

CLAIMS

1. A method for the selection of a sequence optimized nucleic acid sequence from a plurality of nucleic acid sequence variants, wherein said sequence optimized nucleic acid sequence corresponds to a eukaryotic cell with a defined phenotype, said method comprising:
- 5 i) Providing a population of isolated eukaryotic cells, each cell comprising a pre-defined genomic location, which pre-defined genomic location comprises:
- a. a nucleic acid sequence I1 comprising a recognition site for a first DNA enzyme;
- 10 b. a nucleic acid sequence E1 comprising a recognition site for a second DNA enzyme; and
- c. a promotor nucleic acid sequence;
- ii) Providing a plurality of donor vectors, each donor vector comprising:
- a. a nucleic acid sequence I2;
- 15 b. a nucleic acid sequence E2 comprising a recognition site for said second DNA enzyme;
- c. a nucleic acid sequence encoding a first selection marker; and
- d. a nucleic acid sequence region comprising a nucleic acid sequence variant,
- 20 iii) Contacting the plurality of donor vectors with the population of cells in the presence of a first DNA enzyme, wherein the presence of the first DNA enzyme enables recombination between the nucleic acid sequence I2 of a donor vector and the nucleic acid sequence I1 present in the pre-defined genomic location of a cell;
- iv) Selecting and isolating a cell having a donor vector integrated at the pre-defined genomic location by detecting the expression of the first selection marker in the cell, wherein the expression of the first selection marker is activated by the promotor nucleic acid sequence at the pre-defined genomic location; and
- 25 v) Selecting and isolating a cell with a defined phenotype from the cells of step iv), thereby selecting and isolating a sequence optimized nucleic acid sequence from said nucleic acid sequence variants, said sequence optimized nucleic acid sequence corresponding to said defined phenotype.
- 30 2. The method of claim 1, wherein each of the plurality of donor vectors of step ii) further comprises:
- e. an expression cassette encoding a second selection marker; and
- 35 wherein said method further comprises the steps of:

- 5 vi) Excising a nucleic acid sequence flanked by the nucleic acid sequences E1 and E2 from a cell of step iii) or selected and isolated in step iv) or step v) in the presence of a second DNA enzyme, wherein the presence of the second DNA enzyme enables recombination between the nucleic acid sequences E1 and E2, wherein the presence of an expression cassette encoding a second selection marker that is not flanked by the nucleic acid sequences E1 and E2 in a cell is indicative of a stable integration of a donor vector at a genomic location outside the pre-defined genomic location of said cell; and
- 10 vii) Selecting and isolating a cell from step vi) lacking an expression cassette encoding a second selection marker.
3. The method of claim 1 or 2, further comprising obtaining nucleic acid sequence information from the isolated cell of step iv), v) or vii) by sequencing of the sequence optimized nucleic acid sequence present at the pre-defined genomic location of the isolated cell.
- 15 4. The method of any one of claims 1 to 3, wherein a nucleic acid sequence variant of the plurality of donor vectors constitutes a variant of a promotor, a variant of an intron, a variant of a transcription regulatory sequence, a variant of a DNA structure regulatory sequence, a variant of a 5' untranslated region, a variant of a 3' untranslated region, a variant of an internal ribosome entry site, a variant of a gene of interest, a variant of a nucleic acid sequence encoding a signal peptide and/or any combination of such variant.
- 20 5. The method of any one of the preceding claims, wherein the selection of a sequence optimized nucleic acid sequence by the selection of a cell with a defined phenotype in step v) of claim 1 comprises selection based on one or more of the following phenotypic characteristics of said cell:
- 25 (i) The presence or expression level of an endogenous biomolecule in said cell;
(ii) The expression level of a recombinant protein of interest by said cell;
(iii) The growth rate of said cell; and/or
(iv) The functionality of a recombinant protein of interest encoded by said nucleic acid sequence region in said cell.
- 30 6. The method of claim 5, wherein said recombinant protein of interest is a recombinant fusion protein, such as a protein fused to a membrane anchor domain for localization at said cell surface and/or is a protein fused to a fluorescent protein or a fluorescent protein domain.
- 35 7. The method of claim 5, wherein said endogenous biomolecule constitutes a protein, an mRNA, an miRNA, an lncRNA or a metabolite.

8. The method of any one of claims 5 to 6, wherein said functionality of a recombinant protein of interest is measured and determined based on an interaction between said recombinant protein of interest when localized and expressed at the cell surface of said cell, and a target structure, such as a small molecule, a DNA molecule, an RNA molecule, a protein, a protein complex such as a virus particle, an exosome or a cell, optionally wherein said target structure is tagged with a fluorescent moiety.
9. The method of any one of claims 5 to 8, wherein said recombinant protein of interest is an affinity protein candidate and wherein the level of expression is determined by display of said affinity protein candidate on the cell surface of said cell.
10. The method of claim 9, wherein said affinity protein candidate is a single chain polypeptide fused to a membrane anchor domain, optionally wherein said single chain polypeptide is selected from the group consisting of a Z-scaffold protein, a Nanobody scaffold protein, a single chain fragment variable (scfv) scaffold protein, a Fynomer scaffold protein, a DARPin scaffold protein and/or an adnectin scaffold protein.
11. The method of claim 9, wherein said affinity protein candidate comprises two or more polypeptide chains, such as an antibody variant, and wherein a nucleic acid sequence variant corresponding to said affinity protein candidate encodes an affinity protein candidate variant, such as an antibody variant, and wherein one of said two or more polypeptides is fused to a membrane anchor domain.
12. The method of any one of claims 9 to 11, further comprising determining the binding specificity, selectivity, affinity and/or functionality of said affinity protein candidate by providing to said affinity protein candidate a specific target component, optionally labelled with a fluorescent marker, to which the affinity protein candidate is exposed and thereafter detecting binding of said affinity protein candidate to said specific target component.
13. The method of claim 12, wherein said target component is selected from a small molecule, a DNA molecule, an RNA molecule, a protein, a protein complex, such as a virus particle, an exosome or a cell, optionally wherein said target structure is tagged with a fluorescent moiety.
14. The method of any of claims 5 to 13, wherein said expression level or functionality of a recombinant protein of interest and/or said presence or level of an endogenous biomolecule is measured at the level of a single cell, such as by using Flow Cytometry.
15. The method of any one of the preceding claims, wherein said nucleic acid sequence region of a donor vector comprises a nucleic acid sequence variant for expression of one or several recombinant proteins of interest from said donor vector, wherein said

plurality of donor vectors comprises different nucleic acid sequence variants encoding different amino acid sequence variants of said one or several recombinant proteins of interest.

- 5 16. The method of any one of claims 1 to 14, wherein said nucleic acid sequence region of a donor vector comprises a nucleic acid sequence variant for expression of a recombinant protein of interest from said donor vector, wherein nucleic acid sequence variants present in a plurality of donor vectors comprises nucleic acid sequence variants encoding essentially identical amino acid sequence variants of said recombinant protein of interest.
- 10 17. The method of any one of claims 1 to 14 comprising a plurality of donor vectors comprising a nucleic acid sequence region comprising essentially identical nucleic acid sequences to encode essentially identical recombinant proteins of interest but wherein said nucleic acid sequence region comprises different nucleic acid sequence variants of donor vector components, such as nucleic acid sequence variants of
- 15 promoter or enhancer nucleic acid sequences of said donor vector.
18. The method of any of the preceding claims, wherein said method identifies sequence optimized nucleic acid donor vector components for use in recombinant protein expression systems based on a eukaryotic cell system.
- 20 19. A sequence optimized nucleic acid sequence selected by a method of any one of the preceding claims.
20. Use of a sequence optimized nucleic acid sequence of claim 19, for producing a recombinant protein.
21. Use of a sequence optimized nucleic acid sequence of claim 19, for designing further sequence optimized nucleic acid sequences.
- 25 22. An isolated eukaryotic cell with a defined phenotype corresponding to a sequence optimized nucleic acid sequence obtainable by the method of any one of the preceding claims.