targeting and generation of cell lines expressing the GOI. The RTS insertion is generally carried out by RI or into a limited number of specific genomic regions, and thus the resulting cell lines are still subject to instability and reduced production over time. Moreover, SSI generally results in a low number of integrated gene copies that could indirectly limit rP production titres.

[0005] One method to increase integrated copies of recombinant genes is referred to as cumulative or accumulative SSI (see e.g., Kameyama et al. Biotechnol. Bioeng. 105:1106-14 (2010), Kawabe et al. Cytotechnology 64:267-79 (2012) and Turan et al. J. Mol. Biol. 402:52-69 (2010)). Such a method can include repeated rounds of RMCE to load up a single site sequentially with multiple copies of rP expression cassettes.

[0006] What are needed in the art are SSI cell lines that incorporate RTS at transcriptionally active and highly stable loci within the genome of the host cell. Such cell lines would be capable of stable and long-term expression of GOI.

[0007] Publications, patents, and patent applications are cited herein, the disclosures of which are incorporated by reference herein in their entireties.

SUMMARY

[0008] The present disclosure is based upon the recognition that the transcriptional output from a transgene insertion site as well as the stability of the expression system thereof will be strongly influenced by the 3-dimensional (3D) structure of the chromatin in that region. The present disclosure describes methods based on this recognition for determination of the structure and confirmation of a genome in 3 dimensions (3D mapping of a genome). The disclosed 3D mapping methods can be carried out through utilization of techniques such as, e.g., Hi-C and other chromosome conformation capture methods (Elzo de Wit and Wouter de Laat. Genes Dev. 2012 26: 11-24) and Promoter Capture Hi-C (Schoenfelder et al. Genome Res 25:582-97 (2015), among others. Methods of utilizing information obtained by the 3D mapping protocols as well as mammalian cells that can be formed by the methods are also described. This application teaches how to generate multi-level 3D genome maps and then use that information to identify optimal genome integration sites for the expression of heterologous genes. For example, by interrogating the mapped 3D genome structure, integration sites likely to exhibit high performance can be identified.

[0009] In one aspect, the present disclosure refers to a mammalian cell comprising a first recombination target site (RTS) chromosomally integrated at a first high integrating (HI) locus, the first HI locus being within an active genomic compartment of accessible chromatin and within about 30,000 base pairs of a topologically associated domain (TAD) boundary, the first HI locus overlapping a region of the cell genome that interacts with at least one enhancer element. In one example, the present disclosure describes a mammalian cell that includes an RTS at a high integrating (HI) locus. HI loci are high performance genomic sites identified by the inventors through analysis of the 3D hierarchical structure of genomic chromatin. Beneficially, HI loci are in stable, transcriptionally active environments of the

genome and can be repeatedly targeted to deliver predictable and stable levels of GOI expression.

[00010] HI loci can be within an active genomic compartment of accessible chromatin and can also be within about 30,000 base pairs of a topologically associated domain (TAD) boundary. In addition, HI loci can overlap regions of the genome that interact with at least one enhancer element. HI loci can vary depending on whether expression of the GOI will be driven by an in situ endogenous promoter or by a heterologous promoter. For instance, in those cell lines in which expression of the GOI is driven by an in situ endogenous promoter, HI loci can overlap and be downstream of a transcription start site (TSS). Moreover, in this example, HI loci can overlap an active, and in some examples, also fully annotated gene loci, e.g., an active gene the expression product of which or lack thereof is non-vital to the cell. In those cell lines in which expression of the GOI is driven by a heterologous promoter, HI loci can generally be external to active or non-transcribed gene loci. For example, HI loci in such a cell can encompass loci that do not overlap any associated promoter regions of active genes or in one example that do not come within about 1,000 base pairs of any active gene (e.g., within about 1,000 base pairs of any active and fully annotated gene).

[00011] In some examples, a cell can include multiple RTS, e.g., at least two RTS, at least four RTS, or even more in some examples. For instance, a cell can include multiple RTS in a single HI locus, in distinct HI loci, and/or in separate loci (e.g., the FerIL4 locus).

[00012] In some examples, an RTS can include an Frt site, a lox site, a rox site, or an att site. In some examples, an RTS can include a sequence selected from among SEQ ID Nos.: 126-155.

[00013] Cell types encompassed herein can include, without limitation, a mouse cell, a human cell, a Chinese hamster ovary (CHO) cell, a CHO-K1 cell, a CHO-DXB11 cell, a CHO-DG44 cell, a CHOK1SV™ cell including all variants, a CHO glutamine synthetase knockout cell including all variants, a HEK cell, a HEK293 cell including adherent and suspension-adapted variants, a HeLa cell, or a HT1080 cell.

[00014] In one example, a cell can include a GOI, e.g., a chromosomally integrated GOI such as a reporter gene, a selection gene, a gene of therapeutic interest, an ancillary gene, or a combination of genes. A GOI can encode a difficult to express (DtE) protein such as an Fc-fusion protein, an enzyme, a membrane receptor, or a monoclonal antibody (e.g., a bi-specific or a tri-specific monoclonal antibody). In one example, a GOI can be located between two RTS within a single HI locus. A cell can incorporate multiple GOI in some

examples. For instance, a cell can incorporate two or more GOI within a single HI locus, can incorporate multiple GOI, one or more of which being in different HI loci, and/or can incorporate multiple GOI in any combination of HI loci and separate loci. In some examples, a cell can incorporate a recombinase gene, for instance a site-specific recombinase gene that in one example can be chromosomally integrated.

[00015] In another aspect, the present disclosure refers to a method for producing a recombinant cell comprising: a. mapping peaks in accessible chromatin of a cell genome; b. identifying within the mapped peaks a first set of peaks within active genomic compartments of the accessible chromatin and also within about 30,000 base pairs of a topologically associated domain (TAD) boundary; c. defining within the first set of peaks a first high integrating (HI) locus, the first HI locus overlapping a region of the genome that interacts with at least one enhancer element; and d. inserting a first recombination target site (RTS) within the first HI locus. Also disclosed are methods for producing a recombinant cell. For instance, a method can include mapping peaks in accessible chromatin of a cell genome and identifying within the mapped peaks in accessible chromatin a first set of peaks that are within active genomic compartments of the accessible chromatin and also within about 30,000 base pairs of a topologically associated domain (TAD) boundary. In one example, the first set of peaks can be within active genomic compartments (for instance, as defined by Principle Component Analysis Methods (PCA)) and can also be within open chromatin (for instance, as defined by ATAC-seq), but this is not a requirement of a method, and in other examples, the first set of peaks can include those peaks that are within active genomic compartments within the whole of the mapped accessible chromatin. The method can also include identifying among the first set of peaks those that overlap regions of the genome that interact with at least one enhancer element. An HI locus can then be defined among the peaks that fit these criteria. Following identification of an HI locus, an RTS can be inserted into the HI locus. Optionally, a gene encoding a site-specific recombinase can also be inserted into the cell.

[00016] In those examples in which expression of a gene from the HI locus is to be driven by an in situ endogenous promoter, a method can further include identifying among the first set of peaks that overlap regions of the genome that interact with at least one enhancer element a second set of peaks that overlap a TSS, and in particular TSS for active genes the expression product of which or lack thereof is non-vital. The HI locus can be defined within this second set of peaks, the HI locus overlapping an active gene and being downstream of the TSS of the active gene.

[00017] In those examples in which expression of a gene from the HI locus is to be driven by a heterologous promoter, a method can further include identifying within the first set of peaks that overlap regions of the genome that interact with at least one enhancer element those peaks within accessible chromatin that do not overlap active genes or their associated promoter regions and an HI locus can be defined within this second set of peaks.

[00018] A method can also include transfecting the cell with a vector that includes an exchangeable cassette encoding a GOI and integrating the exchangeable cassette into an HI

locus. A cell that includes the exchangeable cassette integrated into the chromosome at an HI locus can then be selected as a recombinant protein producer cell.

[00019] Optionally, methods can include incorporating additional RTS into the cell. For instance, additional RTS can be incorporated into the same HI locus as the first RTS, into one or more additional HI loci, and/or into one or more separate loci.

[00020] In yet another aspect, the disclosure refers to a method for producing a recombinant cell comprising: a. mapping peaks in accessible chromatin of a cell genome; b. identifying within the mapped peaks a first set of peaks within active genomic compartments of the accessible chromatin and also within about 30,000 base pairs of a topologically associated domain (TAD) boundary; c. identifying within the accessible chromatin regions of the genome that interact with at least one enhancer element; d. defining within the first set of peaks a plurality of high integrating (HI) loci, each HI locus of the plurality overlapping an identified region; e. integrating a recombination target site (RTS) into a plurality of cells; and f. selecting from the plurality of cells a cell comprising the RTS integrated at an HI locus. According to another example, a method for producing a recombinant cell is disclosed that includes mapping peaks in accessible chromatin of a cell genome and identifying within the mapped peaks in accessible chromatin a first set of peaks that are within active genomic compartments of the accessible chromatin and also within about 30,000 base pairs of a topologically associated domain (TAD) boundary. In one example, the first set of peaks can be within active genomic compartments (for instance, as defined by Principle Component Analysis Methods (PCA)) and can also be within open chromatin (for instance, as defined by ATAC-seq), but this is not a requirement of a method, and in other examples, the first set of peaks can include those peaks that are within active genomic compartments within the whole of the mapped accessible chromatin. The method can also include identifying within the first set of peaks those that overlap regions of the genome that interact with at least one enhancer element. A plurality of HI loci can then be defined within the resulting set of mapped peaks. A method can further include integrating an RTS into a plurality of cells (e.g., according to an RI protocol), and then selecting from that plurality of cells a cell comprising the RTS integrated into an HI locus. Optionally, a gene encoding a site-specific recombinase can also be inserted into that selected cell.

[00021] In one example, the HI loci identified by the method can be ranked according to effectiveness. For instance, the HI loci can be ranked according to one or more of the expression level of one or more genes associated with each locus, the distance from each locus to the nearest TAD boundary, and the number of predicted enhancer interactions of each locus. In one such example, in which a cell is selected that includes the RTS integrated into an HI locus, the cell(s) can be selected according to the ranking of the HI locus insertions sites.

[00022] In one example, the method of defining the HI loci can also depend upon whether the HI loci are intended to be utilized to express a heterologous gene driven with an in situ endogenous promoter or a heterologous promoter. For instance, in those examples in which expression of genes from the HI loci is to be driven by an in situ endogenous promoter, a method can further include identifying within the resulting set of mapped peaks as defined above those peaks that overlap a TSS for active genes, such as an active gene the

expression product of which or lack thereof is non-vital. A second set of peaks can then be defined that overlap the identified genes and that are downstream of the TSS of these identified genes, and the HI loci can be defined within this second set of peaks.

[00023] In those examples in which expression of genes from the HI loci is to be driven by a heterologous promoter, a method can further include identifying within the resulting set of mapped peaks as defined above a second set of peaks that do not overlap any genes, e.g., any active genes, or their associated promoter regions and the HI loci can be defined within this second set of peaks.

[00024] A method can also include transfecting a selected cell that includes an RTS integrated into an HI locus with a vector that includes an exchangeable cassette encoding a GOI and integrating the exchangeable cassette into the HI locus. A cell that includes the exchangeable cassette integrated into the chromosome can then be selected as a recombinant protein producer cell.

[00025] Optionally, methods can include incorporating additional RTS into the cell. For instance, additional RTS can be incorporated into a first HI locus, into one or more additional HI loci, and/or into one or more separate loci.

BRIEF DESCRIPTION OF THE FIGURES

[00026] A full and enabling disclosure of the present subject matter, including the best mode thereof to one of ordinary skill in the art, is set forth more particularly in the remainder of the specification, including reference to the accompanying figure in which:

[00027] FIG. 1 presents a flow chart showing one embodiment of methods for production of a 3D map of a genome and utilization thereof to define and rank candidate HI loci. The diagram shows a summary of sequential filtering or screening process by which the data used to generate the multi-level 3D genome map can then be used to identify candidate HI loci.

[00028] FIG. 2A shows a section of the genome-wide Hi-C heatmap for data mapped to the LACHESIS assembly at a resolution of individual CHO-K1SV raw scaffolds. Only cis interactions are plotted and the smallest LACHESIS groups 7, 8 and 9 are not included because of visual clarity.

[00029] FIG. 2B shows a 100 % stacked bar chart displaying the average percentage of close cis (<10 kb), far cis (>10 kb) and trans unique, valid di-tags across CHO-K1SV 10E9 Hi-C replicates mapped to individual input CHO-K1SV scaffolds and the final LACHESIS assembly. For comparison, distributions of close cis, far cis and trans di-tags, averaged across replicates of equivalent Hi-C datasets derived from human embryonic stem cells and mouse

What is claimed:

1. A mammalian cell comprising a first recombination target site (RTS) chromosomally integrated at a first high integrating (HI) locus, the first HI locus being within an active genomic compartment of accessible chromatin and within about 30,000 base pairs of a topologically associated domain (TAD) boundary, the first HI locus overlapping a region of the cell genome that interacts with at least one enhancer element.

2. The cell of claim 1, wherein the first HI locus comprises one of SEQ ID NOs: 1-125 or is within or overlapping about 5,000 base pairs of either the 5' or 3' end of any one of SEQ ID NOs: 1-125.

3. The cell of claim 1 or claim 2, wherein the first HI locus overlaps a transcription start site (TSS) within the active genomic compartment, wherein the TSS is optionally operably linked to an active gene, the expression or the lack of expression of the active gene being non-vital to the mammalian cell, or wherein the first HI locus does not overlap a gene locus, or wherein the first HI locus does not overlap an in situ endogenous promoter of a gene locus, for instance wherein the first HI locus is not within about 1,000 base pairs of the promoter.

4. The cell of any one of claims 1 to 3, comprising a second distinct RTS and optionally comprising additional distinct RTS, wherein the first RTS and the second distinct RTS are chromosomally integrated within the first HI locus, or wherein the second distinct RTS is chromosomally integrated within a second HI locus, or wherein the second distinct RTS is chromosomally integrated at a separate locus, for instance wherein the separate locus is the Fer1L4 locus, and optionally, wherein at least one of the RTS is a frt site, a lox site, a rox site, or an att site, such as wherein at least one of the RTS comprises a sequence selected from among SEQ ID NOs: 126-155.

5. The cell of any one of claims 1 to 4, wherein the mammalian cell is a mouse cell, a human cell, a Chinese hamster ovary (CHO) cell, a CHO-K1 cell, a CHO-DXB11 cell, a CHO-DG44 cell, a CHOK1SV™ or variant thereof, a CHO glutamine synthetase knockout cell or variant thereof, a HEK cell, a HEK293 cell or an adherent or suspension-adapted variant thereof, a HeLa cell, or a HT1080 cell.

6. The cell of any one of claims 1 to 5, further comprising a first gene of interest, wherein the first gene of interest is chromosomally integrated, wherein the first gene of interest optionally comprises a reporter gene, a selection gene, a gene of therapeutic interest, an ancillary gene, or a combination thereof, such as wherein the gene of interest comprises a gene of therapeutic interest that comprises a gene encoding a difficult to express protein, for example a difficult to express protein selected from the group consisting of a Fc-fusion protein, an enzyme, a membrane receptor, or a monoclonal antibody, and further wherein the first gene of interest is optionally located between two RTS, or wherein the first gene of interest is located within the first HI locus.

7. The cell of any one of claims 1 to 6, further comprising a second gene of interest and optionally further comprising a third gene of interest, wherein the second gene of interest and the third gene of interest, when present, are chromosomally integrated and, wherein the second gene of interest is located within the first HI locus, or wherein the first gene of interest is located within the first HI locus and the second gene of interest is located within a second HI locus or within a separate locus, or wherein

a. at least one of the first gene of interest, the second gene of interest, and the third gene of interest is within the first HI locus and
b. at least one of the first gene of interest, the second gene of interest, and the third gene of interest is within a second HI locus.

8. The cell of any one of claims 1 to 7, further comprising a site-specific recombinase gene, and optionally wherein the site-specific recombinase gene is chromosomally integrated.

9. A method for producing a recombinant cell comprising:

a. mapping peaks in accessible chromatin of a cell genome;
b. identifying within the mapped peaks a first set of peaks within active genomic compartments of the accessible chromatin and also within about 30,000 base pairs of a topologically associated domain (TAD) boundary;
c. defining within the first set of peaks a first high integrating (HI) locus, the first HI locus overlapping a region of the genome that interacts with at least one enhancer element; and
d. inserting a first recombination target site (RTS) within the first HI locus.

10. A method for producing a recombinant cell comprising:

a. mapping peaks in accessible chromatin of a cell genome;
b. identifying within the mapped peaks a first set of peaks within active genomic compartments of the accessible chromatin and also within about 30,000 base pairs of a topologically associated domain (TAD) boundary;
c. identifying within the accessible chromatin regions of the genome that interact with at least one enhancer element;
d. defining within the first set of peaks a plurality of high integrating (HI) loci, each HI locus of the plurality overlapping an identified region;
e. integrating a recombination target site (RTS) into a plurality of cells; and
f. selecting from the plurality of cells a cell comprising the RTS integrated at an HI locus.

11. The method of claim 9 or 10, wherein the HI loci comprise one of SEQ ID NOs: 1-125 or are within or overlapping about 5,000 base pairs of either the 5' or 3' end of any one of SEQ ID NOs: 1-125.

12. The method of any one of claims 9 to 11, further comprising identifying within the first set of peaks those peaks that overlap any transcription start site (TSS) for a gene, the expression product of which or lack thereof is non-vital, and defining a second set of peaks that overlap the genes and are downstream of the TSS, wherein the HI loci are defined within the second set of peaks.

13. The method of any one or claims 9 to 11, further comprising identifying within the first set of peaks a third set of peaks that that do not overlap any genes, wherein the HI loci are defined within the third set of peaks.

14. The method of any one of claims 9 to 13, further comprising transfecting the cell of claim 9 or a plurality of the cells of claim 10 with a first vector comprising an exchangeable cassette encoding a first gene of interest and integrating the first exchangeable cassette within an HI locus of the cell, the method further comprising selecting a recombinant protein producer cell comprising the first exchangeable cassette integrated into the chromosome.

15. The method of any one of claims 9 to 14, further comprising ranking the HI loci, for instance wherein the HI loci are ranked according to one or more of expression level of one or more genes associated with each locus, distance from each locus to the nearest TAD boundary, number of predicted enhancer interactions at each locus, and expression level of mRNA of one or more genes associated with each locus.