CHROMOSOMAL LANDING PADS AND RELATED USES

REFERENCE TO PRIORITY DOCUMENT

[0001] This application claims the benefit of priority under 35 U.S.C. §119(e) of U.S. Provisional Patent Application Serial No. 61/516,612, filed April 5, 2011. Priority of the aforementioned filing date is hereby claimed and the disclosure of the provisional patent application is hereby incorporated by reference in its entirety.

INCORPORATION OF SEQUENCE LISTING

[0002] This application contains an electronic equivalent paper copy of the sequence listing submitted herewith electronically via EFS web and a computer-readable form of the sequence listing submitted herewith electronically via EFS web and contains the file named "37651505001WOSEQUENCELISTING.txt," which is 213476 bytes in size and which was created on April 5, 2012, are hereby incorporated by reference in their entirety.

BACKGROUND

[0003] Integration of heterologous polynucleotides into the genomes of mammalian cells is routinely practiced for therapeutic purposes (e.g., gene therapy) and in the production of useful proteins or polypeptides in vitro. Insertion at random locations in the genome by non-homologous recombination requires several rounds of selection and clonal expansion to produce an acceptable expression system. The approach also needs to be repeated every time an expression system for a new gene is sought. Due to the random nature of the integration event, some of the locations where recombinant genes are inserted are incapable of supporting transcriptional events at all. This is because expression levels are greatly influenced by the effects of the local genetic environment at the gene locus (position effects). In addition, expression from many chromosomal sites is decreased over time. In some cases, this instability is due to DNA methylation of the transgene. As a result, wide variations in the expression level of integrated genes can occur, depending on the site of integration. In addition, random integration of exogenous DNA into the genome can in some instances disrupt important cellular genes, resulting in an altered phenotype.

[0004] Other than random insertion, recombinase-mediated integration has been described for insertion of transgenes at defined sites in the genome. However, achieving stable, high-efficient expressions of integrated transgenes is still cumbersome and requires large numbers of screened clones in order to select desirable integrated cells.

[0005] There is a need in the art for means for achieving a stable integration and/or high level of gene expression of heterologous polynucleotide in mammalian cells. The present disclosure addresses this and other needs.

SUMMARY

[0006] In one aspect, provided are methods for stable integration and expression of a heterologous polynucleotide in a host cell. The methods involve inserting the heterologous polynucleotide into the genome of the host cell at a native chromosomal site located within or adjacent to a gene selected from the group consisting of ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene and Nephrocystin-1/Mal gene. In some methods, insertion of the heterologous polynucleotide into the host genome is mediated by homologous recombination or by a hybrid recombinase. In some methods, the host cell is a mammalian cell, e.g., a Chinese hamster ovary (CHO) cell. In some of these methods, the native chromosomal insertion site is at or close to positions 130-131 of SEQ ID NO:1 for the Ank2 gene, positions 629-630 of SEQ ID NO:2 for the Cpsf4 gene, positions 272-273 of SEQ ID NO:3 for the C-Mos gene, or positions 239-240 of SEQ ID NO:4 for the Nephrocystin-1/Mal gene. In some methods, the heterologous polynucleotide to be integrated into the host genome can encode a polypeptide, e.g., a therapeutic protein or an industrial protein.

[0007] In a related aspect, provided are recombinant or engineered polynucleotides for stably integrating a heterologous polynucleotide sequence into the genome of a mammalian cell. The recombinant polynucleotides typically contain a first homology arm, the heterologous polynucleotide sequence, and a second homology arm. The first and second homology arms are substantially identical to the 5'- and 3'- sequences, respectively, that flank a native chromosomal insertion site located within or adjacent to a gene selected from the group consisting of ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene and Nephrocystin-1/Mal gene. Typically, the native chromosomal insertion site is capable of supporting stable integration of a foreign gene. In some methods, the heterologous polynucleotide sequence encodes a polypeptide, e.g., a therapeutic protein or an industrial protein. In some other methods, the heterologous polynucleotide sequence comprises a site-specific recombination sequence (chromosomal landing pad). For example, the site-specific recombination sequence can be a recognition sequence recognized by a phage integrase, such as the attP site or the attB site recognized by phiC-31 phage integrase.

In some methods, the host mammalian cell is a Chinese hamster ovary (CHO) cell. In these methods, the native chromosomal insertion site can be located at or close to positions 130-131 of SEQ ID NO:1 for the Ank2 gene, positions 629-630 of SEQ ID NO:2 for the Cpsf4 gene, positions 272-273 of SEQ ID NO:3 for the C-Mos gene, or positions 239-240 of SEQ ID NO:4 for the Nephrocystin-1/Mal gene. In related embodiments, vectors containing the recombinant or engineered polynucleotides are also provided in the invention.

[0008] In another aspect, provided are engineered mammalian cells. The cells harbor a heterologous polynucleotide that is stably integrated into its genome at one or more native chromosomal insertion sites located within or adjacent to a gene selected from the group consisting of ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene and Nephrocystin-1/Mal gene. Typically, the chosen native chromosomal insertion site supports stable integration of a foreign gene. In some of the methods, the heterologous polynucleotide encodes a polypeptide, e.g., a therapeutic protein or an industrial protein. In some other methods, the heterologous polynucleotide contains a site-specific recombination sequence (chromosomal landing pad). For example, the site-specific recombination sequence can be a recognition sequence recognized by a phage integrase, such as the attP site or the attB site recognized by phiC-31 phage integrase. Some preferred embodiments are directed to recombinant or engineered Chinese hamster ovary (CHO) cells. In these embodiments, the heterologous polynucleotide can be preferably integrated at or close to positions 130-131 of SEQ ID NO:1 for the Ank2 gene, positions 629-630 of SEQ ID NO:2 for the Cpsf4 gene, positions 272-273 of SEQ ID NO:3 for the C-Mos gene, or positions 239-240 of SEQ ID NO:4 for the Nephrocystin-1/Mal gene.

[0009] In still another related aspect, provided are methods for stably integrating a heterologous polynucleotide into the genome of a mammalian cell. These methods entail (a) inserting a site-specific recombination sequence into the genome of the cell, wherein the insertion is at a native chromosomal insertion site located within or adjacent to a gene selected from the group consisting of ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene and Nephrocystin-1/Mal gene; and (b) integrating by homologous recombination the heterologous polynucleotide into the genome of the cell at the inserted site-specific recombination sequence. The native chromosomal insertion site chosen for the methods typically supports stable integration of a foreign gene. In some methods, the site-specific recombination sequence is a first recognition sequence recognized by a phage integrase, e.g., the attP site or the attB site of phiC-31 phage integrase. In these

methods, the heterologous polynucleotide is usually attached to a second recognition sequence of the phage integrase which is cognate to the first recognition sequence, e.g., the attB site or the attP site recognized by the phage integrase. In some methods, the employed mammalian host cell is a Chinese hamster ovary (CHO) cell. In these methods, the site-specific recombination sequence can be preferably inserted into the genome at or close to positions 130-131 of SEQ ID NO:1 for the Ank2 gene, positions 629-630 of SEQ ID NO:2 for the Cpsf4 gene, positions 272-273 of SEQ ID NO:3 for the C-Mos gene, or positions 239-240 of SEQ ID NO:4 for the Nephrocystin-1/Mal gene. In some methods, the heterologous polynucleotide contains a target polypeptide-encoding sequence that is operably linked to a promoter sequence. Typically, integration of the heterologous polynucleotide into the host genome occurs in the presence of the phage integrase. In some of these methods, the phage integrase can be expressed from a vector introduced into the cell.

[0010] A further understanding of the nature and advantages of the present invention may be realized by reference to the remaining portions of the specification and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] Figure 1 is a diagram showing the structure of the plasmid used for random integration into CHO genome for identifying native chromosomal insertion sites that support strong transcription activities. The plasmid contains a sv40 promoter driven EGFP expression cassette. The BamHI site is between EGFP and Hygromycin resistence gene (for linearization before stable integration).

[0012] Figure 2 illustrates the plasmid used for introducing an attP site into identified native chromosomal insertion sites in CHO genome. The attP site is in the 5' of Neo gene. The left-homology arm is cloned 5'of attP and Neo gene; the right-homology arm is cloned at 3' of Neo gene. The HSV-TK gene is used for negative selection. The sequences shown are the double strand sequences of the attP and attB sites recognized by phi-C31 phage integrase: attP (SEQ ID NO:5 and SEQ ID NO:6), attB (SEQ ID NO:7 and SEQ ID NO:8).

[0013] Figure 3 shows a native chromosomal insertion site and flanking sequences in the ankyrin 2 (Ank2) gene of CHO cell genome (SEQ ID NO:1).

[0014] Figure 4 shows a native chromosomal insertion site and flanking sequences in the cleavage and polyadenylation specific factor 4 (Cpsf4) gene of CHO cell genome (SEQ ID NO:2).

[0015] Figure 5 shows a native chromosomal insertion site and flanking sequences in the C-Mos gene of CHO cell genome (SEQ ID NO: 3).

[0016] Figure 6 shows a native chromosomal insertion site and flanking sequences in the Nephrocystin-1/Mal gene of CHO cell genome (SEQ ID NO: 4).

DETAILED DESCRIPTION

I. Overview

[0017] Disclosed herein are native chromosomal sites in mammalian cells that are capable of strong transcriptional activity of a recombinant gene and their use as "landing pads" for site-specific integration of recombinant constructs. Specifically, chromosomal locations in several genes in mammalian genomes (e.g., Chinese Hamster Ovary (CHO) genome) were identified that promote strong expression of integrated foreign genes. As described below, identification of these native chromosomal insertion sites involved random integration into the genome of plasmids containing genes for selection (e.g., hygromycin-resistance gene and gene encoding the Enhanced Green Fluorescent Protein, EGFP). Upon random integration, cells were selected for hygromycin-resistance and sorted for EGFP expression using Fluorescent Activated Cell Sorting (FACS) three weeks after initial transfection. Selected cells were allowed to recover and grown without selection for several more weeks. Cells were then FACS-sorted again. Cells with the highest EGFP levels were sorted into individual wells of 96-well plates. Clones were grown from single cells and cultured for several weeks. Cells were then retested for EGFP expression. Cells were further screened to identify those with growth rates that were comparable to or higher than the growth rate of the parental cell line. Sequences at the insertions sites in these genes were then analyzed. These studies resulted in the identification of several genes, ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene and Nephrocystin-1/Mal gene, which harbor native chromosomal sites that enable stable and strong transcription activities from a recombinant gene.

[0018] Also described herein is the indentification of native chromosomal sites as chromosomal landing pads for uniform integration of desired target polynucleotide sequences. To this end, the phage attachment site attP recognized by phage integrase is introduced into the native chromosomal sites through homologous recombination. With the site-specific recombination sequence (i.e., the attP site) inserted into the genome, recombinant genes can then be readily introduced into the cell using vectors containing the

cognate recombination sequence (i.e., attB attachment site) that is recognized by the phage integrase (e.g., the phiC-31 phage integrase) in the presence of the phage integrase. The phage integrase allows the recombination of the two cognate recombination sequences (i.e., attB and attP sites), such that the entire attB-containing vector can be integrated into a single attP site in the chromosome.

[0019] Provided herein are methods for stable integration and/or expression of a heterologous polynucleotide in a host cell. Host cells containing a heterologous polynucleotide stably integrated in or near one or more of the identified genes (i.e., Cpsf4, Ank2, C-Mos, and Nephrocystin-1/Mal genes) are also provided. Further provided are polynucleotides and related vectors which are useful for inserting a heterologous polynucleotide, e.g., a site-specific recombination sequence (chromosomal landing pad), into the genome of a mammalian cell, in particular into one or more of the native chromosomal insertion sites disclosed herein. Additionally provided are engineered mammalian cells which have a heterologous site-specific recombination sequence that is stably integrated into its genome at one or more of the native chromosomal insertion sites disclosed herein. Moreover, provided are methods for stable integration at one or more inserted chromosomal landing pads and expression in a mammalian cell of a heterologous polynucleotide that encodes a target polypeptide of interest. Cells thus generated for expressing the heterologous polynucleotide is also provided herein.

[0020] The particular methodology, protocols, and reagents described here can vary. Unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are within the skill of the art can be employed. Such techniques are explained fully in the literature. For example, exemplary methods are described in the following references, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (3rd ed., 2001); Brent et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (ringbou ed., 2003); Freshney, *Culture of Animal Cells: A Manual of Basic Technique*, Wiley-Liss, Inc. (4th ed., 2000); and Weissbach & Weissbach, *Methods for Plant Molecular Biology*, Academic Press, NY, Section VIII, pp. 42 1-463, 1988.

II. Definitions

[0021] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which this

disclosure pertains. The following references provide one of skill with a general definition of many of the terms used in this disclosure: *Academic Press Dictionary of Science and Technology*, Morris (Ed.), Academic Press (1st ed., 1992); *Oxford Dictionary of Biochemistry and Molecular Biology*, Smith et al. (Eds.), Oxford University Press (revised ed., 2000); *Encyclopaedic Dictionary of Chemistry*, Kumar (Ed.), Anmol Publications Pvt. Ltd. (2002); *Dictionary of Microbiology and Molecular Biology*, Singleton et al. (Eds.), John Wiley & Sons (3rd ed., 2002); *Dictionary of Chemistry*, Hunt (Ed.), Routledge (1st ed., 1999); *Dictionary of Pharmaceutical Medicine*, Nahler (Ed.), Springer-Verlag Telos (1994); *Dictionary of Organic Chemistry*, Kumar and Anandand (Eds.), Anmol Publications Pvt. Ltd. (2002); and *A Dictionary of Biology (Oxford Paperback Reference)*, Martin and Hine (Eds.), Oxford University Press (4th ed., 2000). Further clarifications of some of these terms as they apply specifically to this disclosure are provided herein.

[0022] As used herein, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells, reference to "a protein" includes one or more proteins and equivalents thereof known to those skilled in the art, and so forth.

[0023] The term "agent" includes any substance, molecule, element, compound, entity, or a combination thereof. It includes, but is not limited to, e.g., protein, polypeptide, small organic molecule, polysaccharide, polynucleotide, and the like. It can be a natural product, a synthetic compound, or a chemical compound, or a combination of two or more substances. Unless otherwise specified, the terms "agent", "substance", and "compound" are used interchangeably herein.

[0024] The term "chromosomal landing pad" (or simply "landing pad") refers to a site-specific recognition sequence or a site-specific recombination site (e.g., an attP site) that is stably integrated into the genome of a host cell (e.g., a mammalian cell such as CHO cell). In particular, the site-specific recognition sequence or recombination site is inserted into the host genome at one or more native chromosome insertion sites present in several specific genes disclosed herein, i.e., ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene and Nephrocystin-1/Mal gene. Presence in the host genome of the heterologous site-specific recombination sequence allows a recombinase (e.g., phiC-31 integrase) to mediate site-specific insertion of a heterologous polynucleotide or a transgene into the host genome. Typically, in order to integrate into the landing pad, the heterologous polynucleotide or transgene is attached to a cognate recognition sequence or

recombination site (e.g., an attB site if the inserted site-specific recombination site is an attP site) that is also recognized by the recombinase.

[0025] The phrase "polynucleotide of interest" (or "gene of interest" or "target gene") is intended to include a cistron, an open reading frame (ORF), or a polynucleotide sequence which codes for a polypeptide or protein product ("polypeptide of interest" or "target polypeptide"). For stable integration and expression in an engineered host cell bearing a chromosomal landing pad described herein, a polynucleotide of interest can additionally contain appropriate transcription regulatory elements (e.g., promoter sequences) operably linked to the coding sequence and also a cognate site-specific recombination sequence (e.g., attB or attP site). Various target polypeptides can be encoded by and expressed from a polynucleotide of interest, e.g., therapeutic proteins, nutritional proteins and industrial useful proteins.

[0026] The term "endogenous" as used herein refers to a nucleic acid or polypeptide that is normally found in the wild-type host, while the term "exogenous" refers to a nucleic acid or polypeptide that is not normally found in the wild-type host.

[0027] A "host cell" refers to a living cell into which a heterologous polynucleotide sequence is to be or has been introduced. The living cell includes both a cultured cell and a cell within a living organism. Means for introducing the heterologous polynucleotide sequence into the cell are well known, e.g., transfection, electroporation, calcium phosphate precipitation, microinjection, transformation, viral infection, and/or the like. Often, the heterologous polynucleotide sequence to be introduced into the cell is a replicable expression vector or cloning vector. In some embodiments, host cells can be engineered to incorporate a desired gene on its chromosome or in its genome. Many host cells (e.g., CHO cells) that can serve as hosts are known in the art. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (3rd ed., 2001); and Brent et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (ringbou ed., 2003). In some preferred embodiments, the host cell is a mammalian cell.

[0028] The term "nucleotide sequence," "nucleic acid sequence," "nucleic acid," or "polynucleotide sequence," refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogs of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally-occurring nucleotides. Nucleic acid sequences can be, e.g., prokaryotic sequences, eukaryotic mRNA sequences, cDNA sequences from eukaryotic mRNA, genomic DNA

sequences from eukaryotic DNA (e.g., mammalian DNA), and synthetic DNA or RNA sequences, but are not limited thereto.

[0029] The term "operably linked" or "operably associated" refers to functional linkage between genetic elements that are joined in a manner that enables them to carry out their normal functions. For example, a gene is operably linked to a promoter when its transcription is under the control of the promoter and the transcript produced is correctly translated into the protein normally encoded by the gene. Similarly, an enhancer element is operably associated with a gene of interest if it allows up-regulated transcription of the gene.

[0030] A "substantially identical" nucleic acid or amino acid sequence refers to a nucleic acid or amino acid sequence which comprises a sequence that has at least 75%, 80% or 90% sequence identity to a reference sequence as measured by one of the well known programs described herein (e.g., BLAST) using standard parameters. The sequence identity is preferably at least 95%, more preferably at least 98%, and most preferably at least 99%. In some embodiments, the subject sequence is of about the same length as compared to the reference sequence, i.e., consisting of about the same number of contiguous amino acid residues (for polypeptide sequences) or nucleotide residues (for polynucleotide sequences).

[0031] Sequence identity can be readily determined with various methods known in the art. For example, the BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)). Percentage of sequence identity is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

[0032] As used herein, unidirectional site-specific recombinases (or simply site-specific recombinases) refer to a group of recombinases from bacteria and unicellular yeasts. They encompass both tyrosine recombinases and the resolvase/invertase or serine recombinase

family (e.g., phage integrases such as integrases from phages phiC31, R4, and TP-901). Tyrosine recombinases include tyrosine integrases (e.g., integrases from λ, HK022, P22, HP1 and L5) and other tyrosine recombinases (e.g., Cre and Flp). Examples of serine recombinases include serine integrases (e.g., integrases from phiC-31, R4, TP901) and other serine recombinases (e.g., γδ, Tn3, phage Mu recombinase).

[0033] Preferably, site-specific recombinases can include integrases (especially phage integrases) that mediate unidirectional site-specific recombination between two different DNA recognition sequences, the phage attachment site, attP, and the bacterial attachment site, attB. Integrases of the tyrosine family, e.g., lambda integrase, utilize a catalytic tyrosine to mediate strand cleavage, tend to recognize longer attP sequences, and require other proteins encoded by the phage or the host bacteria. Phage integrases from the serine family (e.g., phiC-31 phage integrase) are larger, use a catalytic serine for strand cleavage, recognize shorter attP sequences, and do not require host cofactors. Because the attB and attP sites are different sequences, recombination will result in a stretch of nucleic acids (called attL or attR for left and right) that is neither an attB sequence or an attP sequence, and is functionally unrecognizable as a recombination site to the relevant integrase enzyme, thus removing the possibility that the enzyme will catalyze a second recombination reaction that would reverse the first. This will result in a unidirectional site-specific integration event.

[0034] Phi-C31 integrase refers to a phage integrase which is capable of catalyzing in mammalian cells genomic recombination with high efficiency and tight sequence specificity. Functional characterization of this enzyme is described in the art, e.g., Kuhstoss and Rao, J. Mol. Biol. 222, 897-908, 1991; Rausch and Lehmann, Nucleic Acids Research 19, 5187-5189, 1991; and Groth et al., Proc. Natl. Acad. Sci. USA 97, 5995-6000, 2000.

[0035] The native attB and attP recognition sites of phage integrases (e.g., phage phi-C31 integrase) are generally about 34 to 40 nucleotides in length. See, e.g., Fig. 2 herein and also Groth et al., Proc. Natl. Acad. Sci. USA 97:5995-6000, 2000. These sites are typically arranged as follows: attB comprises a first DNA sequence attB5', a core region, and a second DNA sequence attB3', in the relative order from 5' to 3' attB5'-core region-attB3'. AttP comprises a first DNA sequence attP5', a core region, and a second DNA sequence attP3', in the relative order from 5' to 3' attP5'-core region-attP3'. The core region of attP and attB of Phi-C31 has the sequence 5'-TTG-3'.

[0036] A transgenic animal or plant refers to a non-human animal or a plant having a transgene or transgenic element integrated in the genome of one or more cells of the animal

or the plant. The term encompasses animals or plants having all or nearly all cells containing a genetic modification (e.g., fully transgenic animals, particularly transgenic animals having a heritable transgene) as well as chimeric transgenic animals or plants, in which a subset of cells of the animal or plants are modified to contain the genomically integrated transgene. A transgenic plant or animal includes an individual animal or plant in all stages of development. For transgenic animals, farm animals (e.g., chickens, pigs, goats, sheep, cows, horses, rabbits and the like), rodents (such as mice), non-human primates (such as rhesus macaques) and domestic pets (e.g., cats and dogs) are considered herein. In some preferred embodiments, the animal is a mouse or a rat.

[0037] "Therapeutic genes" refer to polynucleotide sequences which encode molecules that provide some therapeutic benefit to the host, including proteins (e.g., secreted proteins, membrane-associated proteins (e.g., receptors), structural proteins, cytoplasmic proteins, and the like) functional RNAs (antisense, hammerhead ribozymes), and the like. Secreted proteins include those that may be found in a bodily fluid of a subject (e.g., in blood, lymph, saliva, gastrointestinal secretions, and the like). In some embodiments, the mammalian subject is a human subject and the introduced polynucleotide sequence encodes a human protein or other human gene product.

[0038] The term "vector" or "construct" refers to polynucleotide sequence elements arranged in a definite pattern of organization such that the expression of genes/gene products that are operably linked to these elements can be predictably controlled. Typically, they are transmissible polynucleotide sequences (e.g., plasmid or virus) into which a segment of foreign polynucleotide sequence can be spliced in order to introduce the foreign DNA into host cells to promote its replication and/or transcription.

[0039] A cloning vector is a polynucleotide sequence (typically a plasmid or phage) which is able to replicate autonomously in a host cell, and which is characterized by one or a small number of restriction endonuclease recognition sites. A foreign polynucleotide sequence fragment may be spliced into the vector at these sites in order to bring about the replication and cloning of the fragment. The vector may contain one or more markers suitable for use in the identification of transformed cells. For example, markers may provide tetracycline or ampicillin resistance.

[0040] An expression vector is similar to a cloning vector but is capable of inducing the expression of the polynucleotide sequence that has been cloned into it, after transformation into a host. The cloned polynucleotide sequence is usually placed under the control of (i.e.,

operably linked to) certain regulatory sequences such as promoters or enhancers. Promoter sequences may be constitutive, inducible or repressible.

III. Inserting heterologous polynucleotides at native chromosomal integration sites

[0041] Described herein are several specific genes that contain native chromosomal integration sites which support stable and efficient expression of an inserted heterologous polynucleotide (exogenous gene or transgene). These native chromosomal integration sites are suitable for stable integration and/or expression of a heterologous polynucleotide in a host cell. For example, transgenes or recombinant genes encoding useful polypeptides (e.g., therapeutic or industrial proteins) can be so integrated and expressed in host cells. Additionally, theses sites can be employed for inserting site-specific recombination sequences (chromosomal landing pads) into a host genome. Host cells bearing such inserted chromosomal landing pads can in turn be used for insertion and expression of heterologous polynucleotides.

[0042] A native chromosomal insertion or integration site refers to a chromosomal location or site into which a heterologous polynucleotide can be integrated, e.g., via random integration, and which may occur naturally in the genome of a cell. In other words, the site is not introduced into the genome, for example, by recombinant means. Unless otherwise noted, the term as used herein specifically refers to a position in the genome that supports stable integration of foreign genes and their efficient transcription, and that is located within or adjacent to one of several genes in the CHO genome including: ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene and Mal gene also described herein as the Nephrocystin-1/Mal gene. It also encompasses chromosomal locations in the orthologs of these genes or homologous regions (as determined by sequence alignment) in other mammalian species (e.g., mouse, rat and human) with similar functions or activities.

[0043] As detailed herein, one specific native chromosomal insertion site is described herein for each of the four genes identified in the CHO genome (the "exemplified positions"; see Figures 3-6). However, the native chromosomal insertion sites considered herein are not limited to these specific positions. So long as stable integration and/or efficient transcription of an integrated heterologous polynucleotide is supported, the exact location of the native chromosomal insertion site with respect to the exemplified sites is not essential. Rather, the native chromosomal site can be at any position that is within or adjacent to one of the four

genes. Whether a specific chromosomal location within or adjacent one of the four genes of interest supports stable integration and efficient transcription of an integrated foreign gene can be determined in accordance with standard procedures well known in the art or methods exemplified herein. In some preferred embodiments, the specific positions exemplified herein for CHO genome or corresponding positions (as determined by sequence alignment) in other mammalian genomes (e.g., mouse, rat or human genome) are employed as the native chromosomal insertion sites. In some other embodiments, the native chromosomal sites considered herein are preferably located close to one of the exemplified positions, e.g., within less than about 1 kb, 500 bp, 250 bp, 100 bp, 50 bp, 25 bp, 10 bp, or less than about 5 bp of one of the exemplified positions. In still some other embodiments, the employed native chromosomal site is located at about 1000, 2500, 5000 or more base pairs away from one of the exemplified positions.

[0044] A heterologous polynucleotide (e.g., a recombinant gene or a chromosomal landing pad) can be readily inserted into the native chromosomal integration sites described herein for stable integration and/or expression. The heterologous polynucleotide can be inserted into the native chromosomal integration sites of the host genome by various means, e.g., by homologous recombination or by using a hybrid recombinase that specifically targets sequences at the integration sites. For homologous recombination, homologous polynucleotide molecules line up and exchange a stretch of their sequences. A trans-gene can be introduced during this exchange if the trans-gene is flanked by homologous genomic sequences. For example, as described below, a chromosomal landing pad (an attP site-containing sequence) can be so inserted into the host genome at the native chromosomal integration sites.

[0045] Efficiency of homologous recombination in mammalian cells can be improved by introducing a break in the chromosomal region of homology. This can be achieved by targeting a nuclease to this region. For example, by using a DNA-binding protein that recognizes sequences in the native chromosomal location. One way to achieve this targeting is to use zinc-finger nucleases. These proteins have a modular composition and contain individual zinc finger domains, each of which can recognize a 3-nucleotide sequence in the target sequence (e.g., a native chromosomal integration site described above). Some embodiments can employ zinc finger nucleases with combinations of individual zinc finger domains that target numerous chromosomal locations. For example, the disclosed chromosomal sequences surrounding the exemplified integration sites in the Cpsf4, Ank2, C-

Mos, and Nephrocystin-1/Mal genes contain 8, 6, 7, and 8 candidate sites, respectively, that can be targeted by an engineered zinc finger nuclease.

[0046] Other than homologous recombination, insertion of heterologous polynucleotides into the native chromosomal integration sites in or near the Cpsf4, Ank2, c-Mos, and Nephrocystin-1/Mal genes can also be accomplished via the use of a hybrid recombinase. The recombinant recombinase is an engineered protein that has a recombinase domain (e.g., from phiC31 integrase) linked to a DNA targeting domain (e.g. a zinc finger domain). Such a molecule can be targeted to a site contained in or near the Cpsf4, Ank2, c-Mos, and Nephrocystin-1/Mal genes. Such recombinant proteins would enable integration of a recombinant construct into these chromosomal locations. Advantages of this approach include the ability to target into cell lines without the necessity of prior introduction of a landing pad (as described below), and a higher efficiency than homologous recombination.

[0047] Although zinc finger proteins have been well studied for their ability to bind to DNA and are suitable for the above applications, it may be possible to specifically target the Cpsf4, Ank2, c-Mos, and Nephrocystin-1/Mal genes by using other approaches, for example by mutation of another type of DNA binding domain. Other DNA binding domains include leucine-zippers and helix-turn-helix structures. It may also be possible to specifically target the Cpsf4, Ank2, c-Mos, and Nephrocystin-1/Mal genes by using a nucleic acid moiety to base pair to sequences in these genes.

[0048] Some embodiments include the direct integration of a transgene into the native chromosomal integration sites by either homologous recombination or by using a hybrid recombinase. The transgene can be any recombinant gene that encodes a therapeutic or industrial protein, e.g., a hormone or an enzyme, as detailed below. Some other embodiments are directed to inserting one or more recombinase recognized site specific recombination sequences (chromosomal landing pads) into the native chromosomal integration sites disclosed herein. As detailed herein, the chromosomal landing pads stably inserted into the host genome can in turn be used for integrating and expressing transgenes in the host cell (e.g., a CHO cell or other mammalian cells). Engineered host cells bearing one or more chromosomal landing pads at the native chromosomal integration sites disclosed herein are useful for site-specific integration and stable expression of any desired target gene.

IV. Integrating heterologous polynucleotides via homologous recombination

[0049] In one aspect, disclosed are methods and compositions for stably integrating heterologous polynucleotides into the native chromosomal integration sites via homologous recombination. Provided herein are polynucleotide molecules and vectors ("inserting vector") for inserting a heterologous polynucleotide (a transgene or a site-specific recombination sequence) into a host genome at the native chromosomal integration sites or specific chromosomal locations described herein. The polynucleotides and/or inserting vectors typically include a heterologous polynucleotide sequence (e.g., a recombinant gene or a chromosomal landing pad), a first homology arm, and a second homology arm. The polynucleotide or vector can additionally also include marker genes or sequences for positive and/or negative selections.

[0050] The heterologous polynucleotide sequence to be integrated into the host genome can encode any therapeutically or industrially useful proteins as described herein. It can also be a recombinase recognized integration site (chromosomal landing pad) which is then used for insertion and expression of a trangene, as detailed below. The first and the second homology arms are intended to target the heterologous polynucleotide sequence to a specific chromosomal location (e.g., a native chromosomal insertion site disclosed herein) for homologous recombination. As such, they are sequences that are substantially identical to the 5'- and the 3'- flanking sequences, respectively, of the native chromosomal integration site. As explained above, the native chromosomal integration sites are present within or adjacent to the coding or non-coding regions of one of the 4 specific genes, the Ank2 gene, the Cpsf4 gene, the C-Mos gene, and the Nephrocystin-1/Mal gene. Nephrocystin-1 gene is found 5' to the Mal gene. The insertion site can be 5' to the Mal gene between the Nephrocystin-1 and Mal genes. This genomic region is described herein as "Nephrocystin-1/Mal." As one can readily determine whether insertion of a heterologous polynucleotide at a given position in or around one of these genes leads to stable integration and /or expression, the exact position of the native chromosomal integration site with respect to each of the genes in the genome is not essential. Nevertheless, some preferred native chromosomal integration sites are described herein for Chinese hamster ovary (CHO) cells. As exemplified in the Examples below, the native chromosomal integration sites for CHO cells can be preferably between positions 130-131 of SEQ ID NO:1 for the Ank2 gene, between positions 629-630 of SEQ ID NO:2 for the Cpsf4 gene, between positions 272-273 of SEQ ID NO:3 for the C-Mos gene, between positions 239-240 of SEQ ID NO:4 for the Nephrocystin-1/Mal gene.

The native chromosomal integration sites for CHO cells can also be between positions 26,123-175,773 of NCBI No. NW_003615916.1 for the Ankyrin 2 gene (between positions 23 and 152,773 of SEQ ID NO: 9) or between nucleotides 844-845 of NCBI No. NW_003635654.1 for the Ankyrin2 gene (SEQ ID NO: 10), between positions 858,966-859,967 of NCBI NW_003614125.1 for the Cpsf4 gene (positions 966 – 967 of SEQ ID NO: 11) or between positions 858,533-859,237 of NCBI NW_003614125.1 for the Cpsf4 gene (positions 533 – 1237 of SEQ ID NO: 11), between positions 400,355-400,356 of NCBI NW_003614707.1 for the C-Mos gene (positions 355 – 356 of SEQ ID NO: 12) or between positions 398,595-399,212 of NCBI NW_003614707.1 for the C-Mos gene (SEQ ID NO: 12), and between positions 1,578,738-1,578,739 of NCBI NW_003613665.1 for the Nephrocystin-1/Mal gene (positions 738 – 739 of SEQ ID NO: 13) or between positions 1,574,453-1,625,306 of NCBI NW_003613665.1 for the Nephrocystin-1/Mal gene (SEQ ID NO: 13). The sequences and NCBI Reference numbers are incorporated by reference in their entirety.

[0051] It should be appreciated that the native chromosomal integration sites can also vary from cell line type to cell line type. For example, the nucleotide sequence of Ank2 gene of the CHO DG44 cell line can differ from the nucleotide sequence of the Ank2 gene of the CHO-K1 cell line as can the exact location of the native chromosomal integration site of the two cell line types. Thus, in some embodiments, the chosen native chromosomal integration sites for inserting a heterologous nucleotide sequence (a transgene or a chromosomal landing pad) can be at or close to each of these specific positions in CHO genome. Preferred native chromosomal integration sites between cell line type or for other mammalian cells (e.g., mouse cell, rat cell and human cell) can be determined based on sequence homology among the same gene in different mammalian species.

[0052] Once the exact native chromosomal integration site for inserting a heterologous polynucleotide sequence is determined, the homology arms which are substantially identical to the flanking sequences can then be readily designed and synthesized. Length of the homology arms is not essential, as long as they are capable of directing the homologous recombination at the desired site. Thus, the homology arms can be sequences comprising at least 10 bp, 25 bp, 50 bp, 100 bp, 200 bp, 500 bp, 1 kb, 2, kb, 5 kb, 10 kb or more contiguous nucleotide pairs of the sequences that flank the desired native chromosomal insertion site. In some embodiments, the homology arms comprise sequences identical to sequences that flank one of the exemplified chromosomal insertion sites in CHO genome (Figures 3-6) or

corresponding positions (as determined by sequence alignment) in other mammalian genomes. In some other embodiments, sequences that are substantially identical (e.g., at least 75%, 80%, 90%, 95% or 99% identical) to the flanking sequences of the native integration sites are employed as the homology arms in the polynucleotide molecules and vectors described herein. For example, the homology arms can include part or all of the sequences flanking the exemplified native integration site in each of these genes in CHO cells as shown in Figures 3-6.

[0053] The genes (Ank2, Cpsf4, C-Mos gene, and Nephrocystin-1/Mal gene) in cells from various species (e.g., CHO cells) have also been described in the art. For example, human Ank2 gene (accession nos. NG_009006; NW_003615916.1; NW_003635654.1), Cpsf4 gene (accession nos. EF191081; NW_003614125.1), C-Mos gene (Neel et al., Proc. Natl. Acad. Sci. USA, 79: 7842-6, 1982; and Morris et al., Hum. Genetics 81:339-342; accession no. NW_003614707.1), Nephrocystin-1/Mal gene (Alonso et al., Proc. Natl. Acad. Sci. USA 84:1997-2001, 1987; and Rancano et al., J. Biol. Chem. 269:8159-8164, 1994; accession no. NW_003613665.1) have all been characterized in the art. A skilled artisan can readily design and synthesize appropriate homology arm sequences for various applications. As exemplified in the Examples, sequences flanking one of the identified integration sites with a length of about 1 kb to 5 kb can be employed as the homology arms of the inserting vector for homologous integration of a heterologous polynucleotide (e.g., a landing pad) into a host genome. In some embodiments, the entire gene loci can be employed. In other embodiments, the entire gene loci plus 1, 2 or more kb on at least one of the 5′ and 3′ ends can be employed. In some embodiments, such as for Cpsf4 gene and C-Mos gene, the entire gene loci plus 2kb on each of the 5′ and 3′ ends can be employed. In other embodiments, such as for Nephrocystin-1/Mal gene and Ankyrin-2 gene, the entire gene loci can be employed.

[0054] In some specific embodiments, the heterologous polynucleotide sequence to be integrated into a host genome is site-specific recombination sequence that is recognized by a site-specific recombinase, e.g., a phage integrase such as the phiC-31 phage integrase. The site-specific recombination sequences to be inserted into the native chromosomal integration sites can be any sequence that supports site-specific recombination and is recognized by a unidirectional site-specific recombinase. Preferably, the site-specific recombination sequence comprises the phage attachment site (e.g., attP site) or the bacterial attachment site (e.g., attB site) recognized by an integrase (e.g., a tyrosine integrase or a serine integrase). Examples of

such sequences include attB and attP sequences (as well as pseudo att sites) recognized by several phage integrases, e.g., phiC-31 integrase or λ integrase. Suitable recombination sites also include sequences that are recognized by mutant integrases. During the integration of the phage genome into the genome of its host (e.g., an E. coli cell), the enzyme catalyzes the DNA exchange between the attP site of the phage genome and the attB site of the bacterial genome, resulting in the formation of attL and attR sites. By inserting into the host genome (e.g., at the native chromosomal integration sites disclosed herein) a site-specific recombination site (e.g., attP site) that is recognized by a phage integrase (e.g., phiC-31 integrase), a heterologous polynucleotide attached to the cognate recognition site (e.g., attB site) can be readily inserted into the host genome via site-specific recombination catalyzed by the phage integrase.

[0055] The phage attachment site (attP) and the bacterial attachment site (attB site) recognized by any site-specific recombinase (e.g., serine or tyrosine phage integrases) may be employed as the site-specific recombination sequence described herein. These include both the wildtype (native) attB and attP sites recognized by a given phage integrase as well as pseudo sites. Site-specific recombinases and their respective recognition sequences (attP and attB sites) for various phages and other species have been known and characterized in the art. Examples include λ phage integrase (Enquist et al., Cold Spring Harbor Symp. Quant. Biol. 43:1115–1120, 1979), HK022 phage integrase (Yagil et al., J. Mol. Biol. 207:695–717, 1989), P22 phage integrase (Leong et al., J. Biol. Chem. 260:4468–4477, 1985), HP1 phage integrase (Waldman et al., J. Bacteriol. 165:297-300, 1986), L5 phage integrase (Lee et al., J. Bacteriol. 175:6836–6841, 1993), phiC-31 phage integrase (Kuhstoss and Rao, J. Mol. Biol. 222:897–908, 1991), R4 phage (Groth et al., Proc. Natl. Acad. Sci. USA 97:5995–6000, 2000), TP901 phage integrase (Christiansen et al., J. Bacteriol. 178:5164–5173, 1996), γδ transposon resolvase (Reed et al., Nature 300:381–383, 1982), Tn3 transposon resolvase (Krasnow et al., Cell 32:1313–1324, 1983) and Mu phage invertase Gin (Kahmann et al., Cell 41:771–780, 1985).

[0056] Other than wild type recombination sites that are recognized by site-specific recombinases, the site-specific recombination sequence present in the polynucleotide molecules or vectors for landing pad insertion can also comprise a sequence that is different from the wild-type recognition site (e.g., wild type attP site) by at least one base pair alteration (a substitution, deletion or insertion). Sequence alterations may be at any position within the site-specific recombination sequence. In some embodiments, the modified site-

specific recombination sequences have multiple sequence alterations as compared to a wild type recognition site. When such a modified site-specific recombination sequence (e.g., a modified attP site) is integrated into the genome of an engineered host cell as described herein, the wild type or mutant version of the corresponding integrase (e.g., a mutant phi-C31 integrase) may be needed in order to incorporate a heterologous polynucleotide or transgene into the recombination site. Various mutant integrases (e.g., mutant phiC-31 integrase) are also known in the art. See, e.g., Smith et al., Nuc. Acids Res. 32, 2607-2617, 2004; and Kevarala et al., Mol. Ther. 17, 112–120, 2008.

[0057] For inserting a heterologous polynucleotide sequence (a transgene or a chromosomal landing pad) into the genome of a host cell, the polynucleotide described above is typically present in a vector ("inserting vector"). These vectors are typically circular and linearized before used for homologous recombination. In addition to the homology arms and the heterologous polynucleotide (e.g., a landing pad), the vectors may also contain markers suitable for selection or screening, an origin of replication, and other elements. As exemplified in the Examples herein, the vector can contain both a positive selection marker and a negative selective marker. The positive selection marker, e.g., an antibiotic resistance gene, is used to identify host cells into which the vector has stably integrated. Examples of such markers include antibiotic resistance genes for neomycin, blasticidin, hygromycin and zeocin. The negative selection marker, e.g., a suicide gene, serves to eliminate cells that have randomly integrated the vector sequence while retaining cells that have undergone homologous recombination at the desired location. An Example of such negative selection marker is the HCV-TK gene as exemplified in the Examples herein. The positive screening marker (e.g., enhanced green fluorescent protein) is used to identify host cells into which the vector has stably integrated (e.g., by using fluorescently activated cell sorting, FACS). The negative screening marker, e.g., cyan fluorescent protein, is used to identify cells (e.g., by FACS) that have randomly integrated the vector sequence. FACS for cells containing the positive screening marker but lacking the negative screening marker will identify cells that have undergone homologous recombination at the desired location.

[0058] One more component of the inserting vector (as well as the targeting vector described below) is an origin of replication. Replication origins are unique DNA segments that contain multiple short repeated sequences that are recognized by multimeric origin-binding proteins and that play a key role in assembling DNA replication enzymes at the origin site. Suitable origins of replication for use in the vectors include, e.g., EBV oriP,

SV40, E. coli oriC, colE1 plasmid origin, ARS, and the like. Another useful element in an expression vector is a multiple cloning site or polylinker. Synthetic DNA encoding a series of restriction endonuclease recognition sites is inserted into a plasmid vector, for example, downstream of the promoter element. These sites are engineered for convenient cloning of DNA into the vector at a specific position.

[0059] The polynucleotides or vectors for inserting the heterologous polynucleotide into a host genome can be readily constructed in accordance with standard procedures known in the art of molecular biology (e.g., Sambrook et al., *supra*; and Brent et al., *supra*) and the disclosure herein. To generate the vectors, the above-described polynucleotides comprising the homology arms and the heterologous polynucleotide sequence (e.g., a transgene or a chromosomal landing pad) can be inserted into various known plasmids for transfecting mammalian host cells. Such known plasmids include, e.g., BPV, EBV, vaccinia virus based vector, SV40, 2-micron circle, pcDNA3.1, pcDNA3.1/GS, pYES2/GS, pMT, p IND, pIND(Sp1), pVgRXR (Invitrogen), and the like, or their derivatives. These plasmids are all described and well known in the art (Botstein et al., Miami Wntr. SyTnp. 19:265-274, 1982; Broach, In: *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., p. 445-470, 1981; Broach, Cell 28:203-204, 1982; Dilon et at., J. Clin. Hematol. Oncol. 10:39-48, 1980; and Maniatis, In: *Cell Biology: A Comprehensive Treatise*, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563-608, 1980.

V. Engineered cells with integrated heterologous polynucleotides

[0060] Provided herein are recombinant or engineered host cells which contain heterologous polynucleotides (recombinant genes or chromosomal landing pads) that are stably integrated into the genome at one or more of the native chromosomal integration sites disclosed herein. Cells with recombinant genes integrated at the disclosed sites will allow stable and strong expression of polypeptides encoded by the genes. Cells with integrated chromosomal landing pads allow for efficient site-specific integration and/or expression of a target polynucleotide or gene of interest. Engineered host cells can also include cells which bear such inserted chromosomal landing pads and which then have one or more transgenes integrated into the landing pads, as explained below. Using the polynucleotide molecules or inserting vectors described above, various cells can be modified by inserting recombinant

genes or chromosomal landing pads at one or more of the specific chromosome locations described herein.

[0061] The recombinant polynucleotides or inserting vectors described above (or targeting vectors described below) can be introduced into an appropriate host cell (e.g., a mammalian cell such as CHO cell) by any means known in the art. Typically, after appropriate restriction enzyme digestion to generate free ends of homology to the host chromosome, the polynucleotide can then be transfected into host cells. The linearized inserting vectors can be introduced into the host cell by standard protocols routinely practiced in the art. For example, the vector can be transfected into the host cell by calcium phosphate co-precipitation, by conventional mechanical procedures such as microinjection or electroporation, by insertion of a plasmid encased in liposomes, and by virus vectors. These techniques are all well known and routinely practiced in the art, e.g., Freshney, *supra*; Sambrook et al., *supra*; and Brent et al., *supra*). Host cells which harbor the transfected recombinant inserting vector can be identified and isolated using the selection marker present on the vector. Large numbers of recipient cells may then be grown in a medium which selects for vector-containing cells.

[0062] A specific vector for inserting a site-specific recombination sequence (i.e., attP sequence) into a native chromosomal insertion site is exemplified herein (Fig. 2). In addition to the attP sequence, the vector bears homology arms which support homologous recombination at the noted native chromosomal insertion site and also selection markers. For integrating into the CHO genome at the desired native insertion site, the vector was first linearized via restriction digestion. After transfecting the linearized sequence into a host cell (e.g., CHO cell), the cells are then subjected to positive and negative selections to identify cells which have integrated site-specific recombination site (attP site) via homologous recombination. Cells thus identified can then be further examined to ascertain integration of the heterologous polynucleotide at the chosen native chromosomal insertion site. As disclosed herein, cells with integrated recombinant genes can be directly used for production of therapeutic or industrial proteins encoded by the genes. Alternatively, cells with inserted chromosomal landing pads can be employed for production of a target polypeptide by integrating into the chromosomal landing pad a polynucleotide sequence that encodes the target polypeptide.

[0063] Preferably, host cells for inserting one or more heterologous polynucleotides at the native chromosomal insertion sites are eukaryotic cells. Eukaryotic vector/host systems, and

mammalian expression systems in particular, allow for proper post-translational modifications of expressed mammalian proteins to occur, *e.g.*, proper processing of the primary transcript, glycosylation, phosphorylation and advantageously secretion of expressed product. Therefore, eukaryotic cells such as mammalian cells are the preferred host cells for inserting the heterologous polynucleotides (recombinant genes or chromosomal landing pads) at the native chromosomal locations described herein. Suitable cells include both animal cells (such as cells from insect, rodent, cow, goat, rabbit, sheep, non-human primate, human, and the like) and plant cells (such as rice, corn, cotton, tobacco, tomato, potato, and the like). Specific examples of such host cell lines include CHO, BHK, HEK293, VERO, HeLa, COS, MDCK, PER.C6, and W138.

[0064] In some embodiments, provided are recombinant cells which have a polynucleotide of interest or transgene already stably integrated into a landing pad that has been pre-inserted at a native chromosomal location described herein. Targeting vectors for integrating a target polynucleotide into a chromosomal landing pad that has already been inserted into the host genome are described in more detail below. As described herein, the landing pad comprises a recognition sequence (*e.g.*, attP site) that is recognized by a site-specific recombinase (*e.g.*, a phage integrase such as phi-C31 integrase). By attachment to a cognate recognition sequence (*e.g.*, attB site) that is also recognized by the recombinase, the polynucleotide of interest along with appropriate transcription regulatory elements are integrated into the landing pad via site-specific recombination mediated by the recombinase. The integrated polynucleotides of interest in the recombinant cells can encode any protein or polypeptide useful in industrial or therapeutic applications. Specific examples of such polypeptides and proteins are described above. These include *e.g.*, enzymes (*e.g.*, proteases, phospholipases, and the like), protease inhibitors, hormones (*e.g.*, pituitary hormones), growth factors, cytokines, chemokines, chemotactins, gonadotrophins, lipid-binding proteins, somatamedians, gonadotrophins, and immunoglobulins. Other proteins of interest include antimicrobial polypeptides (e.g., antibacterial, antifungal, antiviral, and/or antiparasitic polypeptides), and antibodies or antigen-binding antibody fragments (*e.g*, FAbs) thereof (*e.g.*, orthoclone OKT-e (anti-CD3), GPIIb/IIa monoclonal antibody).

[0065] Other than mammalian cells, the host cell for inserting heterologous polynucleotides as described herein may also be a yeast cell or a plant cell. Yeast or plant cells thus engineered are suitable for stable integration and expression of a transgene that is introduced into the host via a yeast or plant expression vector. Examples of suitable inset

cells include cells from Drosophila larva. When insect cells are used, the heterologous polynucleotides can be introduced into the cells via appropriate inserting vectors. For example, baculovirus vectors can be employed as described in the art (Jasny, Science 238:1653, 1987; and Miller et al., In: *Genetic Engineering* (1986), Setlow, J. K., et al., eds., Plenum, Vol. 8, pp. 277-297). When insect cells are employed as hosts, the Drosophila-alcohol dehydrogenase promoter can optionally be used in the inserting vector for inserting the heterologous polynucleotides (Rubin, *Science* 240:1453-1459, 1988).

VI. Integrating target polynucleotides into chromosomal landing pads

[0066] As described above, a target polynucleotide or transgene encoding a polypeptide (*i.e.*, a "polynucleotide of interest" or a "gene of interest") can be directly integrated into the native chromosomal integration sites disclosed herein. Stable and efficient expression and production of any of the therapeutic or industrial proteins described below can be achieved in this manner. Alternatively, a target polynucleotide can be integrated into a host genome via a chromosomal landing pad that has already been inserted at a native chromosomal integration site disclosed herein. Employing engineered host cells bearing inserted chromosomal landing pads described herein, also provided are vectors ("targeting vector") and methods for integrating and expressing a heterologous polynucleotide or transgene in the cell. Polynucleotides of interest that encode various useful target polypeptides can be stably integrated into the genome of an engineered host cell described herein. The polynucleotides of interest can be either endogenous or exogenous to the host cell. An exogenous polynucleotide is a nucleic acid molecule having a sequence that is not naturally present in the host cell while an endogenous polynucleotide is a nucleic acid molecule with a sequence that pre-exists in the host cell. Many specific examples of proteins or polypeptides that can be expressed are described below.

[0067] Depending on the engineered host to be used, a variety of targeting vectors are suitable for use. As the preferred host cell bearing the inserted chromosomal landing pad is a mammalian cell (*e.g.*, CHO cell), the targeting vector is preferably a vector for eukaryotic expression. In general, the targeting vector will have the gene of interest attached to a cognate recombination site or a recognition sequence. The cognate recombination site on the vector is also recognized by the site-specific recombinase (*e.g.*, phiC-31 integrase) which recognizes the inserted chromosomal landing pad. As such, the cognate recombination site on the vector will support the recombinase mediated integration of the target polynucleotide

into the landing pad. For example, for integration and expression in an engineered host cell bearing an inserted phage attachment site (attP) of a specific phage integrase, the vector will have the target polynucleotide attached to the cognate bacterial attachment site (attB site) which is also recognized by the same integrase. Similarly, if the inserted landing pad comprises the attB site of a phage integrase, the targeting vector will comprise the cognate attP site recognized by the integrase. Some phage integrases, such as phi-C31 and R4, prefer to integrate into phage attachment sites (attP sites) rather than bacterial attachment sites. With these enzymes, the targeting vector should carry the attB site while the landing pad should comprise the attP site. Other phage integrases preferentially integrate into bacterial attachment sites (*e.g.*, pseudo attB), rather than phage attachment sites. Examples of enzymes with this preference are phiBT1 integrase and A118 integrase. When these integrases are used, the target vector should carry the attP site instead of the attB site while the corresponding host cell should contain the attB site in the inserted landing pad.

[0068] To support expression of the target polynucleotide upon integration at the landing pad, the targeting vector can also contain promoter sequence and other transcription regulatory elements (*e.g.*, enhance sequences) that is operably linked to the target polynucleotide. In general, promoters can be selected such that they are functional in the cell type into which they are being introduced. Many promoters known in the art can be used for expression in mammalian host cells. Examples include, but are not limited to, the promoter of the mouse metallothionein I gene sequence (Hamer et al., *J. Mol. Appl. Gen.* 1:273-288, 1982); the TK promoter of Herpes virus (McKnight, *Cell* 31:355-365, 1982); the SV40 early promoter (Benoist et al., *Nature (London)* 290:304-310, 1981); the yeast gal1 gene sequence promoter (Johnston et al., *Proc. Natl. Acad. Sci. (USA)* 79:6971-6975, 1982); Silver et al., *Proc. Natl. Acad. Sci. (USA)* 81:5951-59SS, 1984), the CMV promoter, the EF-1 promoter, the actin promoter, the phosphoglycerate kinase promoter, the ubiquitin promoter and the thymidine kinase promoter, the ecdysone-responsive promoter(s), tetracycline-responsive promoter, and the like.

[0069] In addition, the targeting vector can have selection or screening marker sequences, an origin of replication, and the like. As with markers used in the inserting vectors described above, the selection or screening markers in the targeting vectors also provide a means to select or screen for growth of only those cells that contain the vector. Such selection markers are typically of two types: drug resistance and auxotrophic. A drug resistance marker enables cells to detoxify an exogenously added drug that would otherwise kill the cell. Auxotrophic

markers allow cells to synthesize an essential component (usually an amino acid) while grown in media that lacks that essential component. Common selectable marker genes include those for resistance to antibiotics such as ampicillin, tetracycline, kanamycin, bleomycin, streptomycin, hygromycin, neomycin, Zeocin™, G418, and the like. Selectable auxotrophic genes include, for example, hisD, that allows growth in histidine free media in the presence of histidinol.

[0070] The selection marker sequences and the transcription regulatory elements should be linked to the target polynucleotide and the cognate recombinase recognition sequence in the vector in such a way that they will co-integrate with the target polynucleotide into the host genome once site-specific recombination at the landing pad takes place. The targeting vectors described herein can be constructed utilizing methodologies known in the art of molecular biology in view of the teachings of the specification. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (3rd ed., 2001); Brent et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (ringbou ed., 2003); and Freshney, *Culture of Animal Cells: A Manual of Basic Technique*, Wiley-Liss, Inc. (4th ed., 2000). Typically, the targeting vectors are assembled by inserting into a suitable vector backbone a recombination site cognate to the landing pad, polynucleotides of interest, sequences encoding selection markers, and other optional elements described herein.

[0071] In addition to an engineered host cell bearing an inserted chromosomal landing pad and the targeting vector, site specific integration of the target polynucleotide at the landing pad (e.g., an attP site) will also require catalytic activities of the corresponding recombinase (e.g., a phage integrase such as phiC-31 integrase). The recombinase (e.g., phiC-31 integrase) can be introduced into a target cell before, concurrently with, or after the introduction of a targeting vector. As explained above, various phage integrases are considered herein. The specific integrase used in integrating a target polynucleotide into an engineered host cell should correspond to and recognize the site-specific recombination sequence in the landing pad of the host genome and the cognate recognition sequence in the targeting vector. In some embodiments, the unidirectional site-specific recombinase is a serine integrase. Serine integrases that may be useful for in vitro and in vivo recombination include, but are not limited to, integrases from phages phi-C31, R4, TP901-1, phiBT1, Bxb1, RV-1, A118, U153, and phiFC1, as well as others in the long serine integrase family. See, e.g., Gregory et al., *J. Bacteriol.*, 185:5320-5323, 2003; Groth and Calos, *J. Mol. Biol.*

335:667-678, 2004; Groth et al., *Proc. Natl. Aacd. Sci.* 97:5995-6000, 2000; Olivares et al., *Gene* 278:167-176, 2001; Smith and Thorpe, *Molec. Microbiol.*, 4:122-129, 2002; and Stoll et al., *J. Bacteriol.*, 184:3657-3663, 2002. In addition to these wild-type integrases, altered integrases that bear mutations are also known in the art (see, *e.g.*, Sclimenti et al., *Nuc. Acid Res.* 29:5044-5051, 2001). Such integrases with altered activity or specificity compared to the wild-type are also useful for the recombination reaction and the integration of target polynucleotides into an engineered host genome.

[0072] In some embodiments, a purified enzyme polypeptide is introduced into the host cell to mediate the integration of the targeting vector. Methods of introducing functional proteins into cells are well known in the art. For example, a phage integrase polypeptide such as phiC-31 integrase can be directly introduced into a cell by many means, including liposomes, coated particles, whiskers, microinjection, electroporation, and peptide transporters (see, *e.g.*, Siprashvili *et al.*, *Mol. Ther.*, 9:721-728, 2004). In some other embodiments, a polynucleotide encoding the integrase can be introduced into the cell using a suitable expression vector. The integrase can be expressed from the same targeting vector expressing the gene of interest. Alternatively, polynucleotide encoding the integrase can be introduced into the host cell via a second vector. In some embodiments, a DNA sequence encoding the integrase is introduced into the host cell on an expression vector. This can be performed as described in the art, *e.g.*, Olivares *et al., Gene*, 278:167-176, 2001; and Thyagarajan *et al., Mol. Cell Biol.* 21:3926-3934, 2001. In some other embodiments, the site specific integration relies on transient presence of a RNA molecule encoding the recombinase polypeptide. For example, an mRNA molecule encoding a phage integrase can be introduced into and expressed in a host cell as described in, *e.g.*, Groth *et al., J. Mol. Biol.* 335:667-678, 2004; and Hollis *et al., Repr. Biol. Endocrin.* 1:79, 2003. It is generally preferred that the integrase be present for only such time as is necessary for insertion of the targeting vector into the genome of the engineered host cell. Introduction of integrase-encoding RNA (*e.g.*, an mRNA) can ensure transient expression and removes the possibility that an integrase-encoding nucleic acid will become permanently incorporated into a target genome. Transient expression of the site-specific recombinase can also be achieved via other means. For example, polynucleotide expressing the enzyme can be placed under the control of a regulatable promoter (*i.e.*, a promoter whose expression can be selectively induced or repressed).

[0073] Any convenient protocol may be employed for *in vitro* or *in vivo* introduction of the targeting vector and/or a second vector expressing a phage integrase into the target cell, depending on the location of the target cell. For example, where the engineered host cell is an isolated cell, the targeting vector may be introduced directly into the cell under cell culture conditions permissive of viability of the target cell, *e.g.,* by using standard transformation techniques. Such techniques include, but are not necessarily limited to: viral infection, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct microinjection, viral vector delivery, and the like. The choice of method is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place *(i.e. in vitro, ex vivo,* or *in vivo*). A general discussion of these methods can be found in, *e.g.*, Brent et al, *supra.*

[0074] Alternatively, where the engineered host cell or cells are part of a multicellular organism, the targeting vector may be administered to the organism or host in a manner such that the targeting vector is able to enter the host cell(s), e.g., via an in vivo or ex vivo protocol. By "in vivo," it is meant in the target construct is administered to a living body of an animal. By "ex vivo" it is meant that cells or organs are modified outside of the body. Such cells or organs are typically returned to a living body. Methods for the administration of nucleic acid constructs are well known in the art. For example, nucleic acid constructs can be delivered with cationic lipids (Goddard, et al, *Gene Therapy*, 4:1231-1236, 1997; Gorman *et al., Gene Therapy* 4:983-992, 1997; Chadwick *et al., Gene Therapy* 4:937-942, 1997; Gokhale *et al., Gene Therapy* 4:1289-1299, 1997; Gao and Huang, *Gene Therapy* 2:710-722, 1995), using viral vectors (Monahan *et al., Gene Therapy* 4:40-49, 1997; Onodera *et al., Blood* 91:30-36, 1998), by uptake of "naked DNA", and the like. Techniques well known in the art for the transfection of cells (see discussion above) can be used for the *ex vivo* administration of nucleic acid constructs. The exact formulation, route of administration and dosage can be chosen empirically. See *e.g.* Fingl *et al.*, 1975, in *The Pharmacological Basis of Therapeutics*, Ch. 1 p 1).

VII. Target polypeptides or proteins to be expressed with engineered host cells

[0075] The engineered host cells described above are useful for stable expression of any polynucleotide of interest. The polynucleotides of interest can encode various polypeptides with medical or industrial applications. In some embodiments, the target polynucleotide or polynucleotide of interest to be integrated into the landing pad in the engineered host cell can

be one that encodes a therapeutic protein. Examples of therapeutic proteins include factor VIII, factor IX, β-globin, low-density lipoprotein receptor, adenosine deaminase, purine nucleoside phosphorylase, sphingomyelinase, glucocerebrosidase, cystic fibrosis transmembrane conductance regulator, α-antitrypsin, CD-18, ornithine transcarbamylase, argininosuccinate synthetase, phenylalanine hydroxylase, branched-chain α-ketoacid dehydrogenase, fumarylacetoacetate hydrolase, glucose 6-phosphatase, α-L-fucosidase, β-glucuronidase, α-L-iduronidase, galactose 1-phosphate uridyltransferase, interleukins, cytokines, small peptides, and the like. Other therapeutic proteins that can be expressed from an intergrated target polynucleotide in the engineered host cell can include, e.g., Herceptin®, polypeptide antigens from various pathogens such as disease causing bacteria or viruses (e.g., E. coli, P. aeruginosa, S. aureus, malaria, HIV, rabies virus, HBV, and cytomegalovirus), and other proteins such as lactoferrin, thioredoxin and beta-caseinvaccines.

[0076] Additional examples of proteins of interest include, but are not necessarily limited to insulin, erythropoietin, tissue plasminogen activator (tPA), urokinase, streptokinase, neutropoesis stimulating protein (also known as filgastim or granulocyte colony stimulating factor (G-CSF)), thrombopoietin (TPO), growth hormone, emoglobin, insulinotropin, imiglucerase, sarbramostim, endothelian, soluble CD4, and antibodies and/or antigen-binding fragments (e.g, FAbs) thereof (e.g., orthoclone OKT-e (anti-CD3), GPIIb/IIa monoclonal antibody), liary neurite transforming factor (CNTF), granulocyte macrophage colony stimulating factor (GM-CSF), brain-derived neurite factor (BDNF), parathyroid hormone(PTH)-like hormone, insulinotrophic hormone, insulin-like growth factor-1 (IGF-1), platelet-derived growth factor (PDGF), epidermal growth factor (EGF), acidic fibroblast growth factor, basic fibroblast growth factor, transforming growth factor β, neurite growth factor (NGF), interferons (IFN) (e.g., IFN-α2b, IFN-α2a, IFN-αN1, IFN-β1b, IFN-γ), interleukins (e.g, IL-1, IL-2, IL-8), tumor necrosis factor (TNF) (e.g, TNF-α, TNF-β), transforming growth factor-α and -β, catalase, calcitonin, arginase, phenylalanine ammonia lyase, L-asparaginase, pepsin, uricase, trypsin, chymotrypsin, elastase, carboxypeptidase, lactase, sucrase, intrinsic factor, vasoactive intestinal peptide (VIP), calcitonin, Ob gene product, cholecystokinin (CCK), serotonin, and glucagon.

[0077] Suitable polypeptides of interest that can be expressed from the integrated target polynucleotides also include specific membrane proteins or other intracellular proteins. Examples of membrane proteins include, but are not necessarily limited to adrenergic receptors, serotonin receptors, low-density lipoprotein receptor, CD-18, sarcoglycans (which

are deficient in muscular dystrophy), etc. Useful intracellular proteins include proteins that are primarily located within the intracellular compartment or which exhibit a desired biological activity within a cell. Such intracellular proteins can include fumarylacetoacetate hydrolase (FAH) which is deficient in subjects with hereditary tyrosinemia Type 1. Other specific examples of intracellular proteins include antiviral proteins (e.g., proteins that can provide for inhibition of viral replication or selective killing of infected cells), structural protein such as collagens, i.e. the type VII collagen COL7A1 gene, defective in Recessive Dystrophic Epidermolysis Bullosa (RDEB) and dystrophin, defective in muscular dystrophy.

VIII. Kits and transgenic animals with integrated transgenes

[0078] Provided herein are kits for using the engineered host cells described above. The kits enable a skilled artisan to site-specifically integrate and/or express a heterologous polynucleotide in an engineered host cell which bears a target transgene or an inserted chromosomal landing pad at one or more native chromosomal integration sites disclosed herein. Some kits described herein contain engineered host cells (*e.g.*, CHO cells) which have a target polynucleotide directly inserted at a native chromosomal integration site in the genome. Some other kits contain engineered host cells which have a target polynucleotide inserted at one or more chromosomal landing pads that have been pre-integrated into native chromosomal integration sites in the genome. Still some other kits described herein contain recombinant cells with inserted chromosomal landing pad at one or more native chromosomal integration sites and other reagents for inserting a target polynucleotide into the chromosomal landing lads.

[0079] As exemplification, some kits described herein contain at least one or more of the following components, an engineered host cell (e.g., a CHO cell line) bearing an inserted landing pad (e.g., an attP site) at one or more of the native chromosomal locations described herein, a targeting vector for cloning and integrating a heterologous polynucleotide, and an integrase component (e.g., phiC-31). The kits can optionally also contain a target polynucleotide that is to be cloned into the targeting vector and expressed in the host cell. Typically, upon cloning into the targeting vector, the heterologous target polynucleotide is attached to a cognate sequence (e.g., an attB site) also recognized by the integrase for integrating at the inserted landing pad. As described herein, the integrase component can be provided in any suitable form (e.g., as a protein formulated for introduction into a target cell or in an integrase vector which provides for expression of the desired integrase following

introduction into the engineered host cell). Thus, some kits can comprise a substantially purified recombinase polypeptide (e.g., phiC-31). Some other kits can contain a second vector that allows expression of the enzyme in the host cell. The kits described herein can optionally contain other components, e.g., restriction enzymes for cloning a targeting polynucleotide, control plasmids, buffers, and etc. The various components of the kit may be present in separate containers or certain compatible components may be pre-combined into a single container, as desired.

[0080] In addition to the various reagents, the kits described herein typically further include instructions for using the components of the kit in integrating and expressing a polynucleotide of interest. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging). In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, etc. In yet some other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, the means for obtaining the instructions is recorded on a suitable substrate.

[0081] Further provided herein are transgenic non-human animals or plants whose genomes have been modified by inserting a heterologous polynucleotide (a transgene or a chromosomal landing pad) at one or more native chromosomal integration sites disclosed herein. The transgenic non-human animals or plants can also have a genome which has inserted chromosomal landing pads and then further modified by integrating one or more target polynucleotides at the inserted landing pads. Examples of transgenic animals that can be produced with methods described herein include mice, rats, chickens, cats, dogs, rabbit, pigs, goats, sheep, cows, horses, as well as non-human primates such as rhesus macaques. The transgenic non-human animals or plants described herein can be produced by integrating a heterologous polynucleotide or transgene into the genome at one or more of the native chromosomal integration sites. Other transgenic animals or plants are produced by integrating a target polynucleotide into a chromosomal landing pad that has already been

inserted into the genome as described herein. The target cell can be any cell amenable to genetic modification using the systems and methods described herein, and which is suitable to produce a transgenic animal described herein. Target cells can be isolated (e.g., in culture) or in a multicellular organism (e.g., in a blastocyst, in a fetus, in a postnatal animal, and the like). Exemplary target cells include, but are not necessarily limited to, primary cells, secondary cells, transformed cells, egg cells, fertilized egg cells, single cell embryos, somatic cells (e.g., muscle, bone, cartilage, ligament, tendon, skin (dermis, epidermis, and the like), cells of the viscera (e.g., lung, liver, pancreas, gastrointestinal tract (mouth, stomach, intestine), and the like), stem cells (e.g., embryonic stem cells (e.g., cells having an embryonic stem cell phenotype), adult stem cells, pluripotent stem cells, hematopoietic stem cells, mesenchymal stem cells, and the like), and germ cells (e.g., primordial germ cells, embryonic germ cells, and the like).

[0082] Transgenic animals or plants can be produced employing the methods that are routinely practiced by the skilled artisans in the art. See, e.g., Brinster, *et al., Proc. Nat. Acad. Sci. USA* 82: 4438, 1985; Houdebine and Chourrout, *Experientia* 47:897-905, 1991; *Teratocarcinomas and Embryonic Stem Cells, A Practical Approach*, E. J. Robertson, ed., IRL Press (1987); Hogan et al., Manipulating the Mouse Embryo (Cold Spring Harbor Press 1986); Krimpenfort et al., Bio/Technology 9:86, 1991; Palmiter et al., *Cell* 41:343, 1985; Kraemer et al., Genetic Manipulation of the Early Mammalian Embryo (Cold Spring Harbor Laboratory Press 1985); Hammer et al., *Nature*, 315:680, 1985; Purcel et al., *Science*, 244:1281, 1986; Pursel, et al., *Science* 244:1281-1288, 1989; Simms, *et al., Bio/Technology* 6:179-183, 1988; and U.S. Pat. No. 5,175,384, and U.S. Patent Nos. 4945050, 5175384 and 5175385.

EXAMPLES

[0083] The following examples are provided to further illustrate, but not to limit in scope, what is described herein. Other variants will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims.

Example 1 Identification of native chromosomal sites with strong transcription activities

[0084] Plasmid for random integration into CHO genome: The plasmid was modified based on the original attP containing plasmid described in Thyagarajan *et al.*, (2001, *Mol Cell Biol.* 21(12):3926-34). Specifically, the original plasmid was modified to replace the Zeocin

marker with a Neomycin marker. In addition, the firefly luciferase gene was replaced with the EGFP gene controlled by the SV40 promoter (Fig. 1).

[0085] Stable transfection: The modified plasmid was purified using a Qiagen midiprep column. Twenty-five μg of DNA was digested overnight with restriction enzyme BamHI to linearize the plasmid. The resulting linear DNA was then transfected into CHO cells to create stable integrations. Two days after transfection, cells were split into new plates and hygromycin was added to growth media for selection of stable integration events. Cells were grown in culture for three weeks before harvesting for FACS analysis.

[0086] FACS: After three weeks of growth in culture under hygromycin selection, stable cells were pooled together for bulk sorting at the Scripps FACS Core Facility; the top 1% of EGFP-expressing cells were collected, returned to culture media and allowed to recover and grow for several couple weeks. When the cell culture plates were confluent, cells were collected and sorted by FACS again. This time, the top 1% of EGFP-expressing cells were sorted as individual cells into the wells of 96-well plates.

[0087] Single cell populations: Cells sorted into 96-well plates were allowed to grow in these plates for two to three weeks before transferred to 24-well plates. After one week, cells were transferred to 6-well plates for expansion cultures. At this stage, EGFP expression was checked to ensure that the single cell populations contained a stably integrated the attP-containing plasmid constructs and maintained EGFP expression.

[0088] Growth rate check: After confirming the EGFP expression, the single cell populations were checked for growth rate along with the parental CHO cell line. Cells were seeded into 6-well plates at 10,000 cells per well. Cell numbers were counted at three time points: 24 hour; 48 hour and 72 hour after plating. Only cells that had growth rates equivalent to or faster than the parental CHO cell line were further cultured for stability studies.

[0089] Stability studies: Single cell populations were further cultured for up to four months to determine the expression stability. Cells were checked for EGFP expression once every month. At the end of the culture period, both of growth rate and EGFP expression were checked to make sure that the single cell populations had maintained a high level of EGFP expression and grew as fast as or even faster than the parental CHO cells. After this stage, twenty single cell populations were chosen as good candidates for identification of chromosome integration sites.

[0090] Identification of integration sites: Genomic DNA was purified from the top 20 single cell populations. Individual DNA samples were checked for concentration and 10 μg

of total DNA was used for enzyme digestion using four blunt end generating restriction enzymes: EcoRV, PvuI, StuI and HincII. The completely digested DNA samples were then subjected to purification with phenol and chloroform. These DNA samples were then precipitated using ethanol and ligated to a double stranded DNA linker molecule (GenomeWalker Adaptors).

[0091] Three gene specific primers (GSPs) were designed based on the hygromycin resistance gene. GSP1 and AP1 (Adaptor Primer 1) were used in primary PCR reactions; the GSP2 (nested gene specific primer) and AP2 (Adaptor Primer 2) were used in secondary PCR reactions. If needed, GSP3 and AP2 were used in tertiary PCR reactions to obtain specific products.

[0092] The results indicate that native chromosomal integration sites in the CHO genome that support stable integration and strong transcription activities are present in the ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene, and Nephrocystin-1/Mal gene. The exact positions of the genes for integration of the heterologous sequence are respectively indicated in Figures 3-6 (SEQ ID NOs:1-4).

Example 2 Inserting landing pads at identified native chromosomal integration sites

[0093] Homologous recombination: Genomic DNA flanking the integration sites was identified and cloned into a plasmid that contains both positive and negative selection markers (Fig. 2). For each site, the longer homology arm (3 to 4 kb in length) is cloned 5' of the neomycin resistance gene. The short homology arm (1.5 to 2 kb in length) is cloned 3' of the neomycin resistance gene. One single phage attachment site attP is located at the end of the long homology arm.

[0094] The homologous recombination plasmid is digested with NotI enzyme to linearize the plasmid; the long homology arm is at the one end of this linear DNA. Upon transfection into CHO cells, neomycin is added to culture media to select for cells that have this linear DNA integrated into the cell chromosome. A pool of resistance clones are obtained after 4 to 6 weeks. Then cells are subjected to negative selection with the addition of ganciclovir to the culture media, which will kill cells that have randomly integrated the plasmid. Only cells that have undergone homologous recombination at target site will survive both positive and negative selection. After both rounds of selection, cells that survive are picked and grown in 24-well culture plates and then expanded to 6-well plates. Genomic DNA is then isolated from these cell clones and checked for attP site integration into the targeted locations.

[0095] Landing pad integration: After verifying that the attP sites are inserted into the desired locations, these cell lines can be used for integration of recombinant genes into the attP sites using the phage Phi-C31 integrase system. Recombinant genes are cloned into a plasmid containing a single attB site. Upon cotransfection of plasmids containing the recombinant gene and the Phi-C31 integrase gene, cells can be selected for specific integration events.

***

[0096] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

[0097] All publications, databases, GenBank sequences, patents, and patent applications cited in this specification are herein incorporated by reference as if each was specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1. A method for stable integration and expression of a heterologous polynucleotide in a host cell, comprising inserting the heterologous polynucleotide into the genome of the host cell at a native chromosomal site located within or adjacent to a gene selected from the group consisting of ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene, and Nephrocystin-1/Mal gene.

2. The method of claim 1, wherein insertion of the heterologous polynucleotide is mediated by homologous recombination or by a hybrid recombinase.

3. The method of claim 1, wherein the host cell is a mammalian cell.

4. The method of claim 1, wherein the host cell is a Chinese hamster ovary (CHO) cell.

5. The method of claim 4, wherein the native chromosomal insertion site is at or close to positions 130-131 of SEQ ID NO: 1 for the Ank2 gene, positions 629-630 of SEQ ID NO: 2 for the Cpsf4 gene, positions 272-273 of SEQ ID NO: 3 for the C-Mos gene, positions 239-240 of SEQ ID NO: 4 for the Nephrocystin-1/Mal gene.

6. The method of claim 1, wherein the heterologous polynucleotide encodes a polypeptide.

7. The method of claim 6, wherein the polypeptide is a therapeutic protein or an industrial protein.

8. A recombinant polynucleotide for stably integrating a heterologous polynucleotide sequence into the genome of a mammalian cell, comprising:

a first homology arm, the heterologous polynucleotide sequence, and a second homology arm;

wherein the first and second homology arms are substantially identical to the 5'- and 3'- sequences, respectively, that flank a native chromosomal insertion site located within or adjacent to a gene selected from the group consisting of ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene, and Nephrocystin-1/Mal gene.

9. The polynucleotide of claim 8, wherein the native chromosomal insertion site supports stable integration of a foreign gene.

10. The polynucleotide of claim 8, wherein the heterologous polynucleotide sequence encodes a polypeptide.

11. The polynucleotide of claim 8, wherein the polypeptide is a therapeutic protein or an industrial protein.

12. The polynucleotide of claim 8, wherein the heterologous polynucleotide sequence comprises a site-specific recombination sequence (chromosomal landing pad).

13. The polynucleotide of claim 12, wherein the site-specific recombination sequence is a recognition sequence recognized by a phage integrase.

14. The polynucleotide of claim 13, wherein the phage integrase is phiC-31 integrase.

15. The polynucleotide of claim 13, wherein the recognition sequence is an attP site or an attB site recognized by the phage integrase.

16. The polynucleotide of claim 8, wherein the mammalian cell is a Chinese hamster ovary (CHO) cell.

17. The polynucleotide of claim 16, wherein the native chromosomal insertion site is at or close to positions 130-131 of SEQ ID NO: 1 for the Ank2 gene, positions 629-630 of SEQ ID NO: 2 for the Cpsf4 gene, positions 272-273 of SEQ ID NO: 3 for the C-Mos gene, positions 239-240 of SEQ ID NO: 4 for the Nephrocystin-1/Mal gene.

18. A vector comprising the recombinant polynucleotide of claim 8.

19. An engineered mammalian cell, comprising a heterologous polynucleotide which is stably integrated into its genome at one or more native chromosomal insertion sites located within or adjacent to a gene selected from the group consisting of ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene and Nephrocystin-1/Mal gene.

20. The cell of claim 19, wherein the native chromosomal insertion site supports stable integration of a foreign gene.

21. The cell of claim 19, wherein the heterologous polynucleotide encodes a polypeptide.

22. The cell of claim 21, wherein the polypeptide is a therapeutic protein or an industrial protein.

23. The cell of claim 19, wherein the heterologous polynucleotide comprises a site-specific recombination sequence (chromosomal landing pad).

24. The cell of claim 23, wherein the site-specific recombination sequence is a recognition sequence recognized by a phage integrase.

25. The cell of claim 24, wherein the phage integrase is phiC-31 integrase.

26. The cell of claim 24, wherein the recognition sequence is an attP site or an attB site recognized by the phage integrase.

27. The cell of claim 19, which is a Chinese hamster ovary (CHO) cell.

28. The cell of claim 27, wherein the heterologous polynucleotide is integrated at or close to positions 130-131 of SEQ ID NO: 1 for the Ank2 gene, positions 629-630 of SEQ ID NO: 2 for the Cpsf4 gene, positions 272-273 of SEQ ID NO: 3 for the C-Mos gene, positions 239-240 of SEQ ID NO: 4 for the Nephrocystin-1/Mal gene.

29. A method for stably integrating a heterologous polynucleotide into the genome of a mammalian cell, comprising:

(a) inserting a site-specific recombination sequence into the genome of the cell, wherein the insertion is at a native chromosomal insertion site located within or adjacent to a gene selected from the group consisting of ankyrin 2 gene (Ank2), cleavage and polyadenylation specific factor 4 gene (Cpsf4), C-Mos gene, and Nephrocystin-1/Mal gene.

(b) integrating by homologous recombination the heterologous polynucleotide into the genome of the cell at the inserted site-specific recombination sequence.

30. The method of claim 29, wherein the native chromosomal insertion site supports stable integration of a foreign gene.

31. The method of claim 29, wherein the site-specific recombination sequence is a first recognition sequence recognized by a phage integrase.

32. The method of claim 31, wherein the phage integrase is phiC-31 integrase.

33. The method of claim 31, wherein the first recognition sequence is an attP site or an attB site.

34. The method of claim 31, wherein the heterologous polynucleotide is attached to a second recognition sequence of the phage integrase which is cognate to the first recognition sequence.

35. The method of claim 34, wherein the second recognition sequence is an attB site or an attP site.

36. The method of claim 29, wherein the cell is a Chinese hamster ovary (CHO) cell.

37. The method of claim 36, wherein the site-specific recombination sequence is inserted into the genome at or close to positions 130-131 of SEQ ID NO: 1 for the Ank2 gene, positions 629-630 of SEQ ID NO: 2 for the Cpsf4 gene, positions 272-273 of SEQ ID NO: 3 for the C-Mos gene, positions 239-240 of SEQ ID NO: 4 for the Nephrocystin-1/Mal gene Mal gene.

38. The method of claim 29, wherein the heterologous polynucleotide comprises a target polypeptide-encoding sequence that is operably linked to a promoter sequence.

39. The method of claim 31, wherein the integration occurs in the presence of the phage integrase.

40. The method of claim 39, wherein the phage integrase is expressed from a vector introduced into the cell.