

USE OF NW_003613638.1 IN CHO CELL GENOME FOR STABLE EXPRESSION OF PROTEIN

TECHNICAL FIELD

[0001] The present invention relates to the field of genetic technology, and
5 specifically relates to a recombinant gene of CHO cells for stable expression of protein.

BACKGROUND

[0002] Chinese Hamster Ovary (CHO) cells, as the main cell line in the field of
bio-pharmaceuticals, has developed many different types of CHO cell lines, including
10 cell lines that can be used to expand the number of gene copies. However, the increase
of transgenic copy number is not positively correlated with the enhancement of target
protein yield. Moreover, even with increased protein expression, the expression levels
of most CHO cells are unstable. The current methods commonly used for constructing
stable transfected cells are time and labor consuming, mainly due to that a large
15 number of monoclonal screening processes need to be repeated. Therefore, it is
widely expected currently in the field of cell line construction to develop a method
capable of obtaining highly and stably expressed cells within a short time, meanwhile
the constructed recombinant cell lines should be ensured to have the same qualities as
those prepared from traditional processes to ensure the approval of regulators.

20 [0003] A traditional process for constructing foreign protein-expressing cell lines is
to obtain cell lines highly expressing foreign proteins by randomly integrating foreign
genes onto a cell genome and then conducting a series of screening on the highly
expressed monoclonal cells. Due to the diversity of site effect differences, the
expression levels of recombinant cells produced by random integration are different,
25 so it will take a long time and many steps to select highly expressed monoclonal cells

in the later stage. Monoclonal cells obtained through random integration cannot guarantee the stable expression of polypeptides/proteins during cell passage, and repeated monoclonal screenings are required for each recombinant cell construction, thus increasing the costs on research and development of biopharmaceuticals.

- 5 **[0004]** Site effect hinders the efficiency of traditional random integration in the construction of recombinant cell lines, and repeated highly expressing monoclonal screenings are time and labor consuming and expensive. How to overcome the site effect and obtain stably expressed monoclonal cells rapidly and efficiently by using a site-specific integration technology has been discussed in academia for years, but no
10 breakthrough has been made.

SUMMARY

- [0005]** Against the above problems in the prior art, applicant of the invention provides a use of NW_003613638.1 in a CHO cell genome for stable expression of protein. According to the present invention, different protein genes are introduced to a
15 fixed position in the CHO cell genome for stable expression; in addition, in the process of realizing the site-specific integration, there is no need for multiple monoclonal screenings to obtain higher-expressing cell lines, saving a lot of time.

[0006] The technical solution of the invention is as below:

- A use for stable expression of protein at a site within a CHO cell genome, the site for
20 stable expression of protein within the CHO cell genome is located at the 1969647th base of CHO cell gene NW_003613638.1;

5' NNNNNNNNNNNNNNNNNNNNNNGG3' within a range of 1969591-1969721 near the site, which can be identified by CRISPR/Cas9 technology, is a target sequence.

- 25 **[0007]** The protein is a protein with a molecular weight of less than 160 KDa.

- [0008] The protein is one of a polypeptide, a functional protein, an antibody, and a fusion protein.
- [0009] The target sequence is bases within a range of 1969633-1969655 near the 1969647th base of the CHO cell gene NW_003613638.1.
- 5 [0010] Further, the target sequence is 5'-CCATGAGGAAACTTTTTGAGGG-3'.
- [0011] The target sequence is 5'-ATTTATATTCAGTATACATTTGG-3'.
- [0012] The target sequence is 5'-TTTATATTCAGTATACATTTGGG-3'.
- [0013] The target sequence is 5'-GTGTATCTTCGAACCAAACTGG-3'.
- [0014] The target sequence is 5'-TGTATCTTCGAACCAAACTGGG-3'.
- 10 [0015] The target sequence is 5'-ACTGGGTCACAAAATGTTTTAGG-3'.
- [0016] The target sequence is 5'-TTTACTCTTTCAACATTCTGAGG-3'.
- [0017] The target sequence is 5'-AACATTCTGAGGACTGCTCAAGG-3'.
- [0018] The target sequence is 5'-AACATTTTGTGACCCAGTTTTGG-3'.
- [0019] A targeting vector, the targeting vector comprises the target sequence of 5'
- 15 NNNNNNNNNNNNNNNNNNNNNNGG3' which can be used for site editing and can be identified by CRISPR/Cas9 technology.
- [0020] The targeting vector is prepared as below: the PSK-u6-gRNA plasmid is digested with BbsI and ligated to a target sequence DNA fragment containing BbsI cohesive ends to obtain a targeting vector.
- 20 [0021] A recombinant donor vector for expressing foreign genes in CHO cells, wherein the recombinant donor vector comprises a homologous arm within a range of 500 bp-800 bp upstream and downstream of the 1969647th base of the site NW_003613638.1.
- [0022] The recombinant donor vector is prepared by inserting protein genes into a
- 25 region between 5' arm and 3' arm of a plasmid, so that a nucleotide sequence is

located downstream of a promoter and is regulated by the promoter, to get a recombinant CHO cell-expressed plasmid.

[0023] The promoter is one of CMV (Human cytomegalovirus-derived strong mammalian expression promoter), EF-1a (Human elongation factor 1 α -derived strong mammalian expression promoter), SV40 (Simian vacuolating virus 40-derived mammalian expression promoter), PGK1 (Phosphoglycerate kinase gene-derived mammalian promoter), UBC (Human ubiquitin C gene-derived mammalian promoter), human beta actin (β -actin gene-derived mammalian promoter), CAG (strong hybrid mammalian promoter).

10 [0024] A CHO recombinant cell line for stably expressing foreign proteins, the above described vector or plasmid is used in the construction process of the recombinant cell line.

[0025] A method of expressing proteins with CHO cell genes comprises the following steps:

15 (1) CHO cells are co-transformed with Cas9, the targeting vector and the recombinant donor vector, to get recombinant CHO cells;

(2) the recombinant CHO cells are cultured on a plate, the supernatant is collected to detect the expression level, and the adherent recombinant CHO cells are suspension domesticated;

20 (3) The suspension domesticated recombinant CHO cells are cultured in a shaking flask, and the expression level of proteins is detected.

[0026] The present invention further provides a method of selecting a stable expression site in CHO cell genome:

1) Constructing lentiviruses with fluorescent tags and deducing their titers. The
25 igk-luc genes are integrated into multiple cloning sites of the

pLVX-CMV-MCS-T2A-Zsreen vector, and then transfected into the HEK-293T cell together with the two plasmids pSPAX2 and pMD2G. The cell supernatants are harvested twice at 48 hour and 72 hour, and collected for ultracentrifugation of the supernatant to get the lentivirus.

5 2) CHO cells are coated on a 6-well plate and cultured overnight. On the next day, the lentivirus is diluted and CHO cells are infected at low MOI ($MOI < 1$) (virus particles corresponding to each cell). After infection for 96 hours, the CHO cells are sorted by a flow cytometer, and the cells with the brightest fluorescence intensity are directly inoculated into a 96-well plate. One week later, when the cells have grown into
10 monoclonal colonies, they are observed under a fluorescence microscope, and the brightest cell colony with normal morphology and growth is selected and transferred into a 24-well plate for expanding the cultivation. When the cells are cultured to a convergence degree near 90%, they are transferred into a 6-well plate for cultivation, and finally expanded into a culture dish of 10 cm for cultivation. A part of cells are
15 cryopreserved and the remaining cells are expanded for cultivation continually.

3) All gene integration sites of lentivirus in CHO cells are searched by a genome walking technology Lenti-X Integration Site Analysis Kit (Clontech: 631263).

[0027] Several cell lines with the brightest fluorescence intensity and normal morphology and growth rate are used as the materials, and genomic DNA is digested
20 with three restriction endonucleases ADraI, SspI, and HpaI overnight. Wherein, a 100 μ L reaction system is formulated with 2.5 μ g of genomic DNA and 80 U of restriction endonucleases, and digested at 37°C overnight (16-18 hours).

[0028] The digested products are purified and recycled with a DNA Purification Kit. The digested genomic DNA 4.8 μ L, together with 1.9 μ L Genome Walker Adaptor (25
25 μ M) and 0.5 μ L T4 ligase, are formulated into a system of 8 μ L for ligation at 16°C

overnight. The ligation system is heated at 70°C for 5 minutes to inactivate the ligase. 32 µL TE buffer is added into each system to constitute a corresponding library.

[0029] The library is subjected to 2 rounds of nested PCR to amplify the LTR region and adjacent genomic regions. The related operating steps of PCR reaction can be conducted following the instruction of Lenti-X Integration Site Analysis Kit (Clontech: 631263).

[0030] Finally, the PCR products are subjected to electrophoresis. The main strips are cut and recovered, and then sent for sequencing. After obtaining all lentivirus integration information for each cell line, the related information of CHO cell lines only containing single copy lentivirus integration is selected, and their sequence information is compared with CHO-K1 genomic information on BLAST to find highly expressed integration sites.

[0031] The present invention has the following beneficial effects:

[0032] In the present invention, a site-specific integration method is employed to site-specifically integrating the target genes into a stable expression region, which can well overcome the problem of integration site uncertainty caused by random integration, effectively avoid multiple rounds of highly expressing monoclonal screening, effectively decreasing the research time for constructing a stably expressing cell line by bio-pharmaceuticals, and reducing the cost.

[0033] According to the present invention, protein genes are introduced to a fixed position in the CHO cell genome for stable expression.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] Fig. 1 is a schematic diagram of the invention;

[0035] Fig. 2 is a diagram showing the identification results of CHO cell gene with NGGH 75KDa gene;

[0036] Fig. 3 is the sequencing of OoPCR_fwd and OoPCR_rev on the CHO cell gene with NGGH 75KDa gene;

[0037] Fig. 4 shows the expression of HSA in cells at different passages;

[0038] Fig. 5 shows the expression of NGGH in cells at different passages;

5 [0039] Fig. 6 shows the mass profile of antibody proteins secreted from each recombinant CHO cell every day.

DESCRIPTION OF THE EMBODIMENTS

[0040] The present invention will be specifically described accompanying with the attached figures and embodiments.

10 [0041] Fig. 1 shows a schematic diagram of the donor plasmid to be integrated to the site and a simulated schematic diagram showing how to site-specifically integrate to the site through homologous recombination. Wherein, GOI is our target gene, which is site-specifically integrated to the target position through two homologous recombination arms, 5'arm and 3'arm, at a puromycin screening pressure of 4 μ g/mL.

15 In addition, the upstream sequence of 5'arm is a negative screening label, which can be used to remove monoclonal cells generated from random integration, thus ensuring that recombinant CHO cell lines generated from site-specific integration can be finally obtained.

Embodiment 1

20 Selection of high expression sites;

[0042] Zsgreen1 gene is integrated at NW_003613638.1, the 1969647th base. This fluorescent cell has been subcultured for no less than 50 generations. The expression level of fluorescence is detected with a flow cytometer. The expression level of green fluorescent protein is still good in the 50th generation of fluorescent cells, and the

fluorescent signals can be retained stably during cell passages.

[0043] Additionally, this fluorescent cell is suspension domesticated, and the expression level of the suspension domesticated fluorescent protein is detected with a flow cytometer again. The detection results show that in recombinant CHO cells that have been suspended for 50 generations, more than 95% of the cells remain the expression of green fluorescent protein after suspension. So it can be considered that this site is extremely stable and will not lose the site-specifically integrated fluorescent protein genes due to cell passages.

Embodiment 2

10 Selection of specific targets;

[0044] On the principle of proximity, the following sequence:

5'TCTAGATCCTGATACTGATGTCCCCCTATTTTAAAATGTCCCTCAAAAA
GTTTTCCTCATGGCCTGAAGACAAAAATCCTGGTATTTTAACTGACCTAATT
TTTCAAGTGTCCAAGTTTCATAGGTTT 3'

15 is input to the CRISPRater system to predict and sort out target sequences with low off-target efficiency. Wherein, the parameters are set as below: 1) The maximum number of mismatched bases at the beginning 15 bp after NGG is 0; 2) The number of mismatched bases at 21 bp after NGG is 2.

[0045] According to the above operation, the following sequence with a score of

20 0.73 is selected as the target sequence according to its scores:

5'-CCATGAGGAAAACCTTTTGAGGG-3';

[0046] According to the CRISPRater system, LOW efficacy (score<0.56); MEDIUM efficacy (0.56<=score<=0.74); HIGH efficacy (score>0.74).

[0047] According to the CRISPRater evaluation system, all the target sequences

25 within the range of 1969591-1969721 near the site NW_006880285.1 have obtained

scores of greater than 0.56, which are in the range of moderate efficiency or high efficiency, and can be used as the 5' NNNNNNNNNNNNNNNNNNNNNNNNGG3' target sequence identified by the CRISPR/Cas9 technology.

Embodiment 3 Selection of promoters

5 **[0048]** The above CMV (human cytomegalovirus-derived strong mammalian expression promoter) at the position of promoter is replaced with different promoters, including common promoters such as EF-1a (human elongation factor 1 α -derived strong mammalian expression promoter), SV40 (Simian vacuolating virus 40-derived mammalian expression promoter), PGK1 (phosphoglycerate kinase gene-derived
10 mammalian promoter), UBC (human ubiquitin C gene-derived mammalian promoter), human beta actin (β -actin gene-derived mammalian promoter) or CAG (strong hybrid mammalian promoter). Upon testing, all the above promoters can regulate the gene sequences of the downstream red fluorescent proteins and express the corresponding red fluorescent proteins.

15 Embodiment 4

[0049] The human serum albumin genes (HSA, 68KDa) are site-specifically integrated at specific sites: To construct CRISPR/Cas9-mediated homologous recombination at a later stage, sgRNA and Donor Plasmid need to be constructed, for which the method is as below:

20 1. Construction of SgRNA, the following sequences are firstly synthesized:

sgRNA-1fwd 5' TTTGCCATGAGGAAAACCTTTTGTAGT 3'

sgRNA-1rev 5' TAAAACTCAAAAAGTTTTCCTCATGG 3'

1) The PSK-u6-gRNA plasmid is digested with BbsI, and the digested vector is recovered;

25 2) The synthesized fragment is annealed to a double strands containing cohesive ends

sgRNA-1 fwd (100uM) 4uL

sgRNA-1 rev (100uM) 4uL

10×NEB buffer2 2uL

H2O 10uL

5 -----

20uL

water bath at 95°C for 5min, naturally cooling to room temperature in the water bath pot;

3) Fragment ligation and construction of the recombinant plasmid

10 Recovered linear vector 50 ng

Annealed fragment 1 uL

10×T4 ligase buffer 1 uL

T4 ligase 1 uL

H2O

15 -----

To 10 uL;

4) Ligation and transformation;

5) Selection of clones, PCR identification, the primers used for identification are:

M13-synthesized R primers, those clones which can be amplified into bands are positive clones.

20

2. Construction of Donor plasmid

The specific information of donor plasmid is as shown in Fig. 1: Except for GOI, the remaining parts are all synthesized; the sequence within 600bp upstream and downstream of the target is the left and right homologous arm sequence information

25 of the donor plasmid, and GOI is a process in which HSA genes are integrated onto

Ligation at room temperature for more than an hour, or ligation at 4°C overnight

the donor plasmid through the existing C115 kit from Vazyme Biotech Co.

3. Three plasmids, Cas9 (Technical University of Denmark, donated by Dr. Helene F Kildegaard), SgRNA and Donor plasmid are co-transfected with CHO cells cultured at 37°C and 5% CO₂, wherein the molar ratio of the three plasmids is 1:1:1, the
5 transfection reagent is Lipofectamine 3000 (Thermo Fisher Scientific), and the specific transfection process with reference to the instruction. After then, 4 µg/ml puromycin is added into the cells for screening, the process lasts for 10 days totally; monoclonal cells are sorted by a MoFloXDP FACS machine (Beckman Coulter), and cells without any fluorescence are selected and inoculated into a 96-well plate.
- 10 4. After growing for 2 weeks, a part of cells are identified by PCR, such as 5'junction PCR, 3'Junction PCR and out-out PCR. Positive cells are reserved.

Embodiment 5

- [0050] The glucagon-like peptide-1-human serum albumin fusion protein genes (NGGH, 75KDa) are site-specifically integrated at specific sites: To construct
15 CRISPR/Cas9-mediated homologous recombination at a later stage, sgRNA and Donor Plasmid need to be constructed, for which the method is as below:

1. The construction of sgRNA is the same as in Embodiment 4.
2. Construction of Donor plasmid.

The specific information of donor plasmid is as shown in Fig. 1: Except for GOI, the
20 remaining parts are all synthesized; the sequence within 600bp upstream and downstream of the target is the left and right homologous arm sequence information of the donor plasmid, and GOI is a process in which NGGH genes are integrated onto the donor plasmid through the existing C115 kit from Vazyme Biotech Co.

3. Three plasmids, Cas9, SgRNA and Donor plasmid are co-transfected with CHO
25 cells cultured at 37°C and 5% CO₂, wherein the molar ratio of the three plasmids is

1:1:1, the transfection reagent is Lipofectamine 3000 (Thermo Fisher Scientific), and the specific transfection process with reference to the instruction. After then, 4 µg/ml puromycin is added into the cells for screening, the process lasts for 10 days totally. Monoclonal cells are sorted by a MoFloXDP FACS machine (Beckman Coulter), and
5 cells without any fluorescence are selected and inoculated into a 96-well plate.

4. After growing for 2 weeks, a part of cells are identified by PCR, such as 5'junction PCR, 3'Junction PCR and out-out PCR. Positive cells are reserved.

[0051] Fig. 2 is a diagram showing the identification results of CHO gene with NGGH 75KDa gene; wherein lanes 1-3 show 5'junction PCR results of three
10 monoclonal cells, and lanes 4-5 show 3'junction PCR results; there are obvious strips in all the lanes, confirming that there is gene knock-in.

[0052] Fig. 3 shows that OoPCR_fwd and OoPCR_rev are utilized for sequencing to determine the sequence accuracy at the junction (for 5'junction and 3'junction, the junction between the 5' upstream and the genome), wherein the sequencing results
15 confirm that GOI is accurately inserted into the target area.

Embodiment 6

[0053] The antibody genes (Avastin, 160KDa) are site-specifically integrated at specific sites: To construct CRISPR/Cas9-mediated homologous recombination at a later stage, sgRNA and Donor Plasmid need to be constructed, for which the method
20 is as below:

1. The construction of sgRNA is the same as in Embodiment 4.
2. Construction of Donor plasmid.

The specific information of donor plasmid is as shown in Fig. 1: Except for GOI, the remaining parts are all synthesized; the sequence within 600bp upstream and
25 downstream of the target is the left and right homologous arm sequence information

of the donor plasmid, and GOI is a process in which Avastin genes are integrated onto the donor plasmid through the existing C115 kit from Vazyme Biotech Co.

3. Three plasmids, Cas9 (Technical University of Denmark, donated by Dr. Helene F Kildegaard), SgRNA and Donor plasmid are co-transfected with CHO cells cultured at 37°C and 5% CO₂, wherein the molar ratio of the three plasmids is 1:1:1, the transfection reagent is Lipofectamine 3000 (Thermo Fisher Scientific), and the specific transfection process with reference to the instruction. After then, 4 µg/mL puromycin is added into the cells for screening, the process lasts for 10 days totally. Monoclonal cells are sorted by a MoFloXDP FACS machine (Beckman Coulter), and cells without any fluorescence are selected and inoculated into a 96-well plate.
4. After growing for 2 weeks, a part of cells are identified by PCR, such as 5'junction PCR, 3'Junction PCR and out-out PCR. Positive cells are reserved.

Test embodiment:

- [0054] The three cell lines obtained in Embodiments 4-6 are detected through ELISA to observe whether there is a target protein expression and whether it is a stable and long-term expression.

- [0055] Detection method: The three detections are all conducted by the ELISA method, and all the sorted positive cells are cultured in a 6-well plate to determine whether there is a long-term stable expression of the target protein. The kits used for the experiments are Human Albumin ELISA Kit (RK00157) and Human IgG (Total) ELISA Kit (RK00393) from Abclonal Co.

- [0056] Figs. 4 and 5 show the expression of HSA and NGGH at different passages respectively, wherein the vertical coordinate indicates the mass of proteins secreted from each cell every day.

- [0057] It can be seen from the figures that both NGGH and HAS can stably express

the corresponding genes in the plate within 50 passages, and the expression levels of the target protein by the sorted three NGGH site-specific integrated cell lines and two HAS site-specific integrated cell lines are close to each other.

[0058] Fig. 6 shows the mass of antibody proteins secreted from each recombinant CHO cell every day. Obviously, cells at different passages can stably express the secreted corresponding proteins, showing good stabilities and being consistent with the previous results of fluorescent cell. The results show that this site can be subjected to CRISPR/Cas9-mediated site-specific integration, and can stably express the corresponding proteins.

[0059] After 5'-CCATGAGGAAACTTTTGGAGGG-3' sequence is selected to subject to the above tests, good results have been obtained. Therefore, the target sequences described in claims 5-12 all can be used to construct site-specifically integrated and stably expressing cell lines, all of which can stably express target proteins.

WHAT IS CLAIMED IS:

1. A use for stable expression of protein at a site within a CHO cell genome, wherein the site for stable expression of protein within the CHO cell genome is located at the 1969647th base of CHO cell gene NW_003613638.1;
- 5 5' NNNNNNNNNNNNNNNNNNNNNNGG3' within a base range of 1969591-1969721 near the site, which can be identified by CRISPR/Cas9 technology, is a target sequence.
2. The use according to claim 1, wherein the protein is a protein with a molecular weight of less than 160 KDa.
- 10 3. The use according to claim 1, wherein the protein is one of a polypeptide, a functional protein, an antibody, and a fusion protein.
4. The use according to claim 1, wherein the target sequence is bases within a range of 1969633-1969655 near the 1969647th base of the CHO cell gene NW_003613638.1; the target sequence is 5'-CCATGAGGAAAACCTTTTGAGGG-3'.
- 15 5. The use according to claim 1, wherein the target sequence is 5'-ATTTATATTCAGTATACATTTGG-3'.
6. The use according to claim 1, wherein the target sequence is 5'-TTTATATTCAGTATACATTTGGG-3'.
7. The use according to claim 1, wherein the target sequence is 5'-GTGTATCTTCGAACCAAACTGG-3'.
- 20 8. The use according to claim 1, wherein the target sequence is 5'-TGTATCTTCGAACCAAACTGGG-3'.
9. The use according to claim 1, wherein the target sequence is 5'-ACTGGGTCACAAAATGTTTTAGG-3'.
- 25 10. The use according to claim 1, wherein the target sequence is

5'-TTTACTCTTTCAACATTCTGAGG-3'.

11. The use according to claim 1, wherein the target sequence is 5'-AACATTCTGAGGACTGCTCAAGG-3'.

12. The use according to claim 1, wherein the target sequence is 5'-AACATTTTGTGACCCAGTTTTGG-3'.

13. A targeting vector, wherein the targeting vector comprises the target sequence of 5' NNNNNNNNNNNNNNNNNNNNNNGG3' according to claim 1 which can be used for site editing and can be identified by CRISPR/Cas9 technology.

14. A recombinant donor vector for expressing foreign genes in CHO cells, wherein the recombinant donor vector comprises a homologous arm within a range of 500 bp-800 bp upstream and downstream of the 1969647th base of the site NW_003613638.1 according to claim 1.

15. The recombinant donor vector according to claim 14, wherein the vector is prepared by inserting protein genes into a region between 5' arm and 3' arm of a plasmid, so that a nucleotide sequence is located downstream of a promoter and is regulated by the promoter, to get a recombinant CHO cell-expressed plasmid.

16. A CHO recombinant cell line for stably expressing foreign proteins, wherein the vector or plasmid according to claim 13, 14 or 15 is used in the construction of the recombinant cell line.

20