[12] Accordingly, it is one object of the present invention to provide a library which encodes protein fragments/peptides wherein such a library can be used in screening methods including but not limited to PPI screens. In other objects, the present invention provides alternative, improved, simpler, cheaper and/or integrated means or methods that address one or more of these or other problems. An object underlying the present invention is solved by the subject matter as disclosed or defined anywhere herein, for example by the subject matter of the attached claims.

Summary

[12a] In an aspect of the present disclosure, there is provided a library of nucleic acids, each nucleic acid comprising a coding region of defined nucleic acid sequence encoding for a peptide having a length of between 25 and 110 amino acids, and having an amino acid sequence being a region of a sequence selected from the amino acid sequence of a naturally occurring protein of one or more organisms; wherein the library comprises nucleic acids that encode for a plurality of at least 10,000 different such peptides, and wherein the amino acid sequence of each of at least 50 of such peptides is a sequence region of the amino acid sequence of a different protein of a plurality of different such naturally occurring proteins, and wherein each peptide encoded by the library is predicted from its amino acid sequence to have an isoelectric point (pI) of greater than 8.0 or less than 6.0.

[12b] In another aspect of the present disclosure, there is provided a library of peptides encoded by the library of nucleic acids disclosed herein.

[12c] In a further aspect of the present disclosure, there is provided a method of identifying a target protein that modulates a phenotype of a mammalian cell, said method comprising:

a. exposing a population of in vitro cultured mammalian cells capable of displaying said phenotype to a library of nucleic acids disclosed herein or a library of peptides disclosed herein,

b. identifying in said cell population an alteration in said phenotype following said exposure,

c. selection of said cells undergoing the phenotypic change and identifying a peptide encoded by (or a peptide of) such library that alters the phenotype of the cell,

d. providing said peptide and identifying the cellular protein that binds to said peptide, said cellular protein being a target protein that modulates the phenotype of the mammalian cell;

optionally wherein the method includes a further step of identifying a compound that binds to said target protein and displaces or blocks binding of said peptide, wherein the compound modulates the phenotype of a mammalian cell.

[12d] In another aspect of the present disclosure, there is provided a use of: (a) a library of nucleic acids disclosed herein; and/or (b) a library of peptides disclosed herein, to identify a peptide that binds to a target; optionally

wherein the target is a protein target; and/or optionally

wherein the identified peptide modulates a phenotype of a mammalian cell**.**

[12e] In a further aspect of the present disclosure, there is provided a use of: (a) a library of nucleic acids disclosed herein; and/or (b) a library of peptides disclosed herein, to identify a compound which binds to a target; optionally

wherein the target is a protein target; and/or optionally

wherein said compound displaces or blocks binding of a peptide to the target; and/or optionally

wherein, the peptide and/or the compound modulates a phenotype of a mammalian cell.

[12f] In another aspect of the present disclosure, there is provided a method disclosed herein, or use disclosed herein, wherein the phenotype is a phenotype related to the modulation of a cell-signalling pathway; optionally

wherein the method or use comprises the identification of peptides which modulate cell-signalling pathways and the identification of protein targets and surface sites on such proteins that participate in signal transduction; and/or optionally

wherein the cell-signalling pathway is active or altered in cancer cells.

Detailed Description

[13] The figures show:

[14] **Figure 1**: depicts a screen for SEPs expressed in the HuPEx library that are able to overcome 6-thioguanine toxicity. Cells carrying a library of inserts that express SEPs are treated with 500nM 6-thioguanine for 6 days. Enrichment between 6-thioguanine treatment (n=3) and DMSO control (n=3) is shown.

[15] **Figure 2**: depicts a screen for SEPs across all three libraries (HuPEx, BugPEx, OmePEx) for SEPs able to selectively kill cells lacking the PTEN tumour suppressor gene.

[16] **Figure 3**: depicts the effect on MCF10A PTEN KO cells of a peptide (7-924) identified from the screen described in Example 4(2) expressed by pMOST25 compared to empty vector and a positive control (shRNA against NLK).

[17] **Figure 4**: depicts the experimental principle of the MNNG-induced parthantos phenotypic screen of Examples 9 and 10.

[18] **Figure 5**: depicts the relative abundance of DNA sequences encoding for SEPs from the HuPEx library that are present in a control aliquot of HuPEX-expressing HeLa cells before (D0) treatment with 6.7uM MNNG, and in a treatment aliquot of such HuPEX-expressing HeLa cells after 8 days (D8) of such MNNG treatment. Peptides showing a significantly increased relative abundance at D8 are marked by triangles.

[19] **Figure 6**: depicts the relative abundance of DNA sequences encoding for SEPs from the BugPEx library that are present in a control aliquot of BugPEx expressing HeLa cells before (D0) treatment with 6.7uM MNNG, and in a treatment aliquot of such BugPEx expressing HeLa cells after 8 days (D8) of such MNNG treatment. Axes are as for Figure 5, and peptides showing a significantly increased relative abundance at D8 are marked by triangles.

[20] **Figure 7**: depicts the relative abundance of DNA sequences encoding for SEPs from the OmePEx library that are present in a control aliquot of OmePEx-expressing HeLa cells before (D0) treatment with 6.7uM MNNG, and in a treatment aliquot of such OmePEx-expressing HeLa cells after 8 days (D8) of such MNNG treatment. Axes are as for Figure 5, and peptides showing a significantly increased relative abundance at D8 are marked by triangles.

[21] **Figure 8 and Figure 9**: Phenotypic screen of PEx libraries for autophagy induction

[22] **Figure 8 (A-C):** HEK293FT Cells were engineered to stably express an GFP-LC3/RFP-LC3DG autophagy reporter (Kaizuka et al Molecular Cell 2016). Subsequently, autophagy reporter cells were infected with the pooled HuPEx (HPX), BugPEx (BPX) and OmePEx (OPX) libraries and selected on puromycin. After selection, cells enriched in the low GFP-LC3 gate, compared to unsorted controls, were flow-sorted and peptide sequences were amplified and sent to NGS analysis as described previously. The graphs (Figure 8A (HPX), Figure 8B (BPX) and Figure 8C (OPX)) show the population of selected hits (autophagy inducers) in the marked region compared to control.

[23] **Figure 9:** Hits were re-run individually in a flow-cytometry experiment after infection with lentivirus carrying either control sequences or putative hits. A selection of candidates is shown with BPX-497507 representing a strong and robust hit able to induce autophagy as measured by GFP-LC3 reduction. Torin1 (250nM) is shown as a positive control.

[24] The present invention, and particular non-limiting aspects and/or embodiments thereof, can be described in

CLAIMS

1. **A library of nucleic acids**, each nucleic acid comprising a coding region of defined nucleic acid sequence encoding for a peptide having a length of between 25 and 110 amino acids, and having an amino acid sequence being a region of a sequence selected from the amino acid sequence of a naturally occurring protein of one or more organisms; wherein the library comprises nucleic acids that encode for a plurality of at least 10,000 different such peptides, and wherein the amino acid sequence of each of at least 50 of such peptides is a sequence region of the amino acid sequence of a different protein of a plurality of different such naturally occurring proteins, and wherein each peptide encoded by the library is predicted from its amino acid sequence to have an isoelectric point (pI) of greater than 8.0 or less than 6.0.

2. The library of nucleic acids of claim 1, wherein each of the plurality of different naturally occurring proteins fulfils one or more pre-determined criteria.

3. The library of nucleic acids of claim 2, wherein either

i) each of the plurality of naturally occurring proteins is associated with a given disease, such as cancer; optionally wherein the disease is breast cancer; or

ii) each of the plurality of naturally occurring proteins is a cytoplasmic protein; optionally wherein each of the plurality of naturally occurring proteins is a cytoplasmic kinase; or

iii) each of the plurality of naturally occurring proteins interacts with a given protein or at least one protein from a (functional) class of proteins; optionally wherein each of the plurality of naturally occurring proteins interacts with KRas.

4. The library of nucleic acids of any one of claims 1 to 3, wherein the library comprises nucleic acids that encode for a plurality of at least 50,000 different such peptides, and wherein the amino acid sequence of each of at least 100 of such peptide is a sequence region of the amino acid sequence of at least 100 different naturally occurring proteins; in particular wherein the library comprises nucleic acids that encode for a plurality of at least 100,000 different such peptides, and wherein the amino acid sequence of each of at least 150 of such peptide is a sequence region of the amino acid sequence of at least 150 different naturally occurring proteins.

5. The library of nucleic acids of any one of claims 1 to 4, wherein the library comprises nucleic acids that encode for a plurality of at least 10,000 different such peptides, and wherein the amino acid sequence of each of at least 1,000 of such peptides is a sequence region of the amino acid sequence of a different protein of such plurality of different naturally occurring proteins.

6. The library of nucleic acids of any one of claims 1 to 5, wherein the library comprises nucleic acids that encode for a plurality of at least 200,000 different such peptides, and wherein the amino acid sequence of each of at least 20,000 of such peptide is a sequence region of the amino acid sequence of at least 20,000 different naturally occurring proteins; in particular wherein the library comprises nucleic acids that encode for a plurality of at least 300,000 different such peptides, and wherein the amino acid sequence of each of at least 25,000 of such peptide is a sequence region of the amino acid sequence of at least 25,000 different naturally occurring proteins.

7. The library of nucleic acids of any one of claims 1 to 6, wherein that in respect of at least about 1% of the naturally occurring proteins a plurality of the nucleic acids encodes for different peptides from the amino acid sequences of such naturally occurring proteins; optionally

wherein that in respect of at least about 50% of the naturally occurring proteins a plurality of the nucleic acids encodes for different peptides from the amino acid sequences of such naturally occurring proteins; and optionally, wherein the plurality of the nucleic acids encodes for different peptides, and the amino acid sequences of which are sequence regions spaced along the amino acid sequence of the naturally occurring protein;

and optionally wherein the sequence regions are spaced by a window of amino acids apart, or by multiples of such window, along the amino acid sequence of the naturally occurring protein wherein, the window is between 1 and about 55 amino acids; in particular wherein the window is between about 5 and about 20 amino acids; most particularly wherein the window of spacing is about 8, 10, 12 or 15 amino acids.

8. The library of nucleic acids of any one of claims 1 to 7 comprising nucleic acids encoding for at least 100,000 different peptides from at least 10,000 different naturally occurring proteins.

9. The library of nucleic acids of any one of claims 1 to 8, wherein each nucleic acid encodes a different peptide.

10. The library of nucleic acids of any one of claims 1 to 9, wherein the mean number of nucleic acids that encode a different peptide from the naturally occurring proteins is greater than 1; in particular between about 1.01 and 1.5 such nucleic acids (peptides) per such protein; optionally

wherein the mean number of nucleic acids that encode a different peptide from the naturally occurring proteins is at least about 5 such nucleic acids (peptides) per such protein, in particular wherein the mean is between about 5 and about 2,000 such nucleic acids (peptides) per such protein or is between about 5 and about 1,000 nucleic acids (peptides) per such protein; and optionally

wherein the mean number of nucleic acids that encode a different peptide from the naturally occurring proteins is between about 100 and about 1,500 such nucleic acids (peptides) per such protein or is between about 250 and about 1,000 such nucleic acids (peptides) per such protein; or

wherein the mean number of nucleic acids that encode a different peptide from the naturally occurring proteins is between about 5 and about 100 such nucleic acids (peptides) per such protein or is between about 5 and about 50 such nucleic acids (peptides) per such protein.

11. The library of nucleic acids of any one of claims 1 to 10, wherein the amino acid sequence of the naturally occurring protein is one selected from the group of amino acids sequences of non-redundant proteins comprised in a reference proteome, suitably, the reference proteome is one or more of the reference proteomes selected from the group of reference proteomes listed in Table A and/or Table B, or an updated version of such reference proteome.

12. The library of nucleic acids of any one of claims 1 to 11, wherein the amino acid sequences of the plurality of encoded peptides are sequence regions selected from amino acid sequences of naturally occurring proteins (or polypeptide chains or domains thereof) with a known three-dimensional structure; in particular wherein the naturally occurring protein (or polypeptide chain or domain thereof) is comprised in the Protein Data Bank, and optionally that has a Pfam annotation; optionally

wherein the sequence region selected from the amino acid sequence of the protein does not include an ambiguous amino acid of such amino acid sequence comprised in the reference proteome or the Protein Data Bank.

13. The library of nucleic acids according to any preceding claim, wherein the library is for expression in a mammalian cell, preferably a human cell; and/or optionally wherein the library is cloned into a lentiviral vector or a retroviral vector.

14. **A library of peptides** encoded by the library of nucleic acids of any one of claims 1 to 13.

15. **A method of identifying a target protein that modulates a phenotype of a mammalian cell**, said method comprising:

a. exposing a population of in vitro cultured mammalian cells capable of displaying said phenotype to a library of nucleic acids according to any of claims 1-13 or a library of peptides according to claim 14,

b. identifying in said cell population an alteration in said phenotype following said exposure,

c. selection of said cells undergoing the phenotypic change and identifying a peptide encoded by (or a peptide of) such library that alters the phenotype of the cell,

d. providing said peptide and identifying the cellular protein that binds to said peptide, said cellular protein being a target protein that modulates the phenotype of the mammalian cell;

optionally wherein the method includes a further step of identifying a compound that binds to said target protein and displaces or blocks binding of said peptide, wherein the compound modulates the phenotype of a mammalian cell.

16. **Use** of: (a) a library of nucleic acids according to any of claims 1-13; and/or (b) a library of peptides according to claim 14, **to identify a peptide that binds to a target**; optionally
wherein the target is a protein target; and/or optionally

wherein the identified peptide modulates a phenotype of a mammalian cell**.**

17. **Use** of: (a) a library of nucleic acids according to any of claims 1-13; and/or (b) a library of peptides according to claim 14, **to identify a compound** which binds to a target; optionally

wherein the target is a protein target; and/or optionally

wherein said compound displaces or blocks binding of a peptide to the target; and/or optionally

wherein, the peptide and/or the compound modulates a phenotype of a mammalian cell.

18. A method according to claim 15, or use according to claim 16 or claim 17, wherein the phenotype is a phenotype related to the modulation of a cell-signalling pathway; optionally

wherein the method or use comprises the identification of peptides which modulate cell-signalling pathways and the identification of protein targets and surface sites on such proteins that participate in signal transduction; and/or optionally

wherein the cell-signalling pathway is active or altered in cancer cells.