus,
5 wherein the engineered target site disrupts the function of the selectable gene and wherein the engineered target site comprises a recognition sequence for a site specific endonuclease.

In some embodiments, the selectable gene is glutamine synthetase (GS) and the locus is methionine sulphoximine (MSX) amplifiable. In some embodiments, the selectable gene is
10 dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amplifiable.

In some embodiments, the selectable gene is selected from the group consisting of Dihydrofolate Reductase, Glutamine Synthetase, Hypoxanthine Phosphoribosyltransferase, Threonyl tRNA Synthetase, Na,K-ATPase, Asparagine Synthetase, Ornithine Decarboxylase, Inosine-5'-monophosphate dehydrogenase, Adenosine Deaminase, Thymidylate Synthetase,
15 Aspartate Transcarbamylase, Metallothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.

In some embodiments, the selectable gene is amplifiable by selection with a selection agent selected from the group consisting of Methotrexate (MTX), Methionine sulphoximine (MSX), Aminopterin, hypoxanthine, thymidine, Borrelidin, Ouabain, Albizziin, Beta-aspartyl
20 hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2'deoxycoformycin, Fluorouracil, N-Phosphonacetyl-L-Aspartate (PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-azauridine, pyrazofuran, hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine and triapine.

In some embodiments, the engineered target site is inserted into an exon of the
25 selectable gene. In some embodiments, the site specific endonuclease is a meganuclease, a zinc finger nuclease or TAL effector nuclease. In some embodiment, the recombinant cell further comprises the site specific endonuclease.

In one aspect, the invention provides a recombinant mammalian cell comprising an engineered target site stably integrated proximal to a selectable gene within an amplifiable

locus, wherein the engineered target site comprises a recognition sequence for a site specific endonuclease.

In some embodiments, the engineered target site is downstream from the 3' regulatory region of the selectable gene. In some embodiments, the engineered target site is 0 to 100,000 base pairs downstream from the 3' regulatory region of the selectable gene. In other
5 embodiments, the engineered target site is upstream from the 5' regulatory region of the selectable gene. In some embodiments, the engineered target site is 0 to 100,000 base pairs upstream from the 5' regulatory region of the selectable gene.

In another aspect, the invention provides a method for inserting an exogenous
10 sequence into an amplifiable locus of a mammalian cell comprising: (a) providing a mammalian cell having an endogenous target site proximal to a selectable gene within the amplifiable locus, wherein the endogenous target site comprises: (i) a recognition sequence for an engineered meganuclease; (ii) a 5' flanking region 5' to the recognition sequence; and
15 (iii) a 3' flanking region 3' to the recognition sequence; and (b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site; (c) contacting the cell with a donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site; (ii) an exogenous sequence; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the
20 endogenous target site; whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a modified cell.

In some embodiments, the method further comprises growing the modified cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene. In some embodiments, the exogenous sequence comprises a
25 gene of interest.

In some embodiments endogenous target site is downstream from the 3' regulatory region of the selectable gene. In some embodiments, the endogenous target site is 0 to 100,000 base pairs downstream from the 3' regulatory region of the selectable gene. In other
30 embodiments, the endogenous target site is upstream from the 5' regulatory region of the selectable gene. In some embodiments, the endogenous target site is 0 to 100,000 base pairs upstream from the 5' regulatory region of the selectable gene.

In one aspect, the invention provides a method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising: (a) providing a mammalian cell having an endogenous target site proximal to a selectable gene within the amplifiable locus, wherein the endogenous target site comprises: (i) a recognition sequence for an engineered
5 meganuclease; (ii) a 5' flanking region 5' to the recognition sequence; and (iii) a 3' flanking region 3' to the recognition sequence; and (b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site; (c) contacting the cell with an engineered target site donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site; (ii) an exogenous
10 sequence comprising an engineered target site; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site; whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a mammalian cell comprising the engineered target site; (d)
15 introducing a double-stranded break between the 5' and 3' flanking regions of the engineered target site; (e) contacting the cell comprising the engineered target site with a sequence of interest donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the engineered target site; (ii) an exogenous sequence comprising a sequence of interest; and (iii) a donor 3' flanking region homologous to the 3' flanking region
20 of the engineered target site; whereby the donor 5' flanking region, the exogenous sequence comprising the sequence of interest and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the engineered target site by homologous recombination to provide an engineered mammalian cell comprising the sequence of interest.

In some embodiments, the method further comprises growing the engineered
25 mammalian cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene. In some embodiments, the sequence of interest comprises a gene.

In another aspect, the invention provides a method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising: (a) providing a
30 mammalian cell having an endogenous target site within a selectable gene within the amplifiable locus, wherein the endogenous target site comprises: (i) a recognition sequence for an engineered meganuclease; (ii) a 5' flanking region 5' to the recognition sequence; and

(iii) a 3' flanking region 3' to the recognition sequence; and (b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site; (c) contacting the cell with an engineered target site donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site; (ii) an exogenous sequence comprising an engineered target site; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site; whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a mammalian cell comprising the engineered target site; (d) introducing a double-stranded break between the 5' and 3' flanking regions of the engineered target site; (e) contacting the cell comprising the engineered target site with a sequence of interest donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the engineered target site; (ii) an exogenous sequence comprising a sequence of interest; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the engineered target site; whereby the donor 5' flanking region, the exogenous sequence comprising the sequence of interest and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the engineered target site by homologous recombination to provide a engineered mammalian cell comprising the sequence of interest.

In some embodiments, the method further comprises growing the engineered mammalian cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene.

In some embodiments, the sequence of interest comprises a gene.

In some embodiments, the endogenous target site is within an intron of the selectable gene. In other embodiments, the endogenous target site is within an exon of the selectable gene.

In one aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 15.

In another aspect, the invention provides a recombinant meganuclease comprising the amino acid sequence of SEQ ID NO: 15.

In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 14. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 14.

5 In another aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 9. In one embodiment, the recombinant meganuclease has the sequence of the meganuclease of SEQ ID NO: 9.

10 In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 7. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 7.

15 In another aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 10. In one embodiment, the recombinant meganuclease comprises the polypeptide of SEQ ID NO: 10.

20 In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 8. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 8.

In another aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 13. In one embodiment, the recombinant meganuclease comprises the polypeptide of SEQ ID NO: 13.

25 In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 12. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 12.

30 In another aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 29. In one embodiment, the recombinant meganuclease comprises the polypeptide of SEQ ID NO: 29.

In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 30. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 30.

5 In another aspect, the invention provides recombinant mammalian cell lines which continue to express a protein product of interest from an exogenous sequence of interest present in an amplified region of the genome (*i.e.*, present in 2-1,000 copies, co-amplified with a selectable gene in an amplifiable locus) for a period of at least 8, 9, 10, 11, 12, 13, or 14 weeks after removal of the amplification selection agent, and with a reduction of
10 expression levels and/or copy number of less than 20, 25, 30, 35 or 40%.

In another aspect, the invention provides methods of producing recombinant cells with amplified regions including a sequence of interest and a selectable gene by subjecting the above-described recombinant cells to selection with a selection agent which causes co-amplification of the sequence of interest and the selectable gene.

15 In another aspect, the invention provides methods of producing a protein product of interest by culturing the above-described recombinant cells, or the above-described recombinant cells with amplified regions, and obtaining the protein product of interest from the culture medium or a cell lysate.

BRIEF DESCRIPTION OF THE FIGURES

20 Figure 1. A general strategy for targeting a sequence of interest to an amplifiable locus.

Figure 2. (A) Schematic of the CHO DHFR locus showing a preferred region for targeting a sequence of interest 5,000-60,000 base pairs downstream of the DHFR gene. (B) Schematic of the CHO GS locus showing a preferred region for targeting a sequence of
25 interest 5,000-55,000 base pairs downstream of the GS gene.

Figure 3. Strategy for inserting a sequence of interest into an amplifiable locus in a two-step process involving a pre-integrated engineered target sequence.

Figure 4. Strategy for inserting an engineered target sequence into an amplifiable locus with concomitant removal of a portion of the selectable gene, followed by insertion of a
30 sequence of interest and reconstitution of the selectable gene.

Figure 5. Strategy for inserting an engineered target sequence into an amplifiable locus with concomitant disruption of the coding sequence of a selectable gene, followed by insertion of a sequence of interest and reconstitution of the selectable gene.

Figure 6. Strategy for inserting an engineered target sequence into an amplifiable locus with concomitant disruption of the mRNA processing, followed by insertion of a sequence of interest and reconstitution of the selectable gene.

Figure 7. (A) A direct-repeat recombination assay for site-specific endonuclease activity. (B) Results of the assay in (A) applied to the CHO-23/24 and CHO-51/52 meganucleases. (C) Alignment of sequences obtained from CHO cells transfected with mRNA encoding the CHO-23/24 meganuclease. (D) Alignment of sequences obtained from CHO cells transfected with mRNA encoding the CHO-51/52 meganuclease.

Figure 8. (A) Strategy for inserting an exogenous DNA sequence into the CHO DHFR locus using the CHO-51/52 meganuclease. (B) PCR products demonstrating insertion of an engineered target sequence.

Figure 9. (A) Strategy for inserting an engineered target sequence into the CHO DHFR locus using the CHO-23/24 meganuclease, followed by Flp recombinase-mediated insertion of a sequence of interest. (B) PCR products from hygromycin-resistant clones produced in (A). (C) GFP expression by the 24 clones produced in (B).

Figure 10. Results of experiments with a GFP-expressing CHO line produced by integrating a GFP gene expression cassette into the DHFR locus using a target sequence strategy as shown in Figure 9.

Figure 11. (A) A direct-repeat recombination assay, as in Figure 5A. (B) The assay in (A) applied to the CHO-13/14 and CGS-5/6 meganucleases. (C) Alignment of sequences obtained from CHO cells transfected with mRNA encoding the CGS-5/6 meganuclease.

DETAILED DESCRIPTION OF THE INVENTION

1.1 Introduction

The present invention depends, in part, upon the development of mammalian cell lines in which exogenous actively transcribed transgenes have been inserted proximal to an endogenous amplifiable locus, and the discovery that (a) the insertion of such exogenous actively transcribed transgenes does not prevent or substantially inhibit amplification of the endogenous amplifiable locus, (b) the exogenous actively transcribed transgene can be co-amplified with the endogenous amplifiable locus, and (c) the resultant cell line, with an amplified region comprising multiple copies of the endogenous amplifiable locus and the exogenous actively transcribed transgene is stable for extended periods even in the absence of the selection regime which was employed to induce amplification. Thus, in one aspect, the

invention provides a method for producing cell lines which can be used for biomanufacturing of a protein product of interest by specifically targeting the insertion of an exogenous gene capable of actively expressing the protein product of interest proximal to an endogenous amplifiable locus. In another aspect, the invention provides engineered cell lines that can be used to produce protein products of interest (*e.g.*, therapeutic proteins such as monoclonal antibodies) at high levels.

1.2 References and Definitions

The patent and scientific literature referred to herein establishes knowledge that is available to those of skill in the art. The entire disclosures of the issued U.S. patents, pending applications, published foreign applications, and scientific and technical references cited herein, including protein and nucleic acid database sequences, are hereby incorporated by reference to the same extent as if each was specifically and individually indicated to be incorporated by reference.

As used herein, the term “meganuclease” refers to naturally-occurring homing endonucleases (also referred to as Group I intron encoded endonucleases) or non-naturally-occurring (*e.g.*, rationally designed or engineered) endonucleases based upon the amino acid sequence of a naturally-occurring homing endonuclease. Examples of naturally-occurring meganucleases include I-SceI, I-CreI, I-CeuI, I-DmoI, I-MsoI, I-AniI, etc. Rationally designed meganucleases are disclosed in, for example, WO 2007/047859 and WO 2009/059195, and can be engineered to have modified DNA-binding specificity, DNA cleavage activity, DNA-binding affinity, or dimerization properties relative to a naturally occurring meganuclease. A meganuclease may bind to double-stranded DNA as a homodimer (*e.g.*, wild-type I-CreI), or it may bind to DNA as a heterodimer (*e.g.*, engineered meganucleases disclosed in WO 2007/047859). An engineered meganuclease may also be a “single-chain meganuclease” in which a pair of DNA-binding domains derived from a natural meganuclease are joined into a single polypeptide using a peptide linker (*e.g.*, single-chain meganucleases disclosed in WO 2009/059195).

As used herein, the term “single-chain meganuclease” refers to a polypeptide comprising a pair of meganuclease subunits joined by a linker. A single-chain meganuclease has the organization: N-terminal subunit – Linker – C-terminal subunit. The two meganuclease subunits will generally be non-identical in amino acid sequence and will recognize non-identical DNA sequences. Thus, single-chain meganucleases typically cleave pseudo-palindromic or non-palindromic recognition sequences. Methods of producing

single-chain meganucleases are disclosed in WO 2009/059195.

As used herein, the term “site specific endonuclease” means a meganuclease, zinc-finger nuclease or TAL effector nuclease.

As used herein, with respect to a protein, the term “recombinant” means having an altered amino acid sequence as a result of the application of genetic engineering techniques to nucleic acids which encode the protein, and cells or organisms which express the protein. With respect to a nucleic acid, the term “recombinant” means having an altered nucleic acid sequence as a result of the application of genetic engineering techniques. Genetic engineering techniques include, but are not limited to, PCR and DNA cloning technologies; transfection, transformation and other gene transfer technologies; homologous recombination; site-directed mutagenesis; and gene fusion. In accordance with this definition, a protein having an amino acid sequence identical to a naturally-occurring protein, but produced by cloning and expression in a heterologous host, is not considered recombinant. As used herein, the term “engineered” is synonymous with the term “recombinant.”

As used herein, with respect to a meganuclease, the term "wild-type" refers to any naturally-occurring form of a meganuclease. The term "wild-type" is not intended to mean the most common allelic variant of the enzyme in nature but, rather, any allelic variant found in nature. Wild-type homing endonucleases are distinguished from recombinant or non-naturally-occurring meganucleases.

As used herein, the term "recognition sequence" refers to a DNA sequence that is bound and cleaved by a meganuclease. A recognition sequence comprises a pair of inverted, 9 base pair “half sites” which are separated by four base pairs. In the case of a homo- or heterodimeric meganucleases, each of the two monomers makes base-specific contacts with one half-site. In the case of a single-chain heterodimer meganuclease, the N-terminal domain of the protein contacts a first half-site and the C-terminal domain of the protein contacts a second half-site. In the case of I-CreI, for example, the recognition sequence is 22 base pairs and comprises a pair of inverted, 9 base pair “half sites” which are separated by four base pairs.

As used herein, the term “target site” refers to a region of the chromosomal DNA of a cell comprising a target sequence into which a sequence of interest can be inserted. As used herein, the term “engineered target site” refers to an exogenous sequence of DNA integrated into the chromosomal DNA of a cell comprising an engineered target sequence into which a sequence of interest can be inserted.

As used herein, the term “target sequence” means a DNA sequence within a target site which includes one or more recognition sequences for a nuclease, integrase, transposase, and/or recombinase. For example, a target sequence can include a recognition sequence for a meganuclease. As used herein, an “engineered target sequence” means an exogenous target sequence which is introduced into a chromosome to serve as the insertion point for another sequence.

As used herein, the term “flanking region” or “flanking sequence” refers to a sequence of >3 or, preferably, >50 or, more preferably, >200 or, most preferably, >400 base pairs of DNA which is immediately 5' or 3' to a reference sequence (*e.g.*, a target sequence or sequence of interest).

As used herein, the terms “amplifiable locus” refers to a region of the chromosomal DNA of a cell which can be amplified by selection with one or more compounds (*e.g.*, drugs) in the growth media. An amplifiable locus will typically comprise a gene encoding a protein which, under the appropriate conditions, is necessary for cell survival. By inhibiting the function of such an essential protein, for example with a small molecule drug, the amplifiable locus is duplicated many times over as a means of increasing the copy number of the essential gene. A gene of interest, if integrated into an amplifiable locus, will also become duplicated with the essential gene. Examples of amplifiable loci include the chromosomal regions comprising the DHFR, GS, and HPRT genes.

As used herein, the term “amplified locus” or “amplified gene” or “amplified sequence” refers to a locus, gene or sequence which is present in 2-1,000 copies as a result of gene amplification in response to selection of a selectable gene. An amplified gene or sequence can be a gene or sequence which is co-amplified due to selection of a selectable gene in the same amplifiable locus. In preferred embodiments, a sequence of interest is amplified to at least 3, 4, 5, 6, 7, 8, 9 or 10 copies.

As used herein, the term “selectable gene” refers to an endogenous gene that is essential for cell survival under some specific culture conditions (*e.g.*, presence or absence of a nutrient, toxin or drug). Selectable genes are endogenous to the cell and are distinguished from exogenous “selectable markers” such as antibiotic resistance genes. Selectable genes exist in their natural context in the chromosomal DNA of the cell. For example, DHFR is a selectable gene which is necessary for cell survival in the presence of MTX in the culture medium. The gene is essential for growth in the absence of hypoxanthine and thymidine. If the endogenous DHFR selectable gene is eliminated, cells are able to grow in the absence of hypoxanthine and thymidine if they are given an exogenous copy of the DHFR gene. This

exogenous copy of the DHFR gene is a selectable marker but is not a selectable gene. An amplifiable locus comprises a selectable gene and a target site. A target site is found outside of a selectable gene such that a selectable gene does not comprise a target site. Examples of selectable genes are given in Table 1.

5 As used herein, when used in connection with the position of a target site, recognition sequence, or inserted sequence of interest relative to the position of a selectable gene, the term “proximal” means that the target site, recognition sequence, or inserted sequence of interest is within the same amplifiable locus as the selectable gene, either upstream (5') or downstream (3') of the selectable gene, and preferably between the selectable
10 gene and the next gene in the region (whether upstream (5') or downstream (3')). Typically, a “proximal” target site, recognition sequence, or inserted sequence of interest will be within <100,000 base pairs of the selectable gene, as measured from the first or last nucleotide of the first or last regulatory element of the selectable gene.

As used herein, the term “homologous recombination” refers to the natural,
15 cellular process in which a double-stranded DNA-break is repaired using a homologous DNA sequence as the repair template (see, *e.g.* Cahill *et al.* (2006), *Front. Biosci.* 11:1958-1976). The homologous DNA sequence may be an endogenous chromosomal sequence or an exogenous nucleic acid that was delivered to the cell. Thus, for some applications of engineered meganucleases, a meganuclease is used to cleave a recognition sequence within a
20 target sequence in a genome and an exogenous nucleic acid with homology to or substantial sequence similarity with the target sequence is delivered into the cell and used as a template for repair by homologous recombination. The DNA sequence of the exogenous nucleic acid, which may differ significantly from the target sequence, is thereby inserted or incorporated into the chromosomal sequence. The process of homologous recombination occurs primarily
25 in eukaryotic organisms. The term “homology” is used herein as equivalent to “sequence similarity” and is not intended to require identity by descent or phylogenetic relatedness.

As used herein, the term “stably integrated” means that an exogenous or heterologous DNA sequence has been covalently inserted into a chromosome (*e.g.*, by homologous recombination, non-homologous end joining, transposition, etc.) and has
30 remained in the chromosome for a period of at least 8 weeks.&&

As used herein, the term “non-homologous end-joining” or “NHEJ” refers to the natural, cellular process in which a double-stranded DNA-break is repaired by the direct joining of two non-homologous DNA segments (see, *e.g.* Cahill *et al.* (2006), *Front. Biosci.* 11:1958-1976). DNA repair by non-homologous end-joining is error-prone and frequently

results in the untemplated addition or deletion of DNA sequences at the site of repair. Thus, for certain applications, an engineered meganuclease can be used to produce a double-stranded break at a meganuclease recognition sequence within an amplifiable locus and an exogenous nucleic acid molecule, such as a PCR product, can be captured at the site of the DNA break by NHEJ (see, *e.g.* Salomon *et al.* (1998), *EMBO J.* 17:6086-6095). In such cases, the exogenous nucleic acid may or may not have homology to the target sequence. The process of non-homologous end-joining occurs in both eukaryotes and prokaryotes such as bacteria.

As used herein, the term “sequence of interest” means any nucleic acid sequence, whether it codes for a protein, RNA, or regulatory element (*e.g.*, an enhancer, silencer, or promoter sequence), that can be inserted into a genome or used to replace a genomic DNA sequence. Sequences of interest can have heterologous DNA sequences that allow for tagging a protein or RNA that is expressed from the sequence of interest. For instance, a protein can be tagged with tags including, but not limited to, an epitope (*e.g.*, c-myc, FLAG) or other ligand (*e.g.*, poly-His). Furthermore, a sequence of interest can encode a fusion protein, according to techniques known in the art (see, *e.g.*, Ausubel *et al.*, Current Protocols in Molecular Biology, Wiley 1999). In preferred embodiments, a sequence of interest comprises a promoter operably linked to a gene encoding a protein of medicinal value such as an antibody, antibody fragment, cytokine, growth factor, hormone, or enzyme. For some applications, the sequence of interest is flanked by a DNA sequence that is recognized by the engineered meganuclease for cleavage. Thus, the flanking sequences are cleaved allowing for proper insertion of the sequence of interest into genomic recognition sequences cleaved by an engineered meganuclease. For some applications, the sequence of interest is flanked by DNA sequences with homology to or substantial sequence similarity with the target site such that homologous recombination inserts the sequence of interest within the genome at the locus of the target sequence.

As used herein, the term “donor DNA” refers to a DNA molecule comprising a sequence of interest flanked by DNA sequences homologous to a target site. Donor DNA can serve as a template for DNA repair by homologous recombination if it is delivered to a cell with a site-specific nuclease such as a meganuclease, zinc-finger nuclease, or TAL-effector nuclease. The result of such DNA repair is the insertion of the sequence of interest into the chromosomal DNA of the cell. Donor DNA can be linear, such as a PCR product, or circular, such as a plasmid. In cases where a donor DNA is a circular plasmid, it may be referred to as a “donor plasmid.”

As used herein, unless specifically indicated otherwise, the word "or" is used in the inclusive sense of "and/or" and not the exclusive sense of "either/or."

2.1 Transgene Targeting to Amplifiable Loci

5 The present invention provides methods for generating transgenic mammalian cell lines expressing a desired protein product of interest, including "high-producer" cell lines, by targeting the insertion of a gene encoding the protein product of interest (*e.g.*, a therapeutic protein gene expression cassette) to regions of the genome that are amplifiable. Such regions in mammalian cells include the DHFR, GS, and HPRT genes, as well as others shown in
10 Table 1.

 The precise mechanism of gene amplification is not known. Indeed, it is very likely that there is no single mechanism by which gene amplification occurs but that a variety of different random chromosomal aberrations, in combination with strong selection for amplification, results in increased gene copy number (reviewed in Omasa (2002), *J. Biosci. Bioeng.* 94:600-605). It is clear that chromosomal location plays a major role in
15 amplification and the stable maintenance of amplified genes (Brinton and Heintz (1995), *Chromosoma* 104:143-51). It has been found that transgenes integrated into chromosomal locations adjacent to telomeres are more easily amplified and, once amplified, tend to be stable at high copy numbers after the selection agent is removed (Yoshikawa *et al.* (2000),
20 *Cytotechnology* 33:37-46; Yoshikawa *et al.* (2000), *Biotechnol Prog.* 16:710-715). This is significant because selection agents such as MTX and MSX are toxic and cannot be included in the growth media in a commercial biomanufacturing process. In contrast, transgenes integrated into regions in the CHO genome that are not adjacent to telomeres amplify inefficiently and rapidly lose copy number following the removal of selection agents from the
25 media. For example, Yoshikawa *et al.* found that randomly-integrated transgenes linked to a DHFR selectable marker amplified to greater than 10-fold higher copy numbers when the integration site was adjacent to a telomere (Yoshikawa *et al.* (2000), *Biotechnol Prog.* 16:710-715). These researchers also found that an amplified transgene integrated into a non-telomeric region will lose >50% of its copies in only 20 days following the removal of MTX
30 from the growth media. None of the selectable genes identified in Table 1 is adjacent to a telomere in the mouse genome (www.ensembl.com) and the similarity in genome organization between mouse and CHO makes it likely that these genes are in non-telomeric regions in CHO as well (Xu *et al.* (2011), *Nat. Biotechnol.* 29:735-741). Thus, the prior art instructs that the loci identified in Table 1, including the DHFR and GS loci, are not preferred

locations to target transgene insertion if the goal is efficient and stable gene amplification.

In addition, in the case of endogenous gene amplification, it is clear that chromosomal sequences outside of the selectable gene sequence play an important role in facilitating amplification and in defining the length of DNA sequence that is co-amplified with the gene under selection (Looney and Hamlin (1987), *Mol. and Cell. Biol.* 7:569-577). In particular, it has been shown that the sequence and location of the DNA replication origin in relation to the selectable gene plays a major role in amplification. For example, it has been shown that amplification of the endogenous CHO DHFR locus is dependent upon a pair of replication origins found in the region 5,000-60,000 base pairs downstream of the DHFR gene coding sequence (Anachkova and Hamlin (1989), *Mol. and Cell. Biol.* 9:532-540; Milbrandt *et al.* (1981), *Proc. Natl. Acad. Sci. USA* 78:6042-6047). Further, Brinton and Heintz have shown that these same replication origins fail to promote gene amplification when incorporated randomly into the genome with a transgenic DHFR sequence (Brinton and Heintz (1995), *Chromosoma*. 104:143-51). This clearly demonstrates the importance of maintaining both the sequence and proper chromosomal context of these replication origins to promote DHFR gene amplification. Thus the art instructs that the region downstream of DHFR is critical to gene amplification and should not be disrupted by, for example, inserting a transgenic gene expression cassette as described in the present invention.

Surprisingly, we have discovered that DNA sequences, including exogenous transcriptionally active sequences, which are inserted proximal to (*e.g.*, within <100,000 base pairs) selectable genes in mammalian cell lines (*e.g.*, CHO-K1) will co-amplify in the presence of appropriate compounds which select for amplification. Thus, the present invention provides methods for reliably and reproducibly producing isogenic cell lines in which transgenes encoding protein products of interest (*e.g.*, biotherapeutic gene expression cassettes) can be amplified but in which it is not necessary to screen a large number of randomly generated cell lines to identify those which express high levels of the protein product of interest and are resistant to gene silencing.

In addition, we have surprisingly found that the mammalian cell lines of the invention, in which a sequence of interest is co-amplified with a selectable gene in an amplifiable locus, are stable with respect to expression of the sequence of interest and/or copy number of the sequence of interest even in the absence of continued selection. That is, whereas the art teaches that amplified sequences will be reduced in copy number over time if selection is not maintained (see, *e.g.*, Yoshikawa *et al.* (2000), *Biotechnol Prog.* 16:710-715), we have found that cell lines produced according to the methods of the invention continue to

produce the protein products of interest (encoded by the sequences of interest) at levels within 20%-25% of the initial levels, even 14 weeks after removal of the selection agent. This is significant, as noted above, because selection agents such as MTX and MSX are toxic, and it would be highly desirable to produce biotherapeutic proteins in cell lines which do not require continued exposure to such selection agents. Therefore, in some embodiments, the invention provides recombinant mammalian cell lines which continue to express a protein product of interest from an exogenous sequence of interest present in an amplified region of the genome (*i.e.*, present in 2-1,000 copies, co-amplified with a selectable gene in an amplifiable locus) for a period of at least 8, 9, 10, 11, 12, 13, or 14 weeks after removal of the amplification selection agent, and with a reduction of expression levels and/or copy number of less than 20, 25, 30, 35 or 40%.

The present invention also provides the products necessary to practice the methods, and to target insertion of sequences of interest into amplifiable loci in mammalian cell lines. A common method for inserting or modifying a DNA sequence involves introducing a transgenic DNA sequence flanked by sequences homologous to the genomic target and selecting or screening for a successful homologous recombination event. Recombination with the transgenic DNA occurs rarely but can be stimulated by a double-stranded break in the genomic DNA at the target site (Porteus *et al.* (2005), *Nat. Biotechnol.* 23: 967-73; Tzfira *et al.* (2005), *Trends Biotechnol.* 23: 567-9; McDaniel *et al.* (2005), *Curr. Opin. Biotechnol.* 16: 476-83). Numerous methods have been employed to create DNA double-stranded breaks, including irradiation and chemical treatments. Although these methods efficiently stimulate recombination, the double-stranded breaks are randomly dispersed in the genome, which can be highly mutagenic and toxic. At present, the inability to target gene modifications to unique sites within a chromosomal background is a major impediment to routine genome engineering.

One approach to achieving this goal is stimulating homologous recombination at a double-stranded break in a target locus using a nuclease with specificity for a sequence that is sufficiently large to be present at only a single site within the genome (see, *e.g.*, Porteus *et al.* (2005), *Nat. Biotechnol.* 23: 967-73). The effectiveness of this strategy has been demonstrated in a variety of organisms using ZFNs (Porteus (2006), *Mol Ther* 13: 438-46; Wright *et al.* (2005), *Plant J.* 44: 693-705; Urnov *et al.* (2005), *Nature* 435: 646-51). Homing endonucleases are a group of naturally-occurring nucleases which recognize 15-40 base-pair cleavage sites commonly found in the genomes of plants and fungi. They are frequently associated with parasitic DNA elements, such as Group I self-splicing introns and

intains. They naturally promote homologous recombination or gene insertion at specific locations in the host genome by producing a double-stranded break in the chromosome, which recruits the cellular DNA-repair machinery (Stoddard (2006), *Q. Rev. Biophys.* 38: 49-95). Homing endonucleases are commonly grouped into four families: the LAGLIDADG family, the GIY-YIG family, the His-Cys box family and the HNH family. These families are characterized by structural motifs, which affect catalytic activity and recognition sequence. For instance, members of the LAGLIDADG family are characterized by having either one or two copies of the conserved LAGLIDADG motif (see Chevalier *et al.* (2001), *Nucleic Acids Res.* 29(18): 3757-3774). The LAGLIDADG homing endonucleases with a single copy of the LAGLIDADG motif form homodimers, whereas members with two copies of the LAGLIDADG motif are found as monomers.

Natural homing endonucleases, primarily from the LAGLIDADG family, have been used to effectively promote site-specific genome modification in plants, yeast, *Drosophila*, mammalian cells and mice, but this approach has been limited to the modification of either homologous genes that conserve the endonuclease recognition sequence (Monnat *et al.* (1999), *Biochem. Biophys. Res. Commun.* 255: 88-93) or to pre-engineered genomes into which a recognition sequence has been introduced (Rouet *et al.* (1994), *Mol. Cell. Biol.* 14: 8096-106; Chilton *et al.* (2003), *Plant Physiol.* 133: 956-65; Puchta *et al.* (1996), *Proc. Natl. Acad. Sci. USA* 93: 5055-60; Rong *et al.* (2002), *Genes Dev.* 16: 1568-81; Gouble *et al.* (2006), *J. Gene Med.* 8(5):616-622).

Systematic implementation of nuclease-stimulated gene modification requires the use of engineered enzymes with customized specificities to target DNA breaks to existing sites in a genome and, therefore, there has been great interest in adapting homing endonucleases to promote gene modifications at medically or biotechnologically relevant sites (Porteus *et al.* (2005), *Nat. Biotechnol.* 23: 967-73; Sussman *et al.* (2004), *J. Mol. Biol.* 342: 31-41; Epinat *et al.* (2003), *Nucleic Acids Res.* 31: 2952-62).

I-CreI (SEQ ID NO: 1) is a member of the LAGLIDADG family of homing endonucleases which recognizes and cuts a 22 base pair recognition sequence in the chloroplast chromosome of the algae *Chlamydomonas reinhardtii*. Genetic selection techniques have been used to modify the wild-type I-CreI cleavage site preference (Sussman *et al.* (2004), *J. Mol. Biol.* 342: 31-41; Chames *et al.* (2005), *Nucleic Acids Res.* 33: e178; Seligman *et al.* (2002), *Nucleic Acids Res.* 30: 3870-9, Arnould *et al.* (2006), *J. Mol. Biol.* 355: 443-58). More recently, a method of rationally-designing mono-LAGLIDADG homing endonucleases was described which is capable of comprehensively redesigning I-CreI and

other homing endonucleases to target widely-divergent DNA sites, including sites in mammalian, yeast, plant, bacterial, and viral genomes (WO 2007/047859).

Thus, in one embodiment, the invention provides engineered meganucleases derived from the amino acid sequence of I-CreI that recognize and cut DNA sites in amplifiable regions of mammalian genomes. These engineered meganucleases can be used in accordance with the invention to target the insertion of gene expression cassettes into defined locations in the chromosomal DNA of cell lines such as CHO cells. This invention will greatly streamline the production of desired cell lines by reducing the number of lines that must be screened to identify a “high-producer” clone suitable for commercial-scale production of a therapeutic glycoprotein.

The present invention involves targeting transgenic DNA “sequences of interest” to amplifiable loci. The amplifiable loci are regions of the chromosomal DNA that contain selectable genes that become amplified in the presence of selection agents (*e.g.*, drugs). For example, the Chinese Hamster Ovary (CHO) cell DHFR locus can be amplified to ~1,000 copies by growing the cells in the presence of methotrexate (MTX), a DHFR inhibitor. Table 1 lists additional examples of selectable genes that can be amplified using small molecule drugs (Kellems, ed. *Gene amplification in mammalian cells: a comprehensive guide*. Marcel Dekker, New York, 1993; Omasa (2002), *J. Biosci. Bioeng.* 94:6 600-605).

Table 1. Amplifiable Genes

Selectable Gene Name	Amplified With
Dihydrofolate Reductase	Methotrexate (MTX)
Glutamine Synthetase	Methionine sulphoximine (MSX)
Hypoxanthine Phosphoribosyltransferase	Aminopterin, hypoxanthine, and thymidine
Threonyl tRNA Synthetase	Borrelidin
Na,K-ATPase	Ouabain
Asparagine Synthetase	Albizziin or Beta-aspartyl hydroxamate
Ornithine Decarboxylase	alpha-difluoromethylornithine (DFMO)
Inosine-5'-monophosphate dehydrogenase	Mycophenolic Acid
Adenosine Deaminase	Adenosine, Alanosine, 2'deoxycoformycin
Thymidylate Synthetase	Fluorouracil
Aspartate Transcarbamylase	N-Phosphonacetyl-L-Aspartate (PALA)
Metallothionein	Cadmium
Adenylate Deaminase (1,2)	Adenine, Azaserine, Coformycin

UMP-Synthetase	6-azauridine, pyrazofuran
Ribonucleotide Reductase	hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine, triapine.

Several considerations must be taken into account when selecting a specific target site for the insertion of a sequence of interest within an amplifiable locus. First, the selected insertion site must be co-amplified with the gene under selection. In many cases, experimental data already exists in the art which delimits the amount of flanking chromosomal sequence that co-amplifies with a selectable gene of interest. This data, which precisely defines the extent of the amplifiable locus, exists for CHO DHFR (Ma *et al.* (1988), *Mol Cell Biol.* 8(6):2316-27), human DHFR (Morales *et al.* (2009), *Mol Cancer Ther.* 8(2):424-432), and CHO GS (Sanders *et al.* (1987), *Dev Biol Stand.* 66:55-63). Where such data does not already exist in the art, we predict that chromosomal DNA sequences <100,000 base pairs upstream or downstream of the selectable gene coding sequence are likely to co-amplify. Hence, these regions could be suitable sites for targeting the insertion of a sequence of interest.

Second, target sites should be selected which will not greatly impact the function of the selectable gene (*e.g.*, the endogenous DHFR, GS, or HPRT gene). Because amplification requires a functional copy of the selectable gene, insertion sites within the promoter, exons, introns, polyadenylation signals, or other regulatory sequences that, if disrupted, would greatly impact transcription or translation of the selectable gene, should be avoided. For example, WO 2008/059317 discloses meganucleases which cleave DNA target sites within the HPRT gene. To the extent WO 2008/059317 discloses the insertion of genes into the HPRT locus, it teaches that the HPRT gene coding sequence should be disrupted in the process of transgene insertion to facilitate selection for proper targeting using 6-thioguanine. 6-thioguanine is a toxic nucleotide analog that kills cells having functional HPRT activity. Because cells produced in accordance with WO 2008/059317 will not have HPRT activity, they will not amplify an inserted transgene in response to treatment with an HPRT inhibitor and, so, cannot be used in the present invention. For the present invention, unless the precise limits of all regulatory sequences are already known for a particular selectable gene, insertion sites >1,000 base pairs, >2,000 base pairs, >3,000 base pairs, >4,000 base pairs, or, preferably, >5,000 base pairs, upstream or downstream of the gene

coding sequence should be selected. However, if the location of the regulatory sequences are known, the sequence of interest can be inserted immediately adjacent to either the most 5' or 3' regulatory sequence (*e.g.*, immediately 3' to the polyadenylation signal).

5 Lastly, target sites should be selected which do not disrupt other chromosomal genes which may be important for normal cell physiology. In general, gene insertion sites should be >1,000 base pairs, >2,000 base pairs, >3,000 base pairs, >4,000 base pairs, or, preferably, >5,000 base pairs, away from any gene coding sequence.

Various methods of the invention are described schematically in the figures as follows:

10 Figure 1 depicts a general strategy for targeting a sequence of interest to an amplifiable locus. In the first step, a site-specific endonuclease introduces a double-stranded break in the chromosomal DNA of a cell at a site that is proximal to an endogenous selectable gene. The cleaved chromosomal DNA then undergoes homologous recombination with a donor DNA molecule comprising a sequence of interest flanked by DNA sequences
15 homologous to sequences flanking the endonuclease recognition sequence in the target site. As a result, the sequence of interest is inserted into the chromosomal DNA of the cell adjacent to the endogenous selectable gene. The modified cell is then grown in the presence of one or more compounds that inhibit the function of the selectable gene to induce an increase in the copy number (*i.e.*, amplification) of the selectable gene. The sequence of
20 interest, which is genetically linked to the selectable gene, will co-amplify with the selectable gene. The result is a stable transgenic cell line comprising multiple copies of the sequence of interest.

Figure 2(A) depicts a schematic of the CHO DHFR locus showing a preferred region for targeting a sequence of interest 5,000-60,000 base pairs downstream of the DHFR
25 gene. Figure 2(B) depicts a schematic of the CHO GS locus showing a preferred region for targeting a sequence of interest 5,000-55,000 base pairs downstream of the GS gene. Promoters are shown as arrows. Exons are shown as rectangles, with non-coding exons in white and protein coding exons in gray.

30 Figure 3 depicts a strategy for inserting a sequence of interest into an amplifiable locus in a two-step process involving a pre-integrated target sequence. In the first step, the chromosomal DNA of a cell is cleaved by a site-specific endonuclease at a site that is proximal to a selectable gene. The cleaved chromosomal DNA then undergoes homologous recombination with a donor DNA molecule comprising an exogenous target sequence flanked by DNA sequences homologous to the sequences flanking the endogenous target site. This

results in the insertion of the new engineered target sequence into the chromosomal DNA of the cell proximal to the selectable gene. A sequence of interest can subsequently be targeted proximal to the same selectable gene using a nuclease, integrase, transposase, or recombinase that specifically recognizes the pre-integrated engineered target sequence. The modified cell is then grown in the presence of one or more compounds that co-amplify the selectable gene and the sequence of interest.

Figure 4 depicts a strategy for inserting an engineered target sequence into a selectable gene (*e.g.*, DHFR) with concomitant removal of a portion of the selectable gene. A site-specific endonuclease is first used to cleave the chromosomal DNA of the cell proximal to or within the selectable gene sequence. As shown in the figure, the endogenous target site is between exons 2 and 3 of the CHO DHFR gene (although the target site could be within any intron or exon, and the selectable gene could be any gene subject to amplification). The chromosomal DNA then undergoes homologous recombination with a first donor DNA (“donor DNA #1”) such that the sequence of the first donor DNA is inserted into the chromosomal DNA of the cell. As shown in the figure, this results in the replacement of the promoter and first two exons of DHFR by the new engineered target sequence (although the first donor DNA could replace more or less of the chromosomal DNA, such as only a portion of one exon). If such a replacement is made to all DHFR alleles in a cell, the resultant cell line is DHFR (-/-). A sequence of interest can subsequently be targeted proximal to the selectable gene in the cell line using an endonuclease, integrase, transposase, or recombinase that recognizes the engineered target sequence. As shown in the figure, the second donor DNA (“donor DNA #2”) comprises a sequence of interest as well as a promoter and the first two exons of DHFR. Proper targeting of this second donor DNA molecule results in the insertion of the sequence of interest at the engineered target sequence while simultaneously reconstituting a functional DHFR gene. Thus, properly targeted cell lines will be DHFR+ and can be selected using media deficient in hypoxanthine/thymidine. In addition, the sequence of interest can be co-amplified with the DHFR gene using MTX selection. The strategy diagrammed here for DHFR can be applied to any selectable gene in an amplifiable locus.

Figure 5 depicts a strategy for inserting an engineered target sequence into an amplifiable locus with concomitant disruption of the coding sequence of a selectable gene. A site-specific endonuclease is first used to cleave the chromosomal DNA of the cell within the selectable gene coding sequence. As shown in the figure, the endogenous target site is in the third exon of the CHO GS gene. The chromosomal DNA then undergoes homologous

recombination with a first donor DNA (“donor DNA #1”) such that the sequence of the first donor DNA is inserted into the chromosomal DNA of the cell. This results in the insertion of a new engineered target sequence into the GS coding sequence. If such an insertion occurs in both alleles of the GS gene and results in a frameshift mutation or otherwise disrupts the function of the GS gene, the resultant cell line will be GS (-/-). A sequence of interest can subsequently be targeted proximal to the amplifiable locus in the cell line using an endonuclease, integrase, transposase, or recombinase that recognizes the engineered target sequence. As shown in the figure, a second donor DNA (“donor DNA #2”) comprises a sequence of interest operably linked to a promoter as well as the 3' portion of the GS coding sequence comprising exons 3, 4, 5, and 6. (The figure shows exons 3, 4, 5, and 6 joined into a single nucleotide sequence (*i.e.*, with introns removed), but a sequence including either the naturally-occurring introns or one or more artificial introns could also be employed). Proper targeting of the second donor DNA molecule results in the insertion of the sequence of interest at the engineered target sequence while simultaneously reconstituting a functional GS gene. Thus, properly targeted cell lines will be GS+ and can be selected using media deficient in L-glutamine. In addition, the sequence of interest can be co-amplified with the GS gene using MSX selection. The strategy diagrammed here for GS can be applied to any selectable gene in an amplifiable locus.

Figure 6 depicts a strategy for inserting an engineered target sequence into an amplifiable locus with concomitant disruption of the mRNA processing of a selectable gene. A site-specific endonuclease is first used to cleave the chromosomal DNA of the cell within an intron in the selectable gene. As drawn, the endogenous target site is in the intron between the third and fourth coding exons of the CHO GS gene. The chromosomal DNA then undergoes homologous recombination with a donor DNA #1 such that the sequence of the donor DNA is inserted in the chromosomal DNA of the cell. This results in the insertion of a new engineered target sequence into the GS coding sequence with an additional sequence that causes the GS mRNA to be processed incorrectly. As drawn, this additional sequence comprises a strong splice acceptor. If such an insertion occurs in both alleles of the GS gene, the artificial splice acceptor will cause the GS mRNA to splice incorrectly, resulting in a loss of GS expression and a requirement for growth in media containing L-glutamine. A sequence of interest can subsequently be targeted to the amplifiable locus in the cell line using an endonuclease, integrase, transposase, or recombinase that recognizes the engineered target sequence. As diagrammed, donor DNA #2 comprises a sequence of interest operably linked to a promoter as well as the 3' portion of the GS coding sequence comprising exons 4,

5, and 6 joined into a single nucleotide sequence. (The figure shows exons 4, 5, and 6 joined into a single nucleotide sequence (*i.e.*, with introns removed), but a sequence including either the naturally-occurring introns or one or more artificial introns could also be employed).

Proper targeting of this donor DNA #2 molecule results in the insertion of the sequence of interest at the engineered target sequence while simultaneously reconstituting a functional GS gene. Thus, properly targeted cell lines will be GS⁺ and can be selected using media deficient in L-glutamine and the sequence of interest can be co-amplified with the GS gene using MSX selection. The strategy diagrammed here for GS can be applied to any selectable gene in an amplifiable locus.

10 Figure 7(A) depicts a direct-repeat recombination assay for site-specific endonuclease activity. A reporter plasmid is produced comprising the 5' two-thirds of the GFP gene ("GF"), followed by an endonuclease recognition sequence, followed by the 3' two-thirds of the GFP gene ("FP"). Mammalian cells are transfected with this reporter plasmid as well as a gene encoding an endonuclease. Cleavage of the recognition sequence
15 by the endonuclease stimulates homologous recombination between direct repeats of the GFP gene to restore GFP function. GFP⁺ cells can then be counted and/or sorted on a flow cytometer.

Figure 7(B) depicts the results of the assay of Figure 7(A) as applied to the CHO-23/24 and CHO-51/52 meganucleases. Light bars indicate the percentage of GFP⁺ cells when cells are transfected with the reporter plasmid alone (-endonuclease). Dark bars indicate the percentage of GFP⁺ cells when cells are co-transfected with a reporter plasmid and the corresponding meganuclease gene (+endonuclease). The assay was performed in triplicate and the standard deviation is shown.

Figure 7(C) depicts alignment of sequences obtained from CHO cells transfected with mRNA encoding the CHO-23/24 meganuclease. The top sequence is from a wild-type (WT) CHO cell with the recognition sequence for CHO-23/24 underlined.

Figure 7(D) depicts alignment of sequences obtained from CHO cells transfected with mRNA encoding the CHO-51/52 meganuclease. The top sequence is from a wild-type (WT) CHO cell with the recognition sequence for CHO-51/52 underlined.

30 Figure 8(A) depicts a strategy for inserting an exogenous DNA sequence into the CHO DHFR locus using the CHO-51/52 meganuclease. CHO cells were co-transfected with mRNA encoding CHO-51/52 and a donor plasmid comprising an EcoRI site flanked by 543 base pairs of DNA sequence homologous to the region upstream of the CHO-51/52 recognition site and 461 base pairs of DNA sequence homologous to the region downstream

of the CHO-51/52 recognition site. 48 hours post-transfection, genomic DNA was isolated and subjected to PCR using primers specific for the downstream region of the DHFR locus (dashed arrows).

Figure 8(B) depicts PCR products that were cloned into pUC-19 and 48 individual plasmid clones and were digested with EcoRI and visualized on an agarose gel. 10 plasmids (numbered lanes) yielded a 647 base pair restriction fragment, consistent with cleavage of a first EcoRI site within the pUC-19 vector and a second EcoRI site in the cloned PCR fragment. These 10 plasmids were sequenced to confirm that they harbor a PCR fragment comprising a portion of the downstream DHFR locus with an EcoRI restriction site inserted into the CHO-51/52 recognition sequence. This restriction pattern was not observed when CHO cells were transfected with the donor plasmid alone.

Figure 9(A) depicts a strategy for inserting an engineered target sequence into the CHO DHFR locus using the CHO-23/24 meganuclease. CHO cells were co-transfected with mRNA encoding CHO-23/24 and a donor plasmid comprising, in 5' to 3' orientation, an SV40 promoter, an ATG start codon, an FRT site, and a Zeocin-resistance (Zeo) gene. Zeocin-resistant cells were cloned by limiting dilution and screened by PCR to identify a clonal cell line in which the donor plasmid sequence integrated into the CHO-23/24 recognition site. After expansion, this cell line was co-transfected with a first plasmid encoding Flp recombinase operably linked to a promoter and second plasmid (donor plasmid #2) comprising a GFP gene under the control of a CMV promoter, an FRT site, and a hygromycin-resistance (Hyg) gene lacking a start codon. Flp-mediated recombination between FRT sites resulted in the integration of the donor plasmid #2 sequence into the engineered target sequence (*i.e.*, the FRT site) such that a functional Hyg gene expression cassette was produced. Figure 9(B) depicts PCR products from hygromycin-resistant clones produced as in (A) that were cloned by limiting dilution. Genomic DNA was extracted from 24 individual clones and PCR amplified using a first primer in the DHFR locus and a second primer in the Hyg gene (dashed lines). All 24 clones yielded a PCR product consistent with Hyg gene insertion into the engineered target sequence. Figure 9(C) depicts GFP expression by the 24 clones produced in (B) using flow cytometry. All clones were found to express high levels of GFP with relatively little clone-to-clone variability.

Figure 10. A GFP-expressing CHO line was produced by integrating a GFP gene expression cassette into the DHFR locus using an engineered target sequence strategy as shown in Figure 9. This cell line was then grown in MTX as described in Example 2 to amplify the integrated GFP gene. (A) Flow cytometry plots showing GFP intensity on the Y-

axis. In the pre-MTX cell line, GFP intensity averages approximately 2×10^3 whereas in the cell line grown in 250 nM MTX, a distinct sub-population is visible (circled) in which GFP intensity approaches 10^4 . (B) MTX treated cell lines were sorted by FACS to identify individual cells expressing higher amounts of GFP. Five such high-expression cells were expanded and GFP intensity was determined by flow cytometry. All five clones were found to have significantly increased GFP expression relative to the pre-MTX cell line. (C) Genomic DNA was isolated from the five clonal cell lines produced in (B) and subjected to quantitative PCR using a primer pair specific for the GFP gene. It was found that the five high-expression clones had significantly more copies of the GFP gene than the pre-MTX cell line. These results demonstrate that the copy number and expression level a transgene integrated downstream of CHO DHFR can amplify in response to MTX treatment.

Figure 11. (A) A direct-repeat recombination assay, as in Figure 5A. (B) The assay in (A) applied to the CHO-13/14 and CGS-5/6 meganucleases. Light bars indicate the percentage of GFP+ cells when cells are transfected with the reporter plasmid alone (- endonuclease). Dark bars indicate the percentage of GFP+ cells when cells are co-transfected with a reporter plasmid and the corresponding meganuclease gene (+endonuclease). The assay was performed in triplicate and standard deviation is shown. (C) Alignment of sequences obtained from CHO cells transfected with mRNA encoding the CGS-5/6 meganuclease. The top sequence is from a wild-type (WT) CHO cell with the recognition sequence for CGS-5/6 underlined. Dashes indicate deleted bases. Bases that are italicized and in bold are point mutations or insertions relative to the wild-type sequence. Note that the mutations observed in at least clones 6d4, 6g5, 3b7, 3d11, 3e5, 6f10, 6hH8, 6d10, 6d7, 3g8, and 3a9 are expected to knockout GS gene function.

2.1.1 Gene Targeting to the CHO DHFR Locus

The CHO DHFR locus is diagrammed in Figure 2A. The locus comprises the DHFR gene coding sequence in 6 exons spanning ~24,500 base pairs. The Msh3 gene is located immediately upstream of DHFR and is transcribed divergently from the same promoter as DHFR. A hypothetical gene, 2BE2121, can be found ~65,000 base pairs downstream of the DHFR coding sequence. Thus, there is a ~65,000 base pair region downstream of the DHFR gene that does not harbor any known genes and is a suitable location for targeting the insertion of sequences of interest. Target sites for insertion of a sequence of interest generally should not be selected which are <1,000 base pairs, and preferably not <5,000 base pairs from either the DHFR or 2BE2121 genes. This limits the

window of preferred target sites to the region 1,000-60,000 base pairs, or 5,000-60,000 base pairs downstream of the DHFR coding sequence. The sequence of this region is provided as SEQ ID NO: 2.

The human and mouse DHFR loci have an organization similar to CHO locus. In both cases, the Msh3 gene is immediately upstream of DHFR but there is a large area devoid of coding sequences downstream of DHFR. In humans, the ANKRD34B gene is ~55,000 base pairs downstream of DHFR while the ANKRD34B gene is ~37,000 base pairs downstream of DHFR in mouse. Therefore, the genomic region downstream of DHFR is an appropriate location to insert genes of interest in CHO, human, and mouse cells and cell lines. Further, gene expression cassettes inserted into this region will be expressed at a high level, resistant to gene silencing, and capable of being amplified by treatment with MTX. Methods for amplifying the CHO cell DHFR locus are known in the art (see, e.g., Kellems, ed., *Gene amplification in mammalian cells: a comprehensive guide*. Marcel Dekker, New York, 1993) and typically involve gradually increasing the concentration of MTX in the growth media from 0 to as high as 0.8 mM over a period of several weeks.

2.1.2 Gene Targeting to the GS Locus

The CHO, human, and mouse glutamine synthetase (also known as “glutamate-ammonia ligase” or “GluL”) loci share a common organization (Fig. 2B). The TEDDM1 gene is immediately upstream of GS in all species (~5,000 bp upstream in the case of human, ~7,000 bp upstream in the case of mouse and CHO). The closest downstream gene, however, is ~46,000 away in the case of human and ~117,000 bp away in the case of mouse and CHO. Therefore, we predict that the chromosomal region 1,000-41,000 bp, or 5,000-41,000 bp downstream of GS in human cells and 1,000-100,000 bp, or 5,000-100,000 bp downstream of GS in mouse and CHO cells are appropriate locations to target the insertion of sequences of interest. Because DNA sites distal to the GS coding sequence are more likely to be susceptible to gene silencing, the chromosomal region 5,000-60,000 bp downstream of GS is a preferred location to target the insertion of a sequence of interest even in mouse or CHO cells. The sequence of this region from the CHO genome is provided as SEQ ID NO: 3. Gene expression cassettes inserted into this region will be expressed at a high level, resistant to gene silencing, and capable of being amplified by treatment with MSX. Less-preferred regions include the chromosomal region between the TEDDM1 and GS genes or the region <10,000 bp downstream of TEDDM1 (see Figure 2B). Methods for amplifying the GS locus are known in the art (Bebbington *et al.* (1992), *Biotechnology (N Y)*. 10(2):169-75).

2.2 Engineered Endonucleases for Gene Targeting

A sequence of interest may be inserted into an amplifiable locus using an engineered site-specific endonuclease. Methods for generating site-specific endonucleases which can target DNA breaks to pre-determined loci in a genome are known in the art. These include zinc-finger nucleases (Le Provost *et al.* (2010), *Trends Biotechnol.* 28(3):134-41), TAL-effector nucleases (Li *et al.* (2011), *Nucleic Acids Res.* 39(1):359-72), and engineered meganucleases (WO 2007/047859; WO 2007/049156; WO 2009/059195). In one embodiment, the invention provides engineered meganucleases derived from I-CreI that can be used to target the insertion of a gene of interest to an amplifiable locus. Methods to produce such engineered meganucleases are known in the art (see, *e.g.*, WO 2007/047859; WO 2007/049156; WO 2009/059195). In preferred embodiments, a “single-chain” meganuclease is used to target gene insertion to an amplifiable region of the genome. Methods for producing such “single-chain” meganucleases are known in the art (see, *e.g.*, WO 2009/059195 and WO 2009/095742). In some embodiments, the engineered nuclease is fused to a nuclear localization signal (NLS) to facilitate nuclear uptake. Examples of nuclear localization signals include the SV40 NLS (amino acid sequence MAPKKKRKV) which can be fused to the C- or, preferably, the N-terminus of the protein. In addition, an engineered nuclease may be tagged with a peptide epitope (*e.g.*, an HA, FLAG, or Myc epitope) to monitor expression levels or localization or to facilitate purification.

2.3 Engineered Cell Lines with Sequences of Interest Targeted to Amplifiable Loci

In some embodiments, the invention provides methods for using engineered nucleases to target the insertion of transgenes into amplifiable loci in cultured mammalian cells. This method has two primary components: (1) an engineered nuclease; and (2) a donor DNA molecule comprising a sequence of interest. The method comprises contacting the DNA of the cell with the engineered nuclease to create a double strand DNA break in an endogenous recognition sequence in an amplifiable locus followed by the insertion of the donor DNA molecule at the site of the DNA break. Such insertion of the donor DNA is facilitated by the cellular DNA-repair machinery and can occur by either the non-homologous end-joining pathway or by homologous recombination (Figure 1).

The engineered nuclease can be delivered to the cell in the form protein or, preferably, as a nucleic acid encoding the engineered nuclease. Such nucleic acid can be DNA (*e.g.*, circular or linearized plasmid DNA or PCR products) or RNA. For embodiments

in which the engineered nuclease coding sequence is delivered in DNA form, it should be operably linked to a promoter to facilitate transcription of the engineered nuclease gene. Mammalian promoters suitable for the invention include constitutive promoters such as the cytomegalovirus early (CMV) promoter (Thomsen *et al.* (1984), *Proc Natl Acad Sci U S A.* 81(3):659-63) or the SV40 early promoter (Benoist and Chambon (1981), *Nature.* 290(5804):304-10) as well as inducible promoters such as the tetracycline-inducible promoter (Dingermann *et al.* (1992), *Mol Cell Biol.* 12(9):4038-45).

5 In some embodiments, mRNA encoding the engineered nuclease is delivered to the cell because this reduces the likelihood that the gene encoding the engineered nuclease will integrate into the genome of the cell. Such mRNA encoding an engineered nuclease can be produced using methods known in the art such as *in vitro* transcription. In some 10 embodiments, the mRNA is capped using 7-methyl-guanosine. In some embodiments, the mRNA may be polyadenylated.

Purified engineered nuclease proteins can be delivered into cells to cleave 15 genomic DNA, which allows for homologous recombination or non-homologous end-joining at the cleavage site with a sequence of interest, by a variety of different mechanisms known in the art. For example, the recombinant nuclease protein can be introduced into a cell by techniques including, but not limited to, microinjection or liposome transfections (see, *e.g.*, Lipofectamine™, Invitrogen Corp., Carlsbad, CA). The liposome formulation can be used to 20 facilitate lipid bilayer fusion with a target cell, thereby allowing the contents of the liposome or proteins associated with its surface to be brought into the cell. Alternatively, the enzyme can be fused to an appropriate uptake peptide such as that from the HIV TAT protein to direct cellular uptake (see, *e.g.*, Hudecz *et al.* (2005), *Med. Res. Rev.* 25: 679-736).

Alternatively, gene sequences encoding the engineered nuclease protein are 25 inserted into a vector and transfected into a eukaryotic cell using techniques known in the art (see, *e.g.*, Ausubel *et al.*, Current Protocols in Molecular Biology, Wiley 1999). The sequence of interest can be introduced in the same vector, a different vector, or by other means known in the art. Non-limiting examples of vectors for DNA transfection include virus vectors, plasmids, cosmids, and YAC vectors. Transfection of DNA sequences can be 30 accomplished by a variety of methods known to those of skill in the art. For instance, liposomes and immunoliposomes are used to deliver DNA sequences to cells (see, *e.g.*, Lasic *et al.* (1995), *Science* 267: 1275-76). In addition, viruses can be utilized to introduce vectors into cells (see, *e.g.*, U.S. Pat. No. 7,037,492). Alternatively, transfection strategies can be utilized such that the vectors are introduced as naked DNA (see, *e.g.*, Rui *et al.* (2002), *Life*

Sci. 71(15): 1771-8).

General methods for delivering nucleic acids into cells include: (1) chemical methods (Graham *et al.* (1973), *Virology* 54(2):536-539; Zatloukal *et al.* (1992), *Ann. N.Y. Acad. Sci.*, 660:136-153; (2) physical methods such as microinjection (Capecchi (1980), *Cell* 5 22(2):479-488, electroporation (Wong *et al.* (1982), *Biochim. Biophys. Res. Commun.* 107(2):584-587; Fromm *et al.* (1985), *Proc. Nat'l Acad. Sci. USA* 82(17):5824-5828; U.S. Pat. No. 5,384,253) and ballistic injection (Johnston *et al.* (1994), *Methods Cell. Biol.* 43(A): 353-365; Fynan *et al.* (1993), *Proc. Nat'l Acad. Sci. USA* 90(24): 11478-11482); (3) viral vectors (Clapp (1993), *Clin. Perinatol.* 20(1): 155-168; Lu *et al.* (1993), *J. Exp. Med.* 10 178(6):2089-2096; Eglitis *et al.* (1988), *Avd. Exp. Med. Biol.* 241:19-27; Eglitis *et al.* (1988), *Biotechniques* 6(7):608-614); and (4) receptor-mediated mechanisms (Curiel *et al.* (1991), *Proc. Nat'l Acad. Sci. USA* 88(19):8850-8854; Curiel *et al.* (1992), *Hum. Gen. Ther.* 3(2):147-154; Wagner *et al.* (1992), *Proc. Nat'l Acad. Sci. USA* 89 (13):6099-6103). In some preferred embodiments, 7-methyl-guanosine capped mRNA encoding the engineered 15 nuclease is delivered to cells using electroporation.

The donor DNA molecule comprises a gene of interest operably linked to a promoter. In many cases, a donor molecule may comprise multiple genes operably linked to the same or different promoters. For example, donor molecules comprising monoclonal antibody expression cassettes may comprise a gene encoding the antibody heavy chain and a 20 second gene encoding the antibody light chain. Both genes may be under the control of different promoters or they may be under the control of the same promoter by using, for example, an internal-ribosome entry site (IRES). Donor molecules may also comprise a selectable marker gene operably linked to a promoter to facilitate the identification of transgenic cells. Such selectable markers are known in the art and include neomycin 25 phosphotransferase (NEO), hypoxanthine phosphoribosyltransferase (HPRT), glutamine synthetase (GS), dihydrofolate reductase (DHFR), and hygromycin phosphotransferase (HYG) genes.

In some embodiments, donor DNA molecules will additionally comprise flanking sequences homologous to the target sequences in the DNA of the cell. Such homologous 30 flanking sequences comprise >3 or, preferably, >50 or, more preferably, >200 or, most preferably, >400 base pairs of DNA that are identical or nearly identical in sequence to the chromosomal locus recognized by the engineered nuclease (Figure 1). Such homologous DNA sequences facilitate the integration of the donor DNA sequence into the amplifiable locus by homologous recombination.

The “donor” DNA molecule can be circular (*e.g.*, plasmid DNA) or linear (*e.g.*, linearized plasmid or PCR products). Methods for delivering DNA molecules are known in the art, as discussed above.

In some embodiments, the engineered nuclease gene and donor DNA are carried
5 on separate nucleic acid molecules which are co-transfected into cells or cell lines. For example, the engineered nuclease gene operably linked to a promoter can be transfected in plasmid form simultaneously with a separate donor DNA molecule in plasmid or PCR product form. In an alternative embodiment, the engineered nuclease can be delivered in mRNA form with a separate donor DNA molecule in plasmid or PCR product form. In a
10 third embodiment, the engineered nuclease gene and donor DNA are carried on the same DNA molecule, such as a plasmid. In a fourth embodiment, cells are co-transfected with purified engineered nuclease protein and a donor DNA molecule in plasmid or PCR product form.

Following transfection with the engineered nuclease and donor DNA, cells are
15 typically allowed to recover from transfection (24-72 hours) before being cloned using methods known in the art. Common methods for cloning a genetically engineered cell line include “limiting dilution” in which transfected cells are transferred to tissue culture plates (*e.g.*, 48 well, 96 well plates) at a concentration of <1 cell per well and expanded into clonal populations. Other cloning strategies include robotic clone identification/isolation systems
20 such as ClonePix™ (Genetix, Molecular Devices, Inc., Sunnyvale, CA). Clonal cell lines can then be screened to identify cell lines in which the sequence of interest is integrated into the intended target site. Cell lines can easily be screened using molecular analyses known in the art such as PCR or Southern Blot. For example, genomic DNA can be isolated from a clonal cell line and subjected to PCR amplification using a first (sense-strand) primer that
25 anneals to a DNA sequence in the sequence of interest and a second (anti-sense strand) primer that anneals to a sequence in the amplifiable locus. If the donor DNA molecule comprises a DNA sequence homologous to the target site, it is important that the second primer is designed to anneal to a sequence in the amplifiable locus that is beyond the limits of homology carried on the donor molecule to avoid false positive results. Alternatively, cell
30 lines can be screened for expression of the sequence of interest. For example, if the sequence of interest encodes a secreted protein such as an antibody, the growth media can be sampled from isolated clonal cell lines and assayed for the presence of antibody protein using methods known in the art such as Western Blot or Enzyme-Linked Immunosorbant Assay (ELISA). This type of functional screen can be used to identify clonal cell lines which carry at least one

copy of the sequence of interest integrated into the genome. Additional molecular analyses such as PCR or Southern blot can then be used to determine which of these transgenic cell lines carry the sequence of interest targeted to the amplifiable locus of interest, as described above.

5 The method of the invention can be used on any culturable and transfectable cell type such as immortalized cell lines and stem cells. In preferred embodiments, the method of the invention is used to genetically modify immortalized cell lines that are commonly used for biomanufacturing. This includes:

- 10 1. Hamster cell lines such as baby hamster kidney (BHK) cells and all variants of Chinese Hamster Ovary (CHO) cells, *e.g.*, CHO-K1, CHO-S (Invitrogen Corp., Carlsbad, CA), DG44, or PotelligentTM (Lonza Group Ltd., Basel, Switzerland). Because the genome sequences of different hamster cell lines are very nearly identical, an engineered meganuclease which can be used to practice the invention in one hamster cell type (*e.g.*, BHK cells) can generally be used to practice the
15 invention in another hamster cell type (*e.g.*, CHO-K1).
2. Mouse cell lines such as mouse hybridoma or mouse myeloma (*e.g.*, NS0) cells. Because the genome sequences of different mouse cell lines are very nearly identical, an engineered meganuclease which can be used to practice the invention in one mouse cell type (*e.g.*, mouse hybridoma cells) can generally be used to practice the
20 invention in another mouse cell type (*e.g.*, NS0).
3. Human cell lines such as human embryonic kidney cells (*e.g.*, HEK-293 or 293S) and human retinal cells (*e.g.*, PER.C6). Because the genome sequences of different human cell lines are very nearly identical, an engineered meganuclease which can be used to practice the invention in one human cell type (*e.g.*, HEK-293 cells) can
25 generally be used to practice the invention in another human cell type (*e.g.*, PER.C6).

2.6 Pre-Engineered Cell Lines with Engineered Target Sequences in Amplifiable Loci.

 In one embodiment, the invention provides cell lines which are pre-engineered to comprise a targetable “engineered target sequence” for gene insertion in an amplifiable locus
30 in a mammalian cell line (Figure 3). An engineered target sequence comprises a recognition sequence for an enzyme which is useful for inserting transgenic nucleic acids into chromosomal DNA sequences. Such engineered target sequences can include recognition sequences for engineered meganucleases derived from I-CreI (*e.g.*, SEQ ID NO 37-87 from WO 2009/076292), recognition sequences for zinc-finger nucleases, recognition sequences

for TAL effector nucleases (TALENs), the LoxP site (SEQ ID NO 4) which is recognized by Cre recombinase, the FRT site (SEQ ID NO: 5) which is recognized by FLP recombinase, the attB site (SEQ ID NO: 6) which is recognized by lambda recombinase, or any other DNA sequence known in the art that is recognized by a site specific endonuclease, recombinase, integrase, or transposase that is useful for targeting the insertion of nucleic acids into a genome. Thus, the invention allows one skilled in the art to use an engineered nuclease (*e.g.*, a meganuclease, zinc-finger nuclease, or TAL effector nuclease) to insert an engineered target sequence into an amplifiable locus in a mammalian cell line. The resulting cell line comprising such an engineered target sequence at an amplifiable locus can then be contacted with the appropriate enzyme (*e.g.*, a second engineered meganuclease, a second zinc-finger nuclease, a second TAL effector nuclease, a recombinase, an integrase, or a transposase) to target the insertion of a gene of interest into the amplifiable locus at the engineered target sequence. This two-step approach can be advantageous because the efficiency of gene insertion that can be achieved using an optimal meganuclease, zinc-finger nuclease, recombinase, integrase, or transposase might be higher than what can be achieved using the initial endonuclease (*e.g.*, meganuclease or zinc-finger nuclease) that cleaves the endogenous target site to promote insertion of the engineered target sequence.

In an alternative embodiment, a cell line is produced by inserting an engineered target sequence into an amplifiable locus with the concomitant removal of all or a portion of the adjacent endogenous marker gene (Figure 4). For example, an engineered meganuclease, zinc-finger nuclease, or TAL-effector nuclease can be used to remove the first two exons of both alleles of the CHO DHFR gene and replace them with an engineered target sequence for a different engineered meganuclease, ZFN, TALEN, recombinase, integrase, or transposase. The resulting cell line will be DHFR deficient and unable to grow in the absence of hypoxanthine/thymidine. Alternatively, for example, an engineered meganuclease, ZFN or TALEN can be used to remove the first exon of both alleles of the CHO GS gene and replace it with an engineered target sequence for a different engineered meganuclease, ZFN, TALEN, recombinase, integrase, or transposase (Figure 4). The resulting cell line will be GS deficient and unable to grow in the absence of L-glutamine. Such a cell line is useful because a gene of interest can be inserted into the engineered target sequence in the pre-engineered cell line while simultaneously reconstituting the selectable gene (*e.g.*, DHFR or GS). Thus, it is possible to select for transfectants harboring the gene of interest at the amplifiable locus using media conditions that select for DHFR⁺ or GS⁺ cells.

In an alternative embodiment, a cell line is produced in which an engineered target

sequence is inserted into an amplifiable locus with disruption of the selectable gene (Figures 5, 6). This can be accomplished, for example, using a meganuclease which recognizes a DNA site in the coding sequence of the selectable gene. Such a meganuclease can be used to target the insertion of an engineered target sequence into the selectable gene coding sequence resulting in disruption of gene function by, for example, introducing a frameshift (Figure 5). Alternatively, for example, an engineered target sequence can be inserted into an intron in the selectable gene sequence with an additional sequence that promotes improper processing of the selectable gene transcript (Figure 6). Such sequences that promote improper processing include, for example, artificial splice acceptors or polyadenylation signals. Splice acceptor sequences are known in the art (Clancy (2008), "RNA Splicing: Introns, Exons and Spliceosome," *Nature Education* 1:1) and typically comprise a 20-50 base pair pyrimidine-rich sequence followed by a sequence (C/T)AG(A/G). SEQ ID NO: 33 is an example of a splice acceptor sequence. Likewise, polyadenylation signals are known in the art and include, for example, the SV40 polyadenylation signal (SEQ ID NO: 34) and the BGH polyadenylation signal (SEQ ID NO: 35). In some embodiments, the resulting cell line harboring the new engineered target sequence in all alleles of the selectable gene will be deficient in the function of the gene due to mis-transcription or mis-translation and will be able to grow only under permissive conditions. For example, an engineered target sequence can be inserted into the GS gene sequence using a meganuclease resulting in a cell line that is GS^{-/-} that can grow only in the presence of L-glutamine in the growth media. In a subsequent step, a gene of interest can be inserted into the engineered target sequence while simultaneously reconstituting the selectable gene (*e.g.*, DHFR or GS). Thus, it is possible to select for transfectants harboring the gene of interest at the amplifiable locus using media conditions that select for DHFR⁺ or GS⁺ cells.

25

2.5 Transgenic Cell Lines for Biomanufacturing.

In some embodiments, the invention provides transgenic cell lines suitable for the production of protein pharmaceuticals. Such transgenic cell lines comprise a population of cells in which a gene of interest, operably linked to a promoter, is inserted into the genome of the cell at an amplifiable locus wherein the gene of interest encodes a protein therapeutic. Examples of protein therapeutics include: monoclonal antibodies, antibody fragments, erythropoietin, tissue-type plasminogen activator, Factor VIII, Factor IX, insulin, colony stimulating factors, interferons (*e.g.*, interferon- α , interferon- β , and interferon- γ), interleukins (*e.g.*, interleukin-2), vaccines, tumor necrosis factor, and glucocerebrosidase. Protein

30

therapeutics are also referred to as “biologics” or “biopharmaceuticals.”

To be used for biomanufacturing, a transgenic cell line of the invention should undergo: (1) adaptation to serum-free growth in suspension; and (2) amplification of the gene of interest. In some embodiments, the invention is practiced on adherent cell lines which can be adapted to growth in suspension to facilitate their maintenance in shaker-flasks or stirred-tank bioreactors as is typical of industrial biomanufacturing. Methods for adapting adherent cells to growth in suspension are known in the art (Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). For regulatory reasons, it is generally necessary to further adapt biomanufacturing cell lines to chemically-defined media lacking animal-derived components (*i.e.*, “serum-free” media). Methods for preparing such media and adapting cell lines to it are known in the art (Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). Such media can also be purchased commercially (*e.g.*, CD-3 media for maintenance of CHO cells, available from Sigma-Aldrich, St. Louis, MO) and cells can be adapted to it by following the manufacturers’ instructions. In some embodiments, the cell line is adapted to growth in suspension and/or serum-free media prior to being transfected with the engineered nuclease.

Lastly, methods for gene amplification are known in the art (Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). In general, the process involves adding an inhibitor of a selectable gene product to the growth media to select for cells that express abnormally high amounts of the gene product due to gene-duplication events. In general, the concentration of inhibitor added to the growth media is increased slowly over a period of weeks until the desired level of gene amplification is achieved. Inhibitor is then generally removed from the media prior to initiating a bioproduction run to avoid the possibility of the inhibitor contaminating the protein therapeutic formulation. For example, the CHO DHFR locus can be amplified by slowly increasing the concentration of MTX in the growth media from 0 mM to as high as 0.8 mM over a period of several weeks. The GS locus can, likewise, be amplified by slowly increasing the concentration of MSX in the media from 0 μ M to as high as 100 μ M over a period of several weeks. Methods for evaluating gene amplification are known in the art and include Southern Blot and quantitative real-time PCR (rtPCR). In addition, or as an alternative, expression levels of the sequence of interest, which are generally correlated to gene copy number, can be evaluated by determining the concentration of protein therapeutic in the growth media using conventional methods such as Western Blot or ELISA.

Following cell line production, adaptation, and amplification, protein therapeutics can

be produced and purified using methods that are standard in the biopharmaceutical industry.

EXAMPLES

This invention is further illustrated by the following examples, which should not be construed as limiting. Those skilled in the art will recognize, or be able to ascertain, using no more than routine experimentation, numerous equivalents to the specific substances and procedures described herein. Such equivalents are intended to be encompassed in the scope of the claims that follow the examples below. Example 1 refers to engineered meganucleases that can be used to target the insertion of a gene of interest downstream of the DHFR gene in CHO cells. Example 2 refers to engineered meganucleases that can be used to target the insertion of an engineered target sequence into the CHO DHFR gene with concomitant removal of DHFR exons 1 and 2. Example 2 also refers to engineered meganucleases that can be used to target the insertion of an engineered target sequence into the CHO GS gene. Example 3 refers to meganucleases that can be used to target the insertion of a gene of interest downstream of the GS gene in CHO cells.

EXAMPLE 1

Targeted Gene Insertion into the CHO DHFR Locus Using Engineered Meganucleases

The CHO genomic DNA sequence 10,000-55,000 base pairs downstream of the DHFR gene was searched to identify DNA sites amenable to targeting with engineered meganucleases. Two sites (SEQ ID NO: 7 and SEQ ID NO: 8) were selected which are, respectively, 35,699 and 15,898 base pairs downstream of the DHFR coding sequence (Table 2).

Table 2. Example Recognition Sites For Engineered Meganucleases in the CHO DHFR Locus.

SEQ ID NO:	Target Site Sequences	Location Relative to CHO DHFR Coding Sequence
7	5' -TAAGGCCTCATATGAAAATATA-3'	35,699 bp downstream
8	5' -ATAGATGTCTTGCATACTCTAG-3'	15,898 bp downstream

1. Meganucleases that recognize SEQ ID NO: 7 and SEQ ID NO: 8

An engineered meganuclease (SEQ ID NO: 9) was produced which recognizes

and cleaves SEQ ID NO: 7. This meganuclease is called "CHO-23/24". A second engineered meganuclease (SEQ ID NO: 10) was produced which recognizes and cleaves SEQ ID NO: 8. This meganuclease is called "CHO-51/52." Each meganuclease comprises an N-terminal nuclease-localization signal derived from SV40, a first meganuclease subunit, a linker sequence, and a second meganuclease subunit.

2. Site-specific cleavage of plasmid DNA by meganucleases CHO-23/24 and CHO-51/52

CHO-23/24 and CHO-51/52 were evaluated using a direct-repeat recombination assay as described previously (Gao *et al.* (2010), *Plant J.* 61(1):176-87, Figure 7A). A defective GFP reporter cassette was generated by first cloning a 5' 480 bp fragment of the GFP gene into NheI/HindIII-digested pcDNA5/FRT (Invitrogen Corp., Carlsbad, CA) resulting in the plasmid pGF. Next, a 3' 480 bp fragment of the GFP gene (including a 240 bp sequence duplicated in the 5' 480 bp fragment) was cloned into BamHI/XhoI-digested pGF. The resulting plasmid, pGFFP, consists of the 5' two-thirds of the GFP gene followed by the 3' two-thirds of the GFP gene, interrupted by 24 bp of the pcDNA5/FRT polylinker. To insert the meganuclease recognition sites, complementary oligonucleotides comprising the sense and anti-sense sequence of each recognition site were annealed and ligated into HindIII/BamHI-digested pGFFP.

The coding sequences of the engineered meganucleases were inserted into the mammalian expression vector pCP under the control of a constitutive (CMV) promoter. Chinese hamster ovary (CHO) cells at approximately 90% confluence were transfected in 96-well plates with 150 ng pGFFP reporter plasmid and 50 ng of meganuclease expression vector or, to determine background, 50 ng of empty pCP, using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen Corp., Carlsbad, CA). To determine transfection efficiency, CHO cells were transfected with 200 ng pCP GFP. Cells were washed in PBS 24 h post-transfection, trypsinized and resuspended in PBS supplemented with 3% fetal bovine serum. Cells were assayed for GFP activity using a Cell Lab Quanta SC MPL flow cytometer and the accompanying Cell Lab Quanta analysis software (Beckman Coulter, Brea, CA).

Results are shown in Figure 7B. It was found that both of the engineered meganucleases were able to cleave their intended recognition sites significantly above background within the context of a plasmid-based reporter assay.

3. Site-specific cleavage of CHO DHFR locus by meganucleases CHO-23/24 and CHO-

51/52

To determine whether or not CHO-23/24 and CHO-51/52 are capable of cleaving their intended target sites in the CHO DHFR locus, we screened genomic DNA from CHO cells expressing either CHO-23/24 or CHO-51/52 to identify evidence of chromosome
5 cleavage at the intended target site. This assay relies on the fact that chromosomal DNA breaks are frequently repaired by NHEJ in a manner that introduces mutations at the site of the DNA break. These mutations, typically small deletions or insertions (collectively known as “indels”) leave a telltale scar that can be detected by DNA sequencing (Gao *et al.* (2010), *Plant J.* 61(1):176-87).

10 CHO cells were transfected with mRNA encoding CHO-23/24 or CHO-51/52. mRNA was prepared by first producing a PCR template for an *in vitro* transcription reaction (SEQ ID NO: 20 and SEQ ID NO: 21). Each PCR product included a T7 promoter and 609 bp of vector sequence downstream of the meganuclease gene. The PCR product was gel purified to ensure a single template. Capped (m7G) RNA was generated using the RibomAX
15 T7 kit (Promega Corp., Fitchburg, WI) according to the manufacturer’s instructions and. Ribo m7G cap analog (Promega Corp., Fitchburg, WI) was included in the reaction and 0.5 µg of the purified meganuclease PCR product served as the DNA template. Capped RNA was purified using the SV Total RNA Isolation System (Promega Corp., Fitchburg, WI) according to the manufacturer’s instructions.

20 1.5x10⁶ CHO-K1 cells were nucleofected with 3x10¹² copies of CHO-23/24 or CHO-51/52 mRNA (2x10⁶ copies/cell) using an Amaxa Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program according to the manufacturer’s instructions. 48 hours post-transfection, genomic DNA was isolated from the cells using a FlexiGene kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. The genomic DNA
25 was then subjected to PCR to amplify the corresponding target site. In the case of cells transfected with mRNA encoding CHO-23/24, the forward and reverse PCR primers were SEQ ID NO: 16 and SEQ ID NO: 17. In the case of cells transfected with mRNA encoding CHO-51/52, the forward and reverse PCR primers were SEQ ID NO: 18 and SEQ ID NO: 19. PCR products were gel purified and cloned into pUC-19. 40 plasmids harboring PCR
30 products derived from cells transfected with CHO-23/24 mRNA were sequenced, 13 of which were found to have mutations in the CHO-23/24 target site (Figure 7C). 44 plasmids harboring PCR products derived from cells transfected with CHO-51/52 mRNA were sequenced, 10 of which were found to have mutations in the CHO-51/52 target site (Figure 7D). These results indicate that CHO-23/24 and CHO-51/52 are able to cut their intended

target sites downstream of the CHO DHFR gene.

4. Site-specific integration into the CHO DHFR locus using an engineered meganuclease

To evaluate the efficiency of DNA insertion into the CHO DHFR locus using an
5 engineered meganuclease, we prepared a donor plasmid (SEQ ID NO: 11) comprising an
EcoRI restriction enzyme site flanked by DNA sequence homologous to the CHO-51/52
recognition site (Figure 8A). Specifically, the donor plasmid of SEQ ID NO: 11 comprises a
pUC-19 vector harboring a homologous recombination cassette inserted between the KpnI
and HindIII restriction sites. The homologous recombination cassette comprises, in 5'- to 3'-
10 order: (i) 543 base pairs of DNA identical to the sequence immediately upstream of the
CHO-51/52 cut site, including the upstream half-site of the CHO-51/52 recognition sequence
and the four base pair "center sequence" separating the two half-sites comprising the CHO-
51/52 recognition sequence; (ii) an EcoRI restriction enzyme site (5'-GAATTC-3'); and iii)
461 base pairs of DNA identical to the sequence immediately downstream of the CHO-51/52
15 cut site, including the downstream half-site of the CHO-51/52 recognition sequence and the
four base pair "center sequence" separating the two half-sites comprising the CHO-51/52
recognition sequence. Note that this results in a duplication of the four base pair "center
sequence" (5'-TTGC-3') to maximize the likelihood of strand invasion by the 3' overhangs
generated by CHO-51/52 cleavage. We have discovered that donor plasmids comprising
20 such a duplication of the center sequence are optimal substrates for gene targeting by
homologous recombination.

mRNA encoding CHO-51/52 was prepared as described above. 1.5×10^6 CHO-K1
cells were nucleofected with 3×10^{12} copies of CHO 51-52 mRNA (2×10^6 copies/cell) and
1.5 μ g of the donor plasmid (SEQ ID NO: 11). Nucleofection was performed using an
25 Amaxa Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program
according to the manufacturer's instructions. 48 hours post-transfection, genomic DNA was
isolated from the cells using a FlexiGene kit (Qiagen, Hilden, Germany) according to the
manufacturer's instructions. The DNA was subjected to PCR using primers flanking the
CHO-51/52 recognition site (SEQ ID NO: 18 and SEQ ID NO: 19). Importantly, these
30 primers are beyond the limits of homologous sequence carried in the donor plasmid and,
therefore, will amplify only the chromosomal DNA sequence and not the donor plasmid.
PCR products were cloned into a pUC-19 plasmid and 48 clones were purified and digested
with EcoRI (Figure 8B). 10 plasmids yielded a restriction pattern consistent with the
insertion of an EcoRI site into the CHO-51/52 recognition sequence. These data demonstrate

that it is possible to use CHO-51/52 to precisely insert DNA downstream of the CHO DHFR gene at SEQ ID NO: 8.

5. Site-specific integration of an engineered target sequence into the CHO DHFR locus

5 A donor plasmid (SEQ ID NO: 25) was produced comprising an FRT sequence (SEQ ID NO: 5) adjacent to a zeocin resistance gene under the control of an SV40 early promoter (Figure 9A). This cassette was flanked by DNA sequence homologous to the CHO DHFR locus immediately upstream or downstream of the CHO-23/24 recognition sequence. CHO cells were co-transfected with this donor plasmid and mRNA encoding CHO-23/24 as described above. 72 hours post-transfection, zeocin-resistant cells were cloned by limiting dilution and expanded for approximately 3 weeks. Clonal populations were then screened by PCR using a first primer in the SV40 promoter (SEQ ID NO: 26) and a second primer in the DHFR locus (SEQ ID NO: 16) to identify cell lines carrying the FRT/Zeocin sequence downstream of the DHFR gene. One such cell line carrying the integrated FRT Insertion target sequence was subsequently co-transfected with a second donor plasmid (SEQ ID NO: 15 27) and a plasmid encoding Flp recombinase. SEQ ID NO: 27 comprises a GFP gene under the control of a CMV promoter, a FRT sequence, and a non-functional hygromycin resistance gene lacking an ATG start codon. Flp-stimulated recombination between FRT sites in the genome and the plasmid resulted in the incorporation of the entire plasmid sequence into the CHO genome at the site of the engineered target sequence. Such recombination restored function to the hygromycin-resistance gene by orientating it downstream of an ATG start codon integrated as part of the engineered target sequence. As such, successful integrations could be selected using hygromycin.

25 Hygromycin-resistant cells were cloned by limiting dilution and 24 individual clonal lines were assayed by PCR using a first primer in the hygromycin-resistance gene (SEQ ID NO: 28). All 24 clones yielded the expected PCR product (Figure 9B), indicating that the GFP gene expression cassette was successfully inserted into the DHFR engineered target sequence in all cases. The 24 cell lines were then evaluated by flow cytometry and were found to express consistent levels of GFP (Figure 9C).

30

6. Transgene Amplification

A GFP-expressing CHO line produced as described above was seeded at a density of 3×10^5 cells/mL in 30 mL of media containing 50nM MTX. Cells were cultured for 14 days before being re-seeded at the same density in media containing 100nM MTX. Cells

were cultured for another 14 days before being re-seeded in media containing 250nM MTX. Following 14 days in culture, GFP expression in the treated cells was evaluated by flow cytometry and compared to GFP expression in the parental (pre-MTX) cell population (Figure 10A). It was found that the MTX-treated cells had a distinct sub-population in which GFP expression was significantly increased. Individual high-expression cells from the MTX-treated population were then isolated using a cell sorter and 5 clones were expanded for 14 days in the absence of MTX. GFP expression in the 5 clonal cell populations was then evaluated by flow cytometry and compared with the parental (pre-MTX) cell population. It was found that the MTX-treated clones had approximately 4-6 times the GFP intensity as the pre-MTX cells. Quantitative PCR was then performed using a primer set specific for the GFP gene and it was found that the MTX-treated clones all had approximately 5-9 times as many copies of the GFP gene as the pre-MTX population. These data provide conclusive evidence that a transgene inserted downstream of the CHO DHFR gene can be amplified by treatment with MTX.

15

7. Stability of Gene Amplification

The five clonal cell lines expressing high levels of GFP that were produced in (6) above were then passaged for a period of 14 weeks in media with or without 250nM MTX to evaluate the stability of gene amplification. GFP intensity was determined on a weekly basis and the quantitative PCR assay used to determine GFP gene copy number described above was repeated at the end of the 14 week evaluation period. As expected, the clones passaged in media with MTX maintained a high level of GFP expression with no clone deviating more than 20% from the GFP intensity determined in week 1. Quantitative PCR revealed that gene copy number likewise deviated by less than 20% for all clones. Surprisingly, gene amplification was equally stable in cell lines grown in media lacking MTX. Contrary to what would have been predicted based on the existing art, GFP gene expression was not reduced by more than 18% in any of the five cell lines over the 14 week evaluation period. Gene copy number determined by quantitative PCR was also stable with less than 24% deviation over time for all of the cell lines. These results indicate that a transgene amplified in the CHO DHFR locus is stable for an extended period of time, obviating the need to grow the cells in toxic selection agents that that could contaminate bioproduct formulations.

30

EXAMPLE 2

Insertion of an engineered target sequence into the CHO DHFR or GS gene coding regions

As diagrammed in Figure 4, an alternative method for targeting a sequence of interest to an amplifiable locus involves the production of a cell line in which a portion of a selectable gene is replaced by an engineered target sequence. The advantage of this approach is that the subsequent insertion of a sequence of interest can be coupled with reconstitution of the selectable gene so that cell lines harboring the properly targeted sequence of interest can be selected using the appropriate media conditions. A cell line harboring such an engineered target sequence can be produced using nuclease-induced homologous recombination. In this case, a site-specific endonuclease which cuts a recognition sequence near or within the selectable gene sequence is preferred.

1. Engineered meganucleases that cut within the DHFR or GS genes.

A meganuclease called "CHO-13/14" (SEQ ID NO: 12) was produced which cuts a recognition sequence in the CHO DHFR gene (SEQ ID NO: 13). The recognition sequence is in an intron between Exon 2 and Exon 3 of CHO DHFR. A meganuclease called "CGS-5/6" (SEQ ID NO: 14) was produced which cuts a recognition sequence in the CHO GS gene (SEQ ID NO: 15). Each meganuclease comprises an N-terminal nuclease-localization signal derived from SV40, a first meganuclease subunit, a linker sequence, and a second meganuclease subunit.

2. Site-specific cleavage of plasmid DNA by meganucleases CHO-13/14 and CGS-5/6

CHO-13/14 and CGS-5/6 were evaluated using a direct-repeat recombination assay as described in Example 1 (Figure 7A). Both meganucleases were found to efficiently cleave their intended recognition sequences within the context of a plasmid-based reporter assay (Figure 7B).

3. Site-specific cleavage of the CHO GS gene by CGS-5/6

CHO cells were transfected with mRNA encoding CGS-5/6. mRNA was prepared by first producing a PCR template for an *in vitro* transcription reaction (SEQ ID NO: 22). Each PCR product included a T7 promoter and 609 bp of vector sequence downstream of the meganuclease gene. The PCR product was gel purified to ensure a single template. Capped (m7G) RNA was generated using the RiboMAX T7 kit (Promega Corp., Fitchburg, WI)

according to the manufacturer's instructions and. Ribo m7G cap analog (Promega Corp., Fitchburg, WI) was included in the reaction and 0.5 µg of the purified meganuclease PCR product served as the DNA template. Capped RNA was purified using the SV Total RNA Isolation System (Promega Corp., Fitchburg, WI) according to the manufacturer's

5 instructions.

1.5x10⁶ CHO-K1 cells were nucleofected with 3x10¹² copies of CGS-5/6 using an Amaxa Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program according to the manufacturer's instructions. 48 hours post-transfection, genomic DNA was isolated from the cells using a FlexiGene kit (Qiagen, Hilden, Germany) according to the

10 manufacturer's instructions. The genomic DNA was then subjected to PCR to amplify the CGS-5/6 target site using the primers of SEQ ID NO: 23 and SEQ ID NO: 24. The PCR products were cloned into a pUC-19 plasmid and 94 plasmids harboring PCR products were digested with the BssSI restriction enzyme, which recognized and cuts the sequence 5'-CTCGTG-3' found within the CGS-5/6 recognition sequence. 17 plasmids were found to be

15 resistant to BssSI, suggesting that the CGS-5/6 recognition site was mutated. These 17 plasmids were sequenced to confirm the existence of indels or point mutations within the CGS-5/6 recognition sequence (Figure 7C). These results indicate that CGS-5/6 is able to cut its intended target site within the CHO GS gene. Because the CGS-5/6 recognition sequence is within an exon in the GS coding sequence, many of the mutations introduced by CGS-5/6

20 are expected to frameshift the GS gene. Therefore, CGS-5/6 is useful for knocking-out CHO GS to produce GS (-/-) cell lines. Such cell lines are useful because they are amenable to GS selection and amplification for producing biomanufacturing cell lines.

EXAMPLE 3

25 Meganucleases for targeting gene insertion to the CHO GS locus

1. Engineered meganucleases that cut downstream of the CHO GS gene.

An engineered meganuclease called "CHOX-45/46" (SEQ ID NO: 29) was produced which recognizes a DNA sequence (SEQ ID NO: 30) approximately 7700 base

30 pairs downstream of the CHO GS coding sequence. CHO cells were transfected with mRNA encoding CHOX-45/46 as described in Example 2. 72 hours post transfection, genomic DNA was extracted from the transfected cell pool and the region downstream of the CHO GS gene was PCR amplified using a pair of primers (SEQ ID NO: 31 and SEQ ID NO: 32) flanking the CHOX-45/46 recognition sequence. PCR products were then cloned and 24 cloned

products were sequenced. It was found that 14 of the 24 clones PCR products (58.3%) had large mutations in the sequence consistent with meganuclease-induced genome cleavage followed by mutagenic repair by non-homologous end-joining. From these data, we conclude that the CHOX-45/46 meganuclease is able to specifically cleave a DNA site downstream of the CHO GS gene coding sequence and will likely be able to target the insertion of transgenes to this amplifiable locus in the genome.

CLAIMS

1. A recombinant mammalian cell comprising an engineered target site stably integrated within selectable gene within an amplifiable locus, wherein the engineered target site disrupts the function of the selectable gene and wherein the engineered target site comprises a recognition sequence for a site specific endonuclease.
2. The recombinant cell of claim 1, wherein the selectable gene is glutamine synthetase (GS) and the locus is methionine sulphoximine (MSX) amplifiable.
3. The recombinant cell of claim 1, wherein the selectable gene is dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amplifiable.
4. The recombinant cell of any one of claims 1-3, wherein the engineered target site is inserted into an exon of the selectable gene.
5. The recombinant cell of any one of claims 1-4, wherein the site specific endonuclease is a meganuclease, a zinc finger nuclease or TAL effector nuclease.
6. The recombinant cell of any one of claims 1-5, further comprising the site specific endonuclease.
7. The recombinant cell of any one of claims 1 and 4-6, wherein the selectable gene is selected from the group consisting of Dihydrofolate Reductase, Glutamine Synthetase, Hypoxanthine Phosphoribosyltransferase, Threonyl tRNA Synthetase, Na,K-ATPase, Asparagine Synthetase, Ornithine Decarboxylase, Inosine-5'-monophosphate dehydrogenase, Adenosine Deaminase, Thymidylate Synthetase, Aspartate Transcarbamylase, Metallothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.
8. The recombinant cell of claim 7, wherein the selectable gene is amplifiable by selection with a selection agent selected from the group consisting of Methotrexate (MTX), Methionine sulphoximine (MSX), Aminopterin, hypoxanthine, thymidine, Borrelidin, Ouabain, Albizziin, Beta-aspartyl hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2'-deoxycoformycin, Fluorouracil, N-Phosphonacetyl-L-Aspartate (PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-

azauridine, pyrazofuran, hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine and triapine.

9. A recombinant mammalian cell comprising an engineered target site stably integrated proximal to a selectable gene within an amplifiable locus, wherein the engineered target site comprises a recognition sequence for a site specific endonuclease.

10. The recombinant cell of claim 9, wherein the engineered target site is downstream from the 3' regulatory region of the selectable gene.

11. The recombinant cell of claim 10, wherein the engineered target site is 0 to 100,000 base pairs downstream from the 3' regulatory region of the selectable gene.

12. The recombinant cell of claim 10, wherein the engineered target site is upstream from the 5' regulatory region of the selectable gene.

13. The recombinant cell of claim 12, wherein the engineered target site is 0 to 100,000 base pairs upstream from the 5' regulatory region of the selectable gene.

14. The recombinant cell of any one of claims 9-13, wherein the selectable gene is glutamine synthetase (GS) and the locus is methionine sulphoximine (MSX) amplifiable.

15. The recombinant cell of any one of claims 9-13, wherein the selectable gene is dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amplifiable.

16. The recombinant cell of any one of claims 9-15, wherein the site specific endonuclease is a meganuclease, a zinc finger nuclease, or a TAL effector nuclease.

17. The recombinant cell of any one of claims 9-16, further comprising the site specific endonuclease.

18. The recombinant cell of any one of claims 9-13 and 16-17, wherein the selectable gene is selected from the group consisting of Dihydrofolate Reductase, Glutamine Synthetase, Hypoxanthine Phosphoribosyltransferase, Threonyl tRNA Synthetase, Na,K-ATPase, Asparagine Synthetase, Ornithine Decarboxylase, Inosine-5'-monophosphate dehydrogenase, Adenosine Deaminase, Thymidylate Synthetase, Aspartate Transcarbamylase, Metallothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.

19. The recombinant cell of claim 18, wherein the selectable gene is amplifiable by selection with a selection agent selected from the group consisting of Methotrexate (MTX), Methionine sulphoximine (MSX), Aminopterin, hypoxanthine, thymidine, Borrelidin, Ouabain, Albizziin, Beta-aspartyl hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2' deoxycoformycin, Fluorouracil, N-Phosphonacetyl-L-Aspartate (PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-azauridine, pyrazofuran, hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine and triapine.

20. A method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising:

(a) providing a mammalian cell having an endogenous target site proximal to a selectable gene within the amplifiable locus, wherein the endogenous target site comprises:

(i) a recognition sequence for an engineered meganuclease;

(ii) a 5' flanking region 5' to the recognition sequence; and

(iii) a 3' flanking region 3' to the recognition sequence; and

(b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site;

(c) contacting the cell with a donor vector comprising from 5' to 3':

(i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site;

(ii) an exogenous sequence; and

(iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site;

whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a modified cell.

21. The method of claim 20, further comprising growing the modified cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene.
22. The method of any one of claims 20-21, wherein the exogenous sequence comprises a gene of interest.
23. The method of any one of claims 20-22, wherein the endogenous target site is downstream from the 3' regulatory region of the selectable gene.
24. The method of claim 23, wherein the endogenous target site is 0 to 100,000 base pairs downstream from the 3' regulatory region of the selectable gene.
25. The method of any one of claims 20-22, wherein the endogenous target site is upstream from the 5' regulatory region of the selectable gene.
26. The method of claim 25, wherein the endogenous target site is 0 to 100,000 base pairs upstream from the 5' regulatory region of the selectable gene.
27. The method any one of claims 20-26, wherein the selectable gene is glutamine synthetase (GS) and the locus is methionine sulphoximine (MSX) amplifiable.
28. The method any one of claims 20-26, wherein the selectable gene is dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amplifiable.
29. The method any one of claims 20-26, wherein the selectable gene is selected from the group consisting of Dihydrofolate Reductase, Glutamine Synthetase, Hypoxanthine Phosphoribosyltransferase, Threonyl tRNA Synthetase, Na,K-ATPase, Asparagine Synthetase, Ornithine Decarboxylase, Inosine-5'-monophosphate dehydrogenase, Adenosine Deaminase, Thymidylate Synthetase, Aspartate Transcarbamylase, Metallothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.
30. The method of claim 29, wherein the selectable gene is amplifiable by selection with a selection agent selected from the group consisting of Methotrexate (MTX), Methionine sulphoximine (MSX), Aminopterin, hypoxanthine, thymidine, Borrelidin, Ouabain, Albizziin, Beta-aspartyl hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2'deoxycoformycin, Fluorouracil, N-Phosphonacetyl-L-Aspartate

(PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-azauridine, pyrazofuran, hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine and triapine.

31. A method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising:

(a) providing a mammalian cell having an endogenous target site proximal to a selectable gene within the amplifiable locus, wherein the endogenous target site comprises:

- (i) a recognition sequence for an engineered meganuclease;
- (ii) a 5' flanking region 5' to the recognition sequence; and
- (iii) a 3' flanking region 3' to the recognition sequence; and

(b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site;

(c) contacting the cell with an engineered target site donor vector comprising from 5' to 3':

- (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site;
- (ii) an exogenous sequence comprising an engineered target site; and
- (iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site;

whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a mammalian cell comprising the engineered target site;

(d) introducing a double-stranded break between the 5' and 3' flanking regions of the engineered target site;

(e) contacting the cell comprising the engineered target site with a sequence of interest donor vector comprising from 5' to 3':

(i) a donor 5' flanking region homologous to the 5' flanking region of the engineered target site;

(ii) an exogenous sequence comprising a sequence of interest; and

(iii) a donor 3' flanking region homologous to the 3' flanking region of the engineered target site;

whereby the donor 5' flanking region, the exogenous sequence comprising the sequence of interest and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the engineered target site by homologous recombination to provide an engineered mammalian cell comprising the sequence of interest.

32. The method of claim 31, further comprising growing the engineered mammalian cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene.

33. The method of any one of claims 31-32, wherein the sequence of interest comprises a gene.

34. The method of any one of claims 31-33, wherein the endogenous target site is downstream from the 3' regulatory region of the selectable gene.

35. The method of claim 34, wherein the endogenous target site is 0 to 100,000 base pairs downstream from the 3' regulatory region of the selectable gene.

36. The method of any one of claims 31-33, wherein the endogenous target site is upstream from the 5' regulatory region of the selectable gene.

37. The method of claim 36, wherein the endogenous target site is 0 to 100,000 base pairs upstream from the 5' regulatory region of the selectable gene.

38. The method any one of claims 31-36, wherein the selectable gene is glutamine synthetase (GS) and the locus is methionine sulphoximine (MSX) amplifiable.

39. The method any one of claims 31-36, wherein the selectable gene is dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amplifiable.
40. The method any one of claims 31-37, wherein the selectable gene is selected from the group consisting of Dihydrofolate Reductase, Glutamine Synthetase, Hypoxanthine Phosphoribosyltransferase, Threonyl tRNA Synthetase, Na,K-ATPase, Asparagine Synthetase, Ornithine Decarboxylase, Inosine-5'-monophosphate dehydrogenase, Adenosine Deaminase, Thymidylate Synthetase, Aspartate Transcarbamylase, Metallothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.
41. The method of claim 40, wherein the locus is Methotrexate (MTX), Methionine sulphoximine (MSX), Aminopterin, hypoxanthine, thymidine, Borrelidin, Ouabain, Albizziin, Beta-aspartyl hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2'-deoxycoformycin, Fluorouracil, N-Phosphonacetyl-L-Aspartate (PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-azauridine, pyrazofuran, hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine and/or triapine amplifiable.
42. A method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising:
- (a) providing a mammalian cell having an endogenous target site within a selectable gene within the amplifiable locus, wherein the endogenous target site comprises:
 - (i) a recognition sequence for an engineered meganuclease;
 - (ii) a 5' flanking region 5' to the recognition sequence; and
 - (iii) a 3' flanking region 3' to the recognition sequence; and
 - (b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site;
 - (c) contacting the cell with an engineered target site donor vector comprising from 5' to 3':
 - (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site;

(ii) an exogenous sequence comprising an engineered target site; and

(iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site;

whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a mammalian cell comprising the engineered target site;

(d) introducing a double-stranded break between the 5' and 3' flanking regions of the engineered target site;

(e) contacting the cell comprising the engineered target site with a sequence of interest donor vector comprising from 5' to 3':

(i) a donor 5' flanking region homologous to the 5' flanking region of the engineered target site;

(ii) an exogenous sequence comprising a sequence of interest; and

(iii) a donor 3' flanking region homologous to the 3' flanking region of the engineered target site;

whereby the donor 5' flanking region, the exogenous sequence comprising the sequence of interest and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the engineered target site by homologous recombination to provide a engineered mammalian cell comprising the sequence of interest.

43. The method of claim 42, further comprising growing the engineered mammalian cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene.

44. The method of any one of claims 42-43, wherein the sequence of interest comprises a gene.

45. The method of any one of claims 41-44, wherein the endogenous target site is within an intron of the selectable gene.

46. The method of any one of claims 41-44, wherein the endogenous target site is within an exon of the selectable gene.
47. The method any one of claims 42-46, wherein the selectable gene is glutamine synthetase (GS) and the locus is methionine sulphoximine (MSX) amplifiable.
48. The method any one of claims 42-46, wherein the selectable gene is dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amplifiable.
49. The method any one of claims 42-46, wherein the selectable gene is selected from the group consisting of Dihydrofolate Reductase, Glutamine Synthetase, Hypoxanthine Phosphoribosyltransferase, Threonyl tRNA Synthetase, Na,K-ATPase, Asparagine Synthetase, Ornithine Decarboxylase, Inosine-5'-monophosphate dehydrogenase, Adenosine Deaminase, Thymidylate Synthetase, Aspartate Transcarbamylase, Metallothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.
50. The method of claim 49, wherein the locus is Methotrexate (MTX), Methionine sulphoximine (MSX), Aminopterin, hypoxanthine, thymidine, Borrelidin, Ouabain, Albizziin, Beta-aspartyl hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2'deoxycoformycin, Fluorouracil, N-Phosphonacetyl-L-Aspartate (PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-azauridine, pyrazofuran, hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine and/or triapine amplifiable.
51. A recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 15.
52. A recombinant meganuclease comprising the amino acid sequence of SEQ ID NO: 15.
53. A recombinant meganuclease which recognizes and cleaves a recognition site having 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 14.
54. The recombinant meganuclease of claim 53, wherein the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 14.
55. A recombinant meganuclease comprising a polypeptide having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 9.

56. The recombinant meganuclease of 55, having the sequence of the meganuclease of SEQ ID NO: 9.
57. A recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 7.
58. The recombinant meganuclease of claim 57, wherein the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 7.
59. A recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 10.
60. The recombinant meganuclease of 59, comprising the polypeptide of SEQ ID NO: 10.
61. A recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 8.
62. The recombinant meganuclease of claim 61, wherein the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 8.
63. A recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 13.
64. The recombinant meganuclease of 63, comprising the polypeptide of SEQ ID NO: 13.
65. A recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 12.
66. The recombinant meganuclease of claim 65, wherein the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 12.
67. A recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 29.
68. The recombinant meganuclease of 67, comprising the polypeptide of SEQ ID NO: 29.
69. A recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 30.
70. The recombinant meganuclease of claim 69, wherein the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 30.

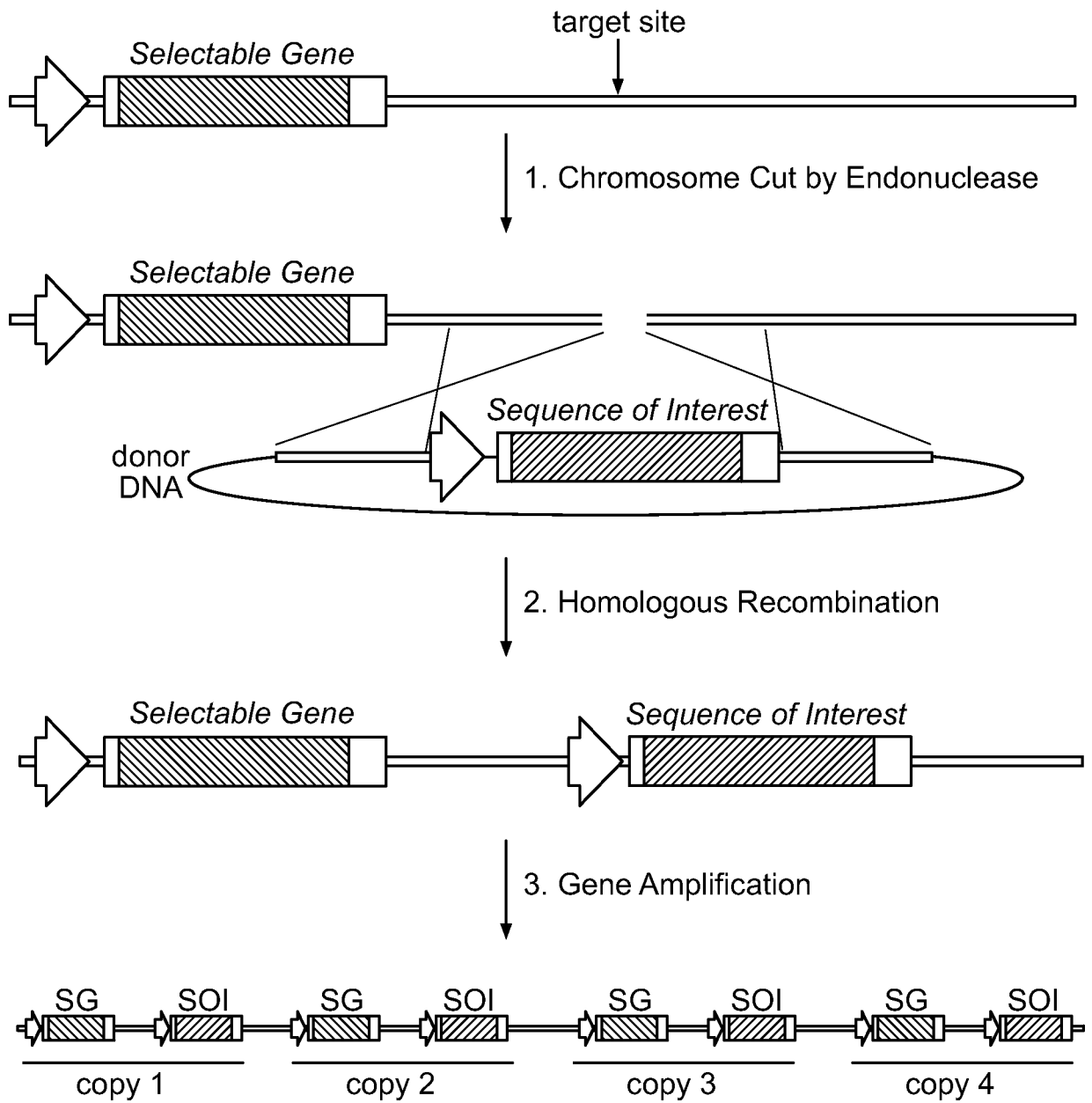


FIG. 1

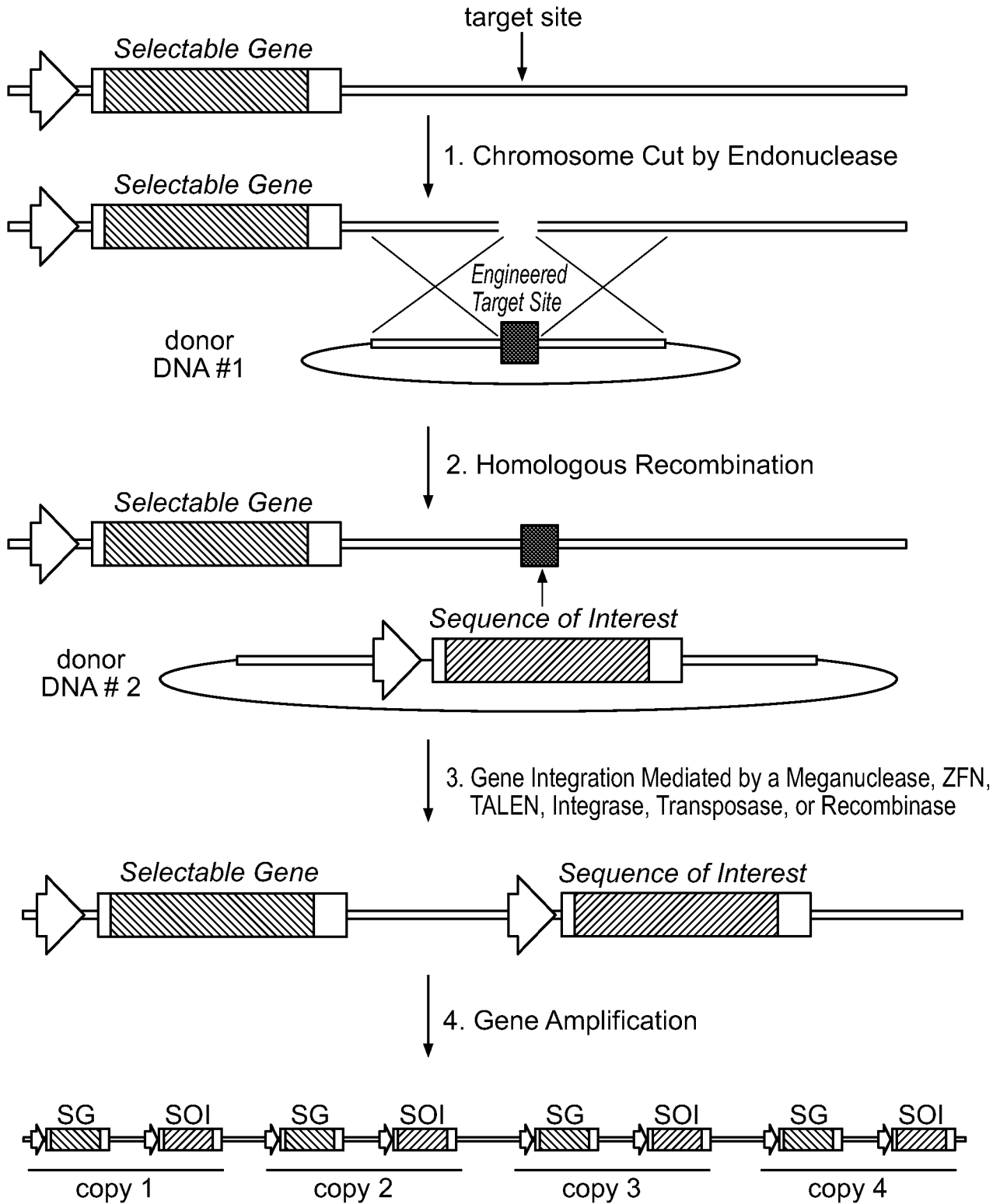


FIG. 3

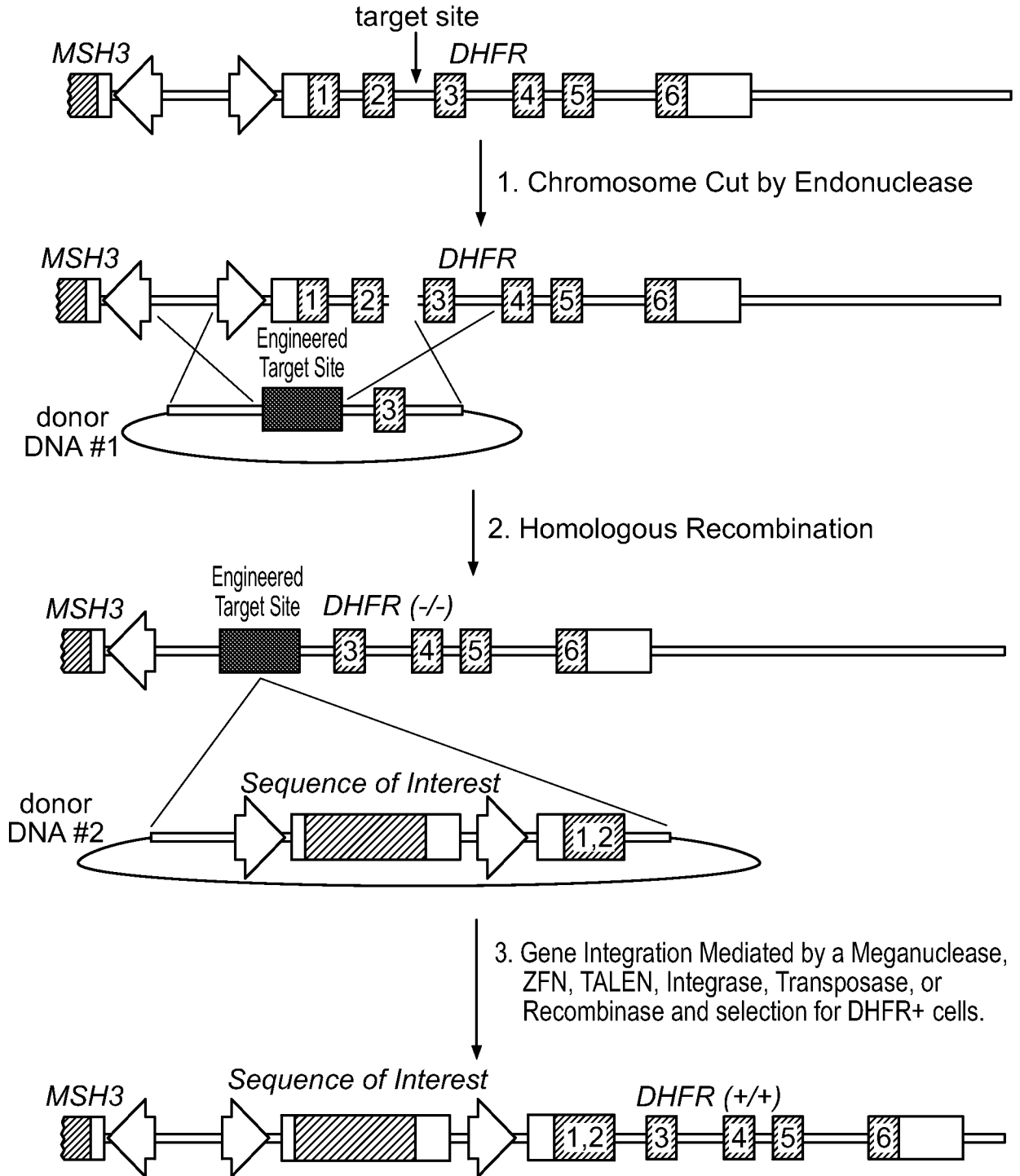


FIG. 4

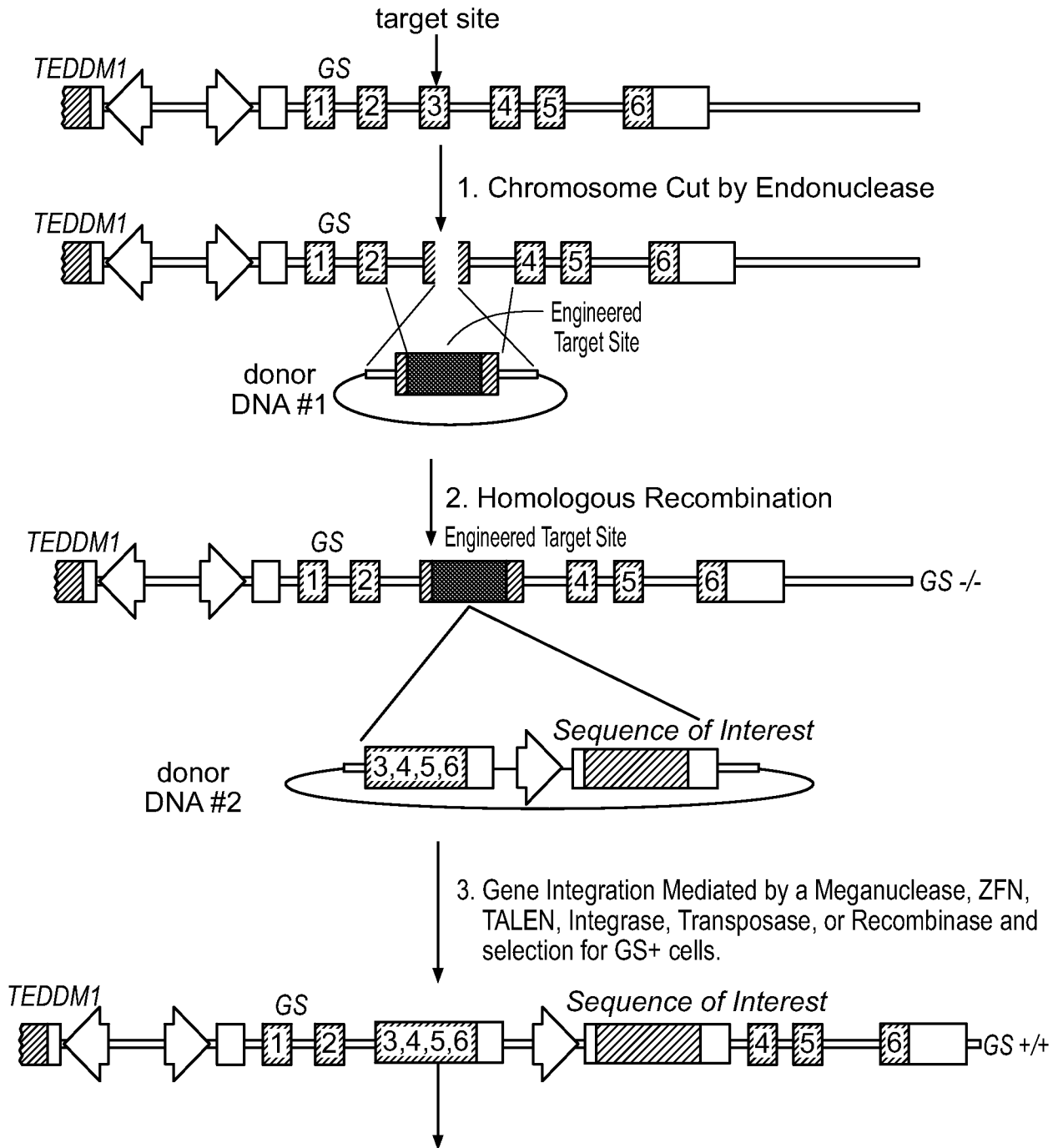


FIG. 5

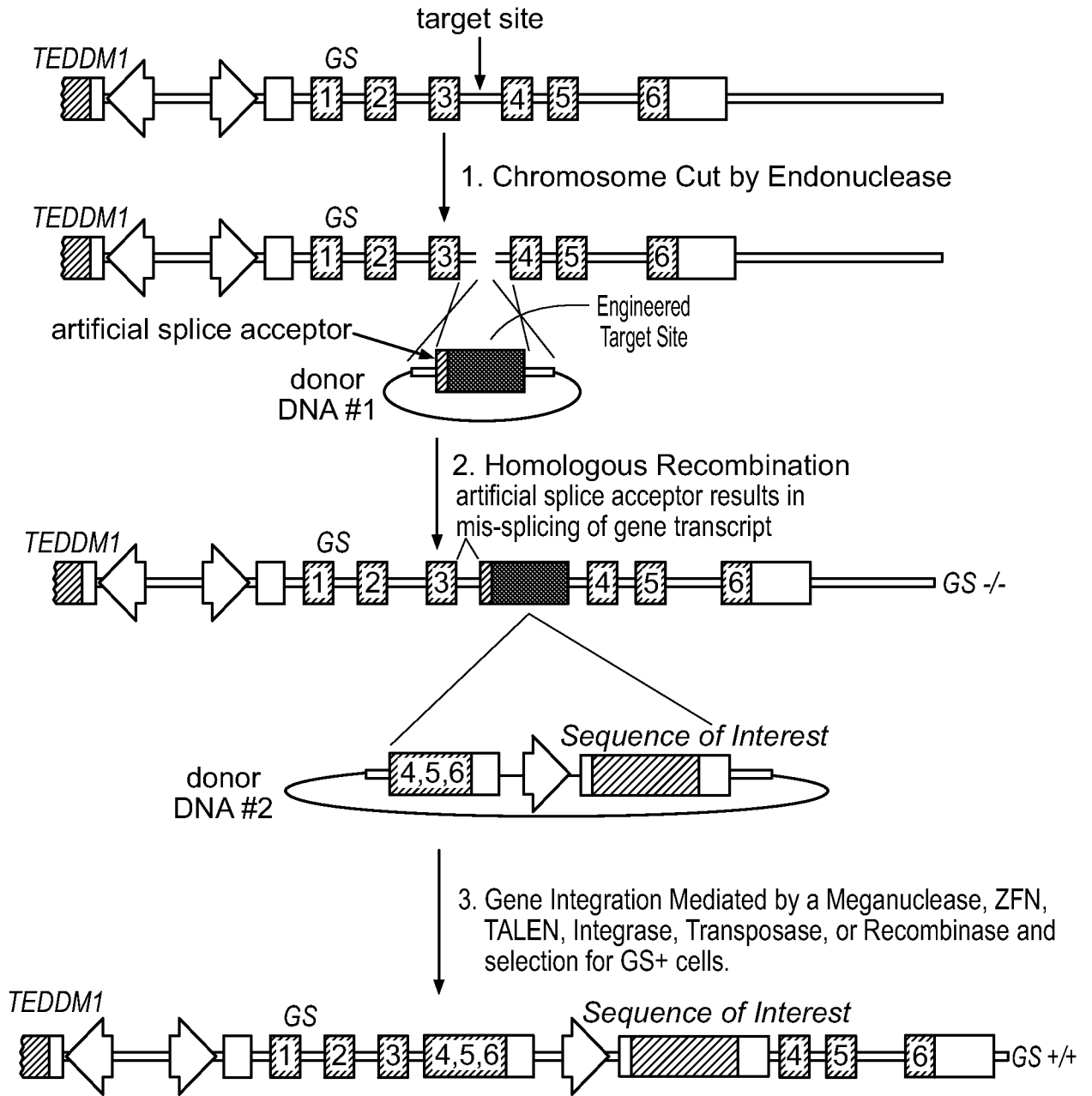


FIG. 6

7/18

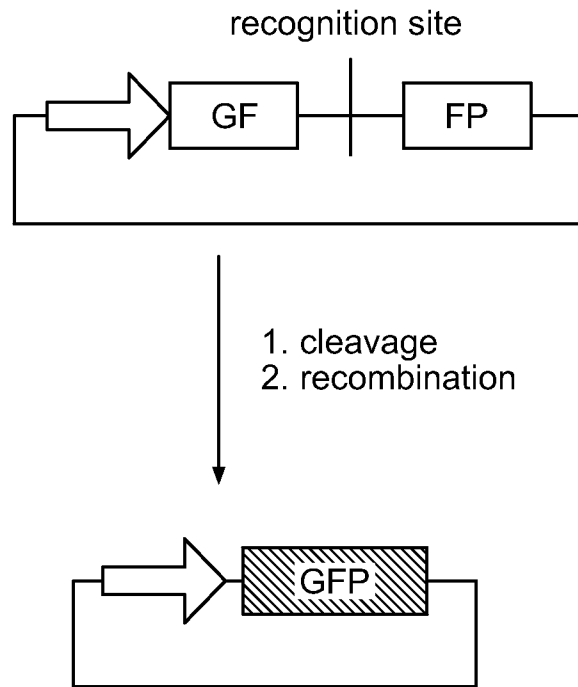


FIG. 7A

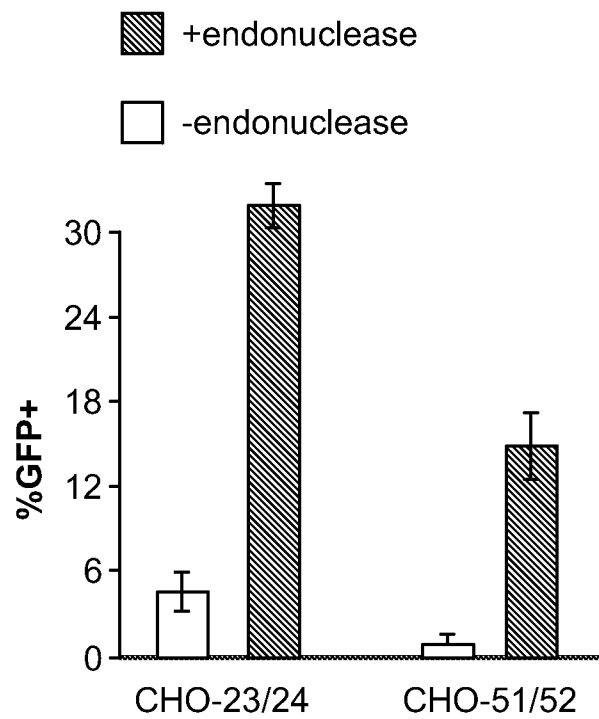


FIG. 7B

CHO-23/24 site

WT: GAA TGG GAA GTT CCAG AAT TTA AGG CCT CAT ATG AAA AATA TAA AGG CTT TCT
F8: GAA TGG GAA GTT CCAG AAT TTA A-----TAA AGG CTT TCT
F10: GAA TGG GAA GTT CCAG A-----AGG CTT TCT
G2: GAA TGG GAA GTT CCAG AAT TTA A-----TAA AGG CTT TCT
H4: GAA TGG GAA GTT CCAG A-----AGG CTT TCT
H5: GAA TGG GAA GTT CCAG AAT TTA A-----TAA AGG CTT TCT
G1: GAA TGG GAA GTT CCAG AAT TTA A-----TAA AGG CTT TCT
G3: -----561 bp deletion-----
G8: -----348 bp deletion-----
F1: -----562 bp deletion-----
G8: -----348 bp deletion-----
F1: -----562 bp deletion-----
E8: -----568 bp deletion-----
G12: -----587 bp deletion-----

FIG. 7C

CHO-51/52 site

WT TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGCTTGCATACTCTAGAGACTACAGATGCCGGCCCCAG
F3 TCTGAAAGCCAGAAAGAGCCCTAGACA-----TGCATACTCTAGAGACTACAGATGCCGGCCCCAG
G10 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGTC-----AGATGCCGGCCCCAG
F1 TCTGAAAGCCAGAAAGAGCCCTA-----CAGATGCCGGCCCCAG
E3 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGCTTGC-----ATGCCGGCCCCAG
E9 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGCTTGC-----TCTAGAGACTACAGATGCCGGCCCCAG
F11 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGCTT--ATACTCTAGAGACTACAGATGCCGGCCCCAG
F6 TCTGAAAGCCAGAAAGAGCCCTAGACAGAT-----GCCGGCCCCAG
G5 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGCTT-----TACAGATGCCGACCCCCAG
H10 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGCTTGC----TCTAGAGACTACAGATGCCGGCCCCAG
F11 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGCTT**T**GCATACTCTAGAGACTACAGATGCCGGCCCA

FIG. 7D

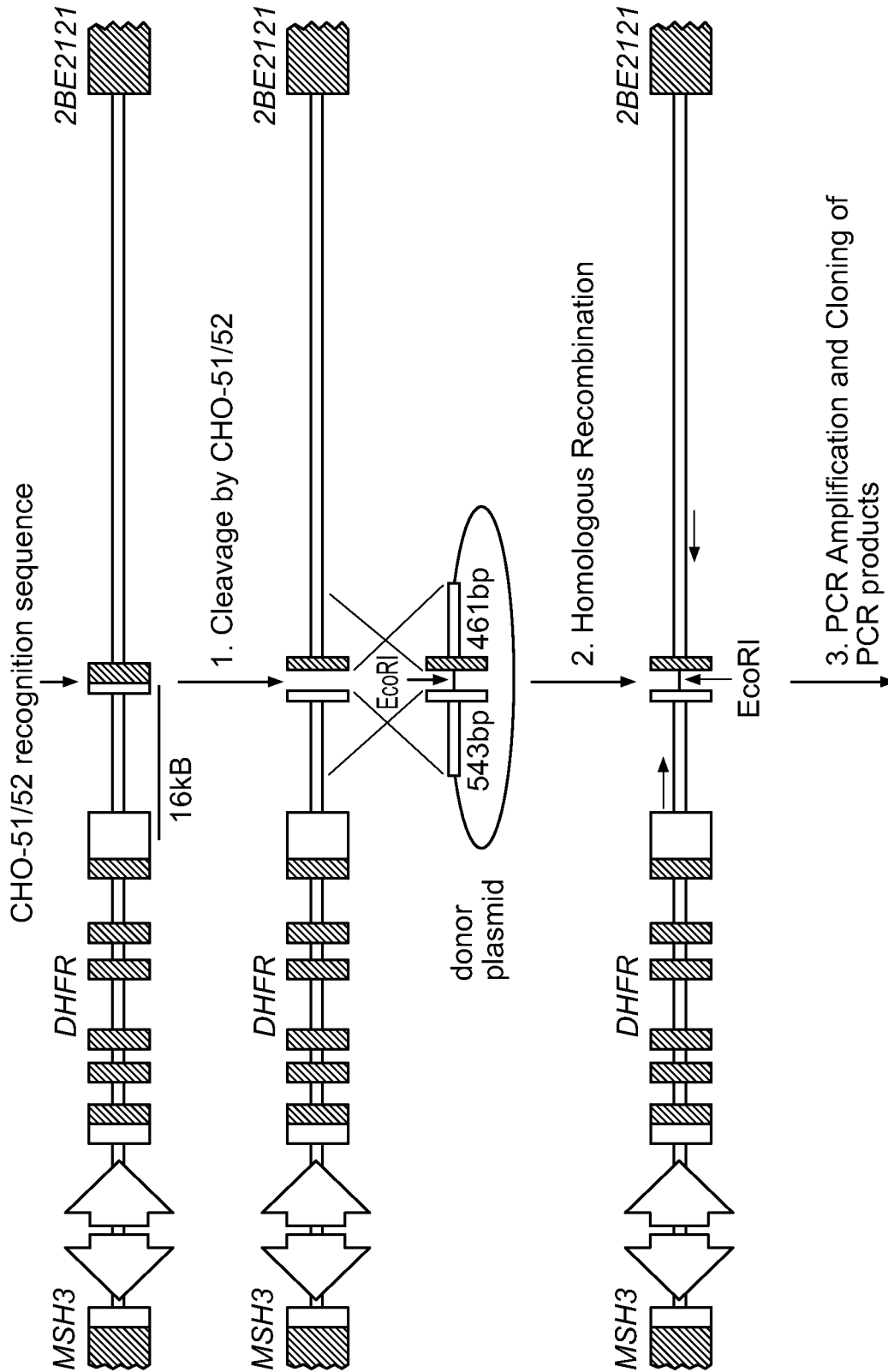


FIG. 8A

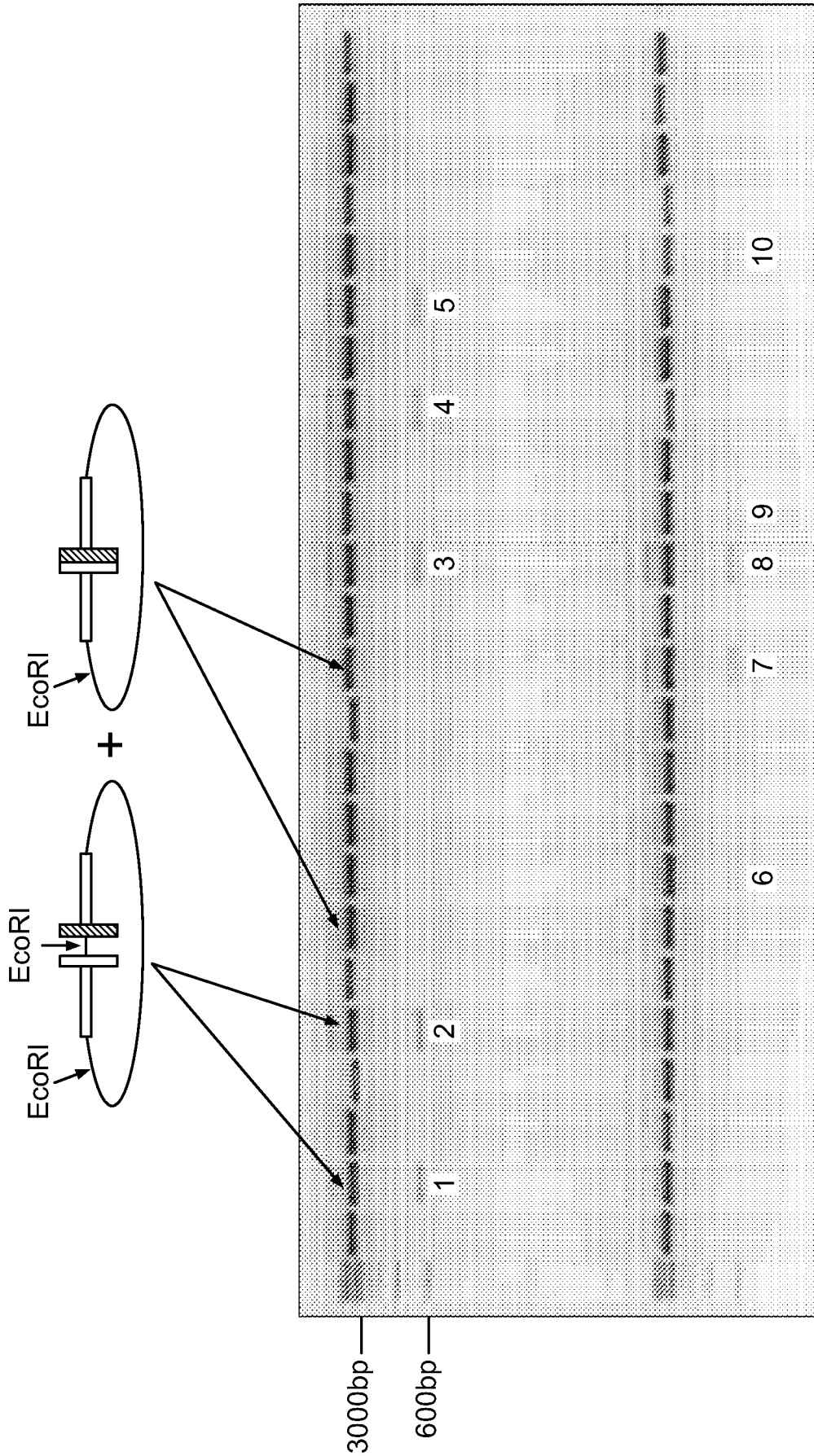


FIG. 8B

12/18

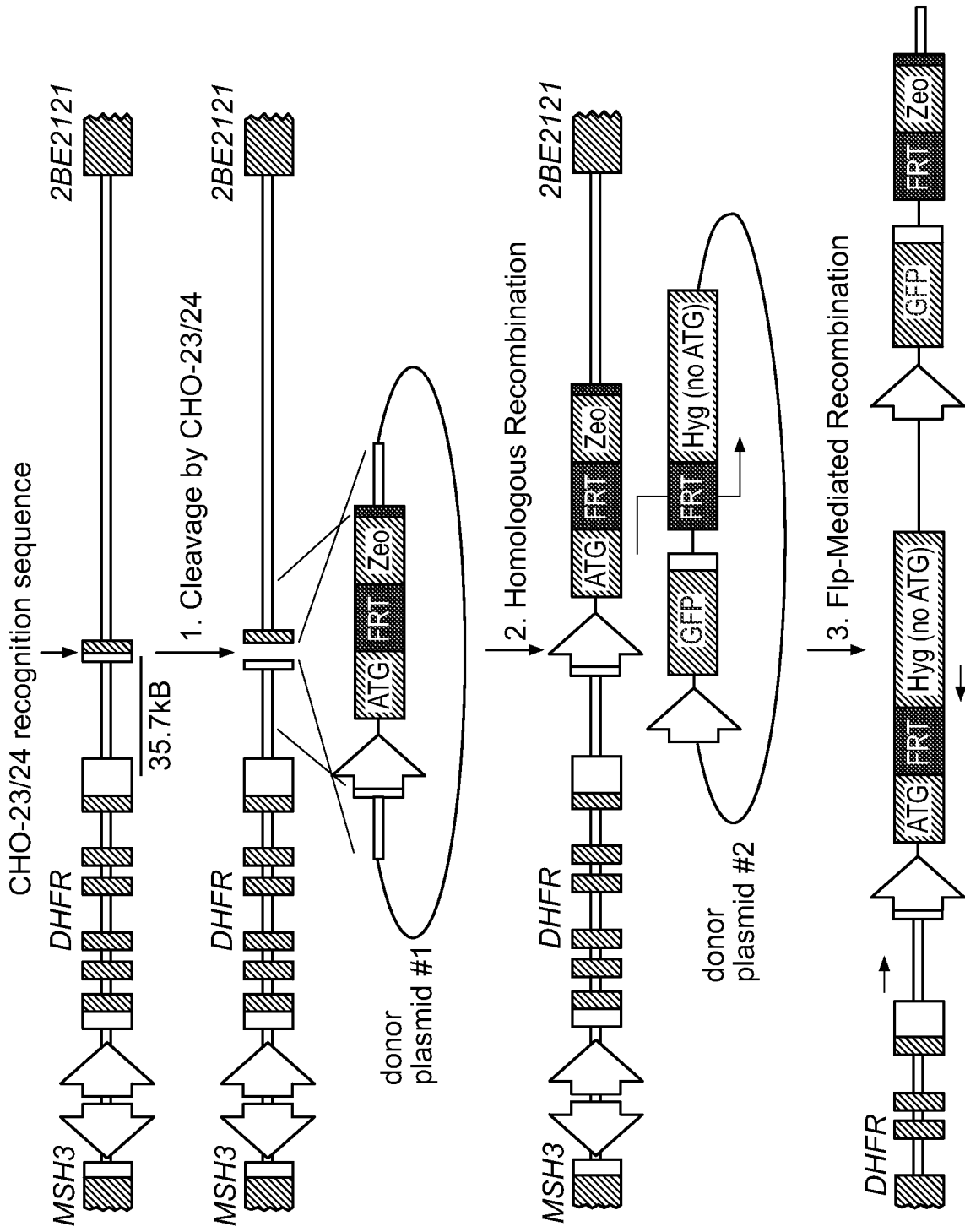


FIG. 9A

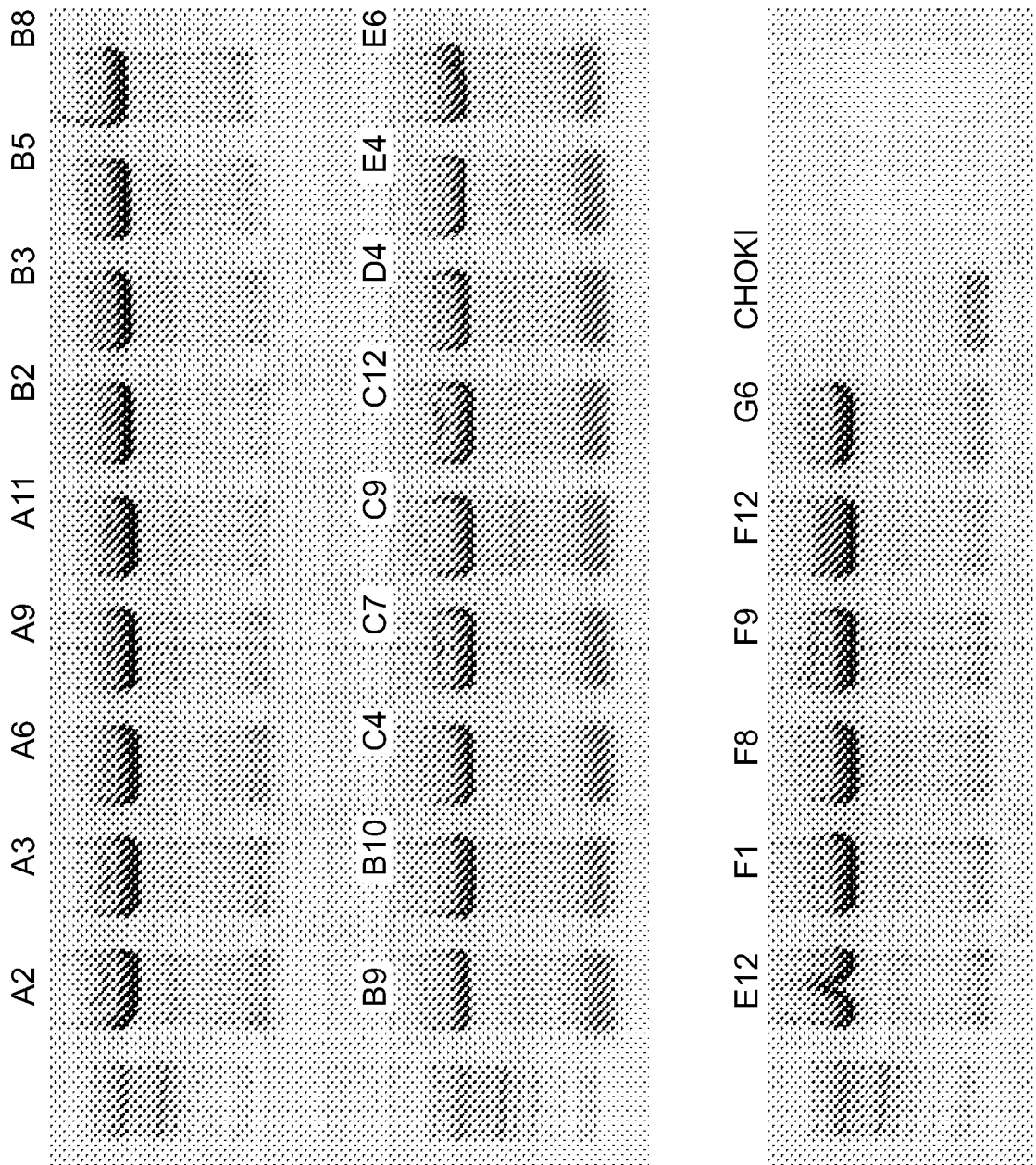


FIG. 9B

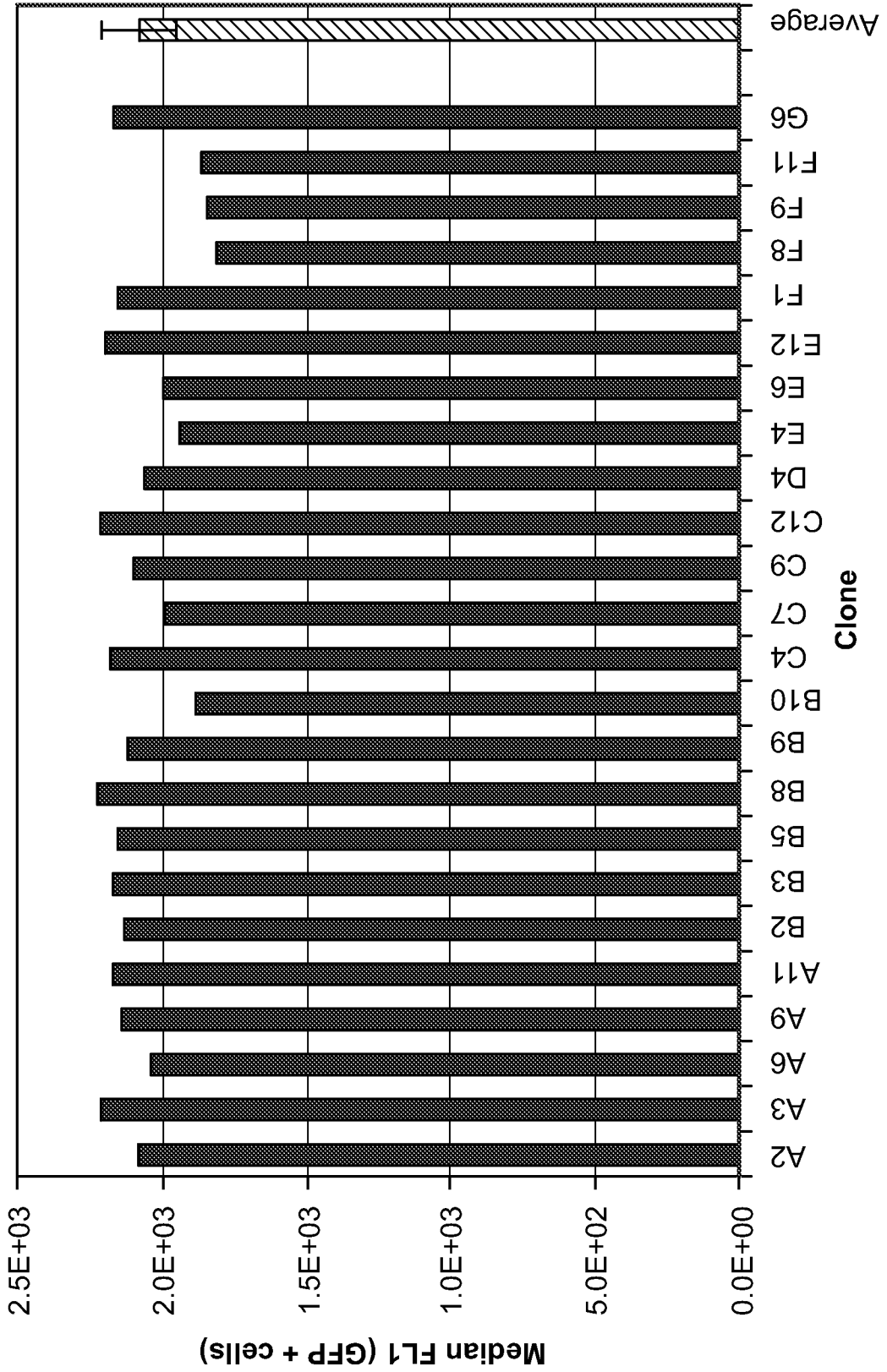


FIG. 9C

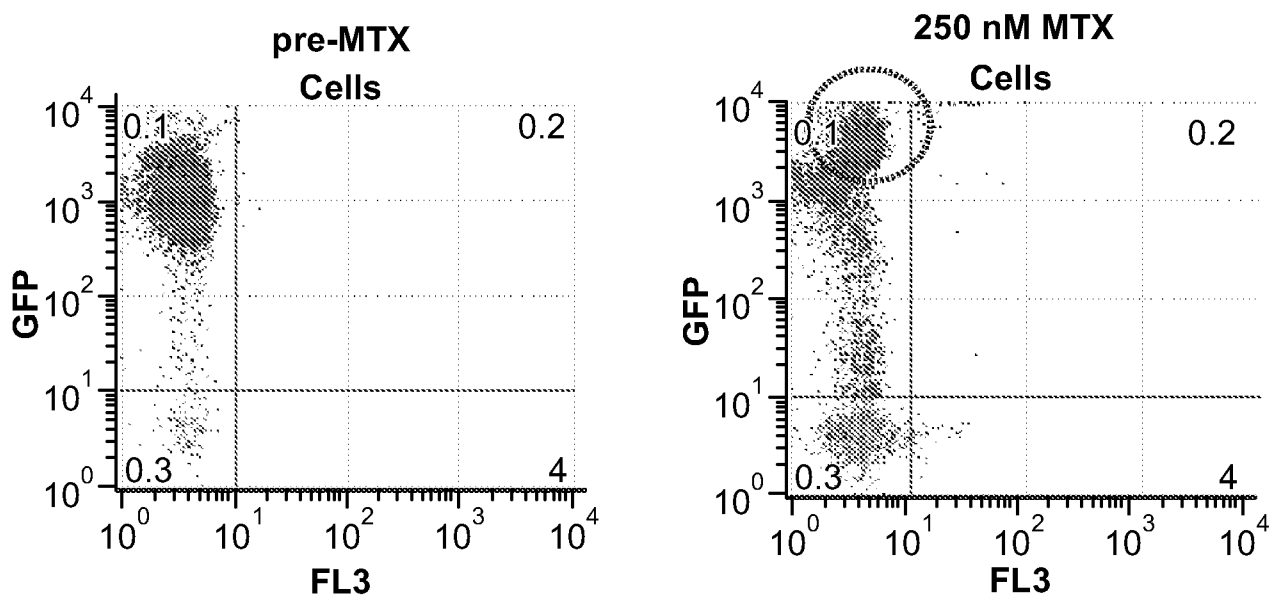


FIG. 10A

16/18

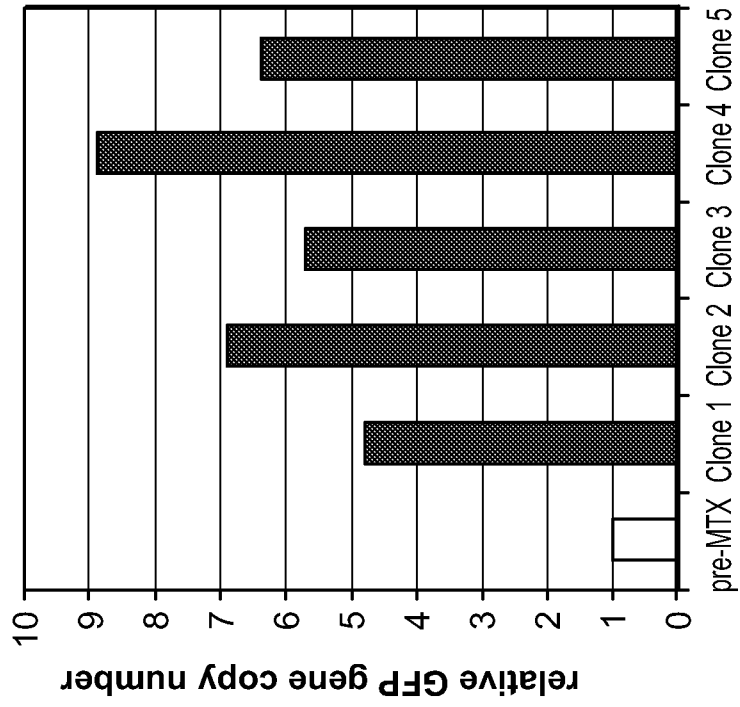


FIG. 10C

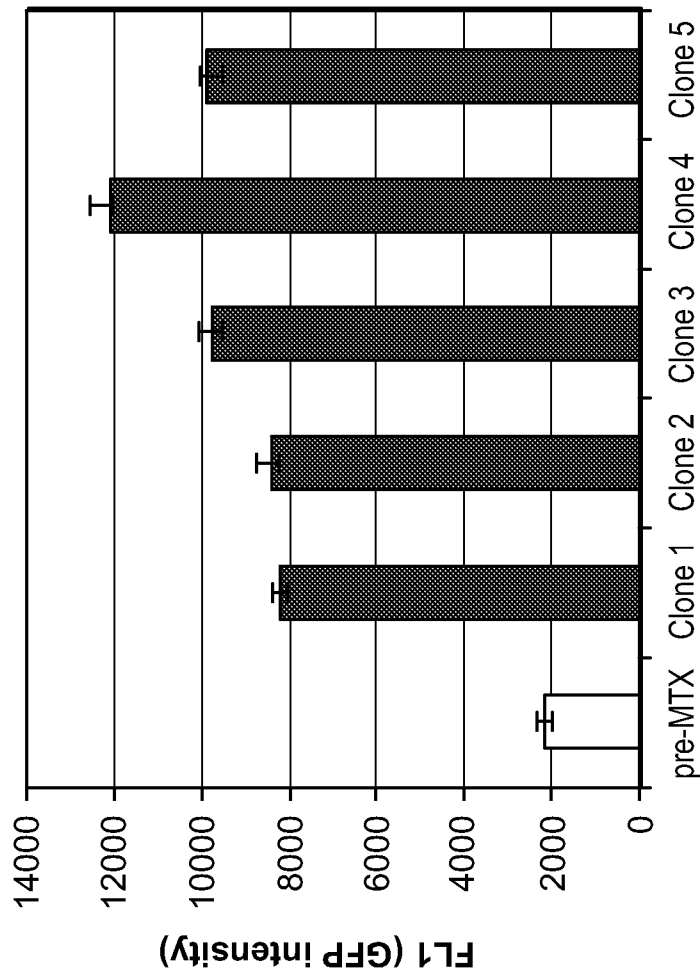


FIG. 10B

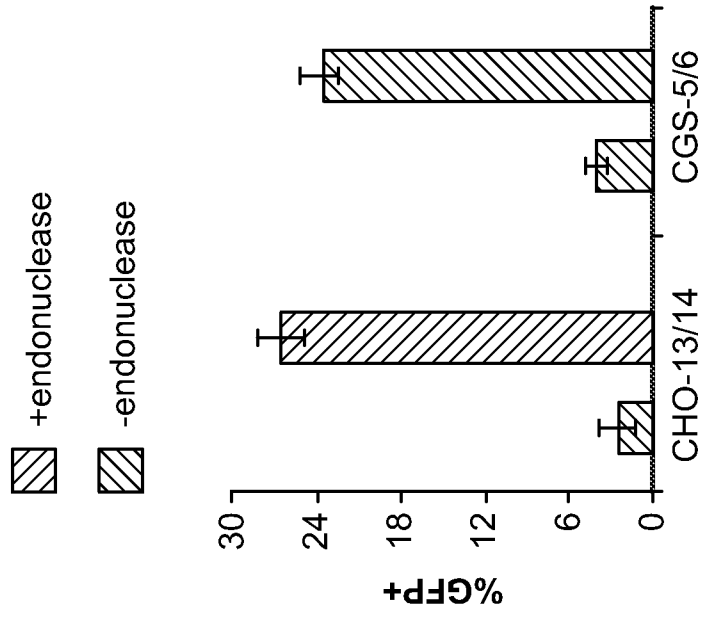


FIG. 11B

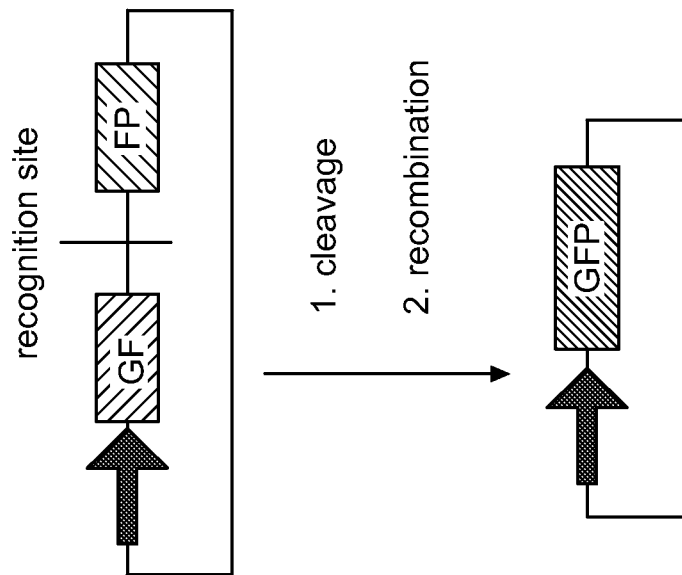


FIG. 11A

CGS-5/6 site

WT GATGCTTTATTCTAGAGACCAATTTAAGGCACCTCGTGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCC
3d2 GATGCTTTATTCTAGAGACCAATTTAAGGCACCTCGTGT----G-TAATGGACATGGTGAGCAACCAGCACCCC
6d4 GATGCTTTATTCTAGAGACCAATTTAAGGCACCTCGTGTAA-CGGATAATGGACATGGTGAGCAACCAGCACCCC
6g5 GATGCTTTATTCTAGAGACCAATTTAAGGCACCTC---AAACGGATAATGGACATGGTGAGCAACCAGCACCCC
3b7 GATGCTTTATTCTAGAGACCAATTTAAGGCACCTCGT--AAACGGATAATGGACATGGTGAGCAACCAGCACCCC
3d11 GATGCTTTATTCTAGAGACCAATTTAAGGCACCTCGT--AAACGGATAATGGACATGGTGAGCAACCAGCACCCC
3e5 GATGCTTTATTCTAGAGACCAATTTAAGGCACCTCGT--AAACGGATAATGGACATGGTGAGCAACCAGCACCCC
6f10 GATGCTTTATTCTAGAGACCAATTTAAGGCCT-GTGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCC
6a2 GATGCTTTATTCTAGAGACCAATTTAAGGCA**T**TCGTGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCC
6b3 GATGCTTTATTCTAGAGACCAATTTAAGGCAC**T**ATGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCC
6c9 GATGCTTTATTCTAGAGACCAATTTAAGGC**G**CTCGTGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCC
6f12 GATGCTTTATTCTAGAGACCAATTTAAGGCAC**T**CGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCC
6h7 GATGCTTTATTCTAGAGACCAATTTAAGGCAC**A**CGTGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCC
6H8 GATGCTTTATTCTAGAGACCAATTTAAGGCAC**T**CGTGT**G**TAAACGGATAATGGACATGGTGAGCAACCAGCAC
6d10 GATGCTTTATTCTAGAGACCAATTTAAGGCAC**T**CGTGT**G**TAAACGGATAATGGACATGGTGAGCAACCAGCAC
6d7 GATGCTTTATTCTAGAGACCAATTTAAGGCAC**T**CGTGT**G**TAAACGGATAATGGACATGGTGAGCAACCAGCAC
3g8 GATGCTTTATTCTAGAGACCAATTTAAGGCAC**T**CGTGT**G**TAAACGGATAATGGACATGGTGAGCAACCAGCAC
3a9 GATGCTTTATTCTAGAGACCAATTTAAGGCA---**TGGTGAGTAACC**-----GAGCAACCAGCAC

FIG. 11C