(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 January 2014 (23.01.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/015227 A1

(51) International Patent Classification:
C12N 15/79 (2006.01) *C07K 19/00* (2006.01)
C12N 15/65 (2006.01) *C12N 15/62* (2006.01)
C12N 15/85 (2006.01)

(21) International Application Number:
PCT/US2013/051232

(22) International Filing Date:
19 July 2013 (19.07.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/673,399 19 July 2012 (19.07.2012) US

(71) Applicant: **THE SCRIPPS RESEARCH INSTITUTE** [US/US]; 10550 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors; and
(71) Applicants : **PENG, Yingjie** [CN/US]; 5150 Balboa Arms Drive, Apt. G12, San Diego, CA 92117 (US). **LERNER, Richard** [US/US]; 7750 Roseland Drive, La Jolla, CA 92037 (US).

(74) Agent: **FITTING, Thomas**; The Scripps Research Institute, 10550 North Torrey Pines Road, TPC-8, La Jolla, California 92037 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

[Continued on next page]

(54) Title: ENGINEERING CELL SURFACES FOR ORTHOGONAL SELECTABILITY

(57) Abstract: The invention provides methods and compositions for endowing eukaryotic cell surfaces with orthogonal selection markers.

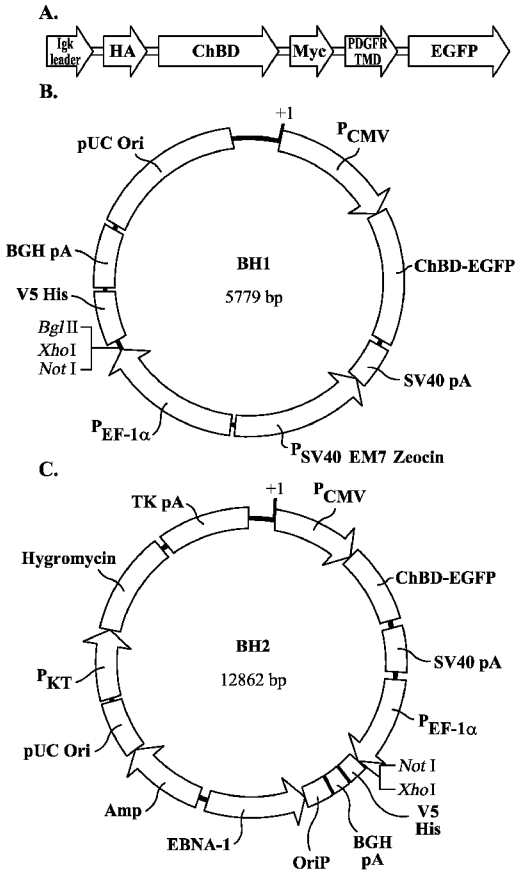

FIG. 1

WO 2014/015227 A1

BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

Engineering Cell Surfaces for Orthogonal Selectability

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The subject patent application claims the benefit of priority to U.S. Provisional Patent Application Number 61/673,399 (filed July 19, 2012). The full disclosure of the priority application is incorporated herein by reference in its entirety and for all purposes.

BACKGROUND OF THE INVENTION

[0002] The selective addition of markers to expressed proteins has become a standard procedure in molecular biology. The most frequently used markers are fluorescent proteins and epitope tags. Fluorescent markers such as green fluorescent protein (GFP) allows confirmation that the protein to which it is appended is expressed and enables study of the sub-cellular localization of the protein by fluorescent microscopy as well as isolation of the cells expressing the protein by fluorescent cell sorting. Similarly, in protein tagging, an antigen for which one has a cognate antibody or a peptide that has affinity for a ligand such as nickel is appended to a protein. It is used to study the localization of a given protein in cells and allows identification of the molecule after analytical procedures and isolation by affinity chromatography. Thus, at the level of individual proteins the artisans have a choice of methods to study the expression patterns of the marked protein and even isolate the cells expressing it by cell sorting. However, there is a need in the art for more robust and cost-effective genetic methods for marking the surface of the cells. The present invention addresses this and other needs by providing novel methods and compositions useful for cellular selections.

SUMMARY OF THE INVENTION

[0003] In one aspect, the invention provides methods for engineering a surface selection marker on eukaryotic cells. The methods entail expressing on the surface of the cells a polypeptide comprising a chitin-binding domain (ChBD) or cellulase-binding domain (CBD). Typically, the polypeptide selection marker is expressed from a vector introduced into the cell. In some embodiments, the expressed polypeptide additionally contains a eukaryotic transmembrane domain fused to the ChBD or CBD. For example, the expressed surface polypeptide can contain platelet-derived growth factor receptor (PDGFR). In various

embodiments, the expressed polypeptide can further harbor an enhanced green fluorescent protein (EGFP). In some embodiments, the methods employ a ChBD that is obtained from *Bacillus circulans* WL-12 chitinase A1 or ChBD from *Vibrio harveyi* chitinase A. Some methods of the invention are directed to conferring selectability to mammalian cells.

[0004] In a related aspect, the invention provides engineered or recombinant eukaryotic cells (e.g., mammalian cells) which have been modified to provide a surface marker for cellular selection. Typically, the surface marker is a heterologous polypeptide comprising a chitin-binding domain (ChBD) or cellulase-binding domain (CBD). In some preferred embodiments, the heterologous polypeptide is expressed from a vector introduced into the cell. In some embodiments, the heterologous polypeptide further comprises a eukaryotic transmembrane domain that is fused to the ChBD or CBD. For example, the ChBD or CBD can be fused to platelet-derived growth factor receptor (PDGFR). In some embodiments, the polypeptide can additionally comprise an enhanced green fluorescent protein (EGFP) that is fused to the ChBD. In some engineered cells of the invention, the heterologous surface selection marker comprises a ChBD that is obtained from *Bacillus circulans* WL-12 chitinase or ChBD from *Vibrio harveyi* chitinase A.

[0005] In another related aspect, the invention provides isolated or recombinant polypeptides that contain a chitin-binding domain (ChBD) fused to a transmembrane domain of a eukaryotic cell. For example, the polypeptide can have a ChBD obtained from *Bacillus circulans* WL-12 chitinase or *Vibrio harveyi* chitinase A. In some embodiments, the transmembrane domain is from platelet-derived growth factor receptor (PDGFR). Polynucleotides encoding the isolated or recombinant polypeptides are also provided in the invention. In other aspects, the invention provides expression vectors comprising the polynucleotides, host cells harboring the vectors, as well as transgenic non-human animals which have integrated such vectors.

[0006] 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

[0007] Figures 1A-1C are schematic diagrams of ChBD-EGFP construct (A), BH1 vector (B), and BH2, Epstein Barr virus based vector (C). The construct used the IgK leader which is the murine Ig kappa-chain V-J2-C signal peptide and included a gene encoding the Hemagglutinin A (HA) epitope before the N-terminus of the ChBD. The ChBD is from

Bacillus circulans WL-12; PDGFR TMD is the platelet-derived growth factor receptor transmembrane domain; EGFP is the enhanced green fluorescent protein. A gene encoding the Myc epitope is located between the ChBD and the transmembrane domain.

[0008] Figures 2A-2B show cell surface expression of ChBD-EGFP in BH1 transfected HEK 293T cells allows attachment of cells to chitin-coated beads. A: confocal images show that no ChBD-EGFP is expressed in mock-transfected HEK 293T cells (left upper), while ChBD is uniformly expressed in BH1 transfected HEK 293T cells (right upper). The ChBD co-localizes with wheat germ agglutinin that marks the external surface of the plasma membrane (left bottom and right bottom). B: after selection, non ChBD-EGFP expressing cells did not bind to chitin-coated beads (left), while ChBD-EGFP expressing cells could specifically bind to the chitin-coated beads (right).

[0009] Figure 3 shows cell membrane purification facilitated by ChBD. Cell membrane fragments from ChBD-EGFP transfected HEK 293T cells (293T-CG) were studied. Antibodies against two cell membrane marker proteins, α1 sodium/potassium ATPase and pan Cadherin were used to mark membrane fractions attached to chitin beads (B), the total cellular homogenate (H), and flow-through fraction (FT). As a control, cell membrane fragments from mock-transfected HEK 293T cells (293T) were studied. Antibody against β Actin was used to monitor the efficiency of the membrane purification.

[0010] Figure 4 shows expression of EGFP in mock (left), pBudCE4.1-EGFP (middle), and BH1 (right)-transfected HEK 293T cells.

DETAILED DESCRIPTION OF THE INVENTION

[0011] The invention relates to methods and compositions which enable efficient and robust selection of modified cells from a population of cells. Methods of the invention can find wide uses not only for selection of cells but also for isolating cells in vitro and in vivo in any situation where one wishes to trace the fate and ultimately recover a minor population of cells. As exemplification, the present invention was employed for specific recovery of transformed cells and isolation of plasma membrane fragments. With the methods, certain cells can be easily and simply isolated from a population where they may be a minor component. Amongst other uses, methods of the invention find various important applications, e.g., in enriching for transformants, especially when the transformation frequency is low. The methods can facilitate combinatorial antibody selections of phage that bind to cell surfaces where, as for transformants, the target cell can be a minor component of

an otherwise large population. Also, such a method would allow for rapid affinity based isolation of plasma membranes for biochemical studies.

[0012] Methods of the invention typically entail engineered expression on the surface of eukaryotic cells of a heterologous ligand binding domain obtained from some unusual enzymes such as chitin-binding domain (ChBD), thereby conferring orthogonal selectability to the target cells. In some embodiments, the heterologous ligand binding domain is expressed as a fusion protein with a transmembrane domain on the cell surface. As exemplification, a cellular selection system for affinity selection of cells and plasma membranes was generated via expression of the chitin-binding domain (ChBD) as a fusion with a transmembrane domain on the surface of eukaryotic cells. Similar selection systems can be based on engineered expression on the cell surface of other ligand binding domains, e.g., or cellulase binding domain (CBD). Conservatively modified variants of various known ChBDs or CBDs, as well as variant polypeptides with substantially identical sequences, can also be used in the practice of the present invention.

[0013] The enzymes chitinase and cellulase hydrolyze chitin and cellulose respectively, which are the two most abundant insoluble polymers on earth. These enzymes are unusual because, in addition to the canonical enzyme format where the active site has affinity for the substrate, they each have an appended small domain that binds to their insoluble substrates. In the case of the chitin-binding domain, this small appendage has nearly covalent binding energy for the substrate, which is a readily abundant and cheap insoluble polymer of N-acetyl glucosamine. By expressing ChBD or CBD on the eukaryotic cell surface, the cells would have affinity for insoluble polymers that are not present in animal cells.

[0014] Transmembrane domain usually denotes a single transmembrane alpha helix of a transmembrane protein. It is called a "domain" because an alpha-helix in a membrane can fold independently from the rest of the protein, similar to domains of water-soluble proteins. More broadly, a transmembrane domain is any three-dimensional protein structure which is thermodynamically stable in a membrane. This may be a single alpha helix, a stable complex of several transmembrane alpha helices, a transmembrane beta barrel, a beta-helix of gramicidin A, or any other structure. Depending on the target eukaryotic cells to be modified for conferring selectability, various eukaryotic transmembrane domain polypeptides can be used in the practice of the present invention. As exemplified herein, the transmembrane domain of platelet-derived growth factor receptor (PDGFR) is suitable for the invention.

[0015] 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 invention pertains. The following references provide one of skill with a general definition of many of the terms used in this invention: *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). In addition, the following definitions are provided to assist the reader in the practice of the invention.

[0016] The singular terms "a," "an," and "the" include plural referents unless the context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise.

[0017] As used herein, the term "amino acid" of a peptide refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an a carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid.

[0018] Cellulase refers to a suite of enzymes produced chiefly by fungi, bacteria, and protozoans that catalyze cellulolysis (i.e. the hydrolysis of cellulose). However, there are also cellulases produced by a few other types of organisms, such as some termites and the microbial intestinal symbionts of other termites. Several different kinds of cellulases are

known, which differ structurally and mechanistically. Cellulases hydrolyze the 1,4-beta-D-glycosidic linkages in cellulose, lichenin and cereal beta-D-glucans.

[0019] Chitinases are hydrolytic enzymes that break down glycosidic bonds in chitin. As chitin is a component of the cell walls of fungi and exoskeletal elements of some animals (including worms and arthropods), chitinases are generally found in organisms that either need to reshape their own chitin or dissolve and digest the chitin of fungi or animals. Chitinivorous organisms include many bacteria (e.g., Aeromonads, Bacillus, Vibrio), which may be pathogenic or detritivorous. They attack living arthropods, zooplankton or fungi or they may degrade the remains of these organisms. Fungi, such as Coccidioides immitis, also possess degradative chitinases related to their role as detritivores and also to their potential as arthropod pathogens. Chitinases are also present in plants (e.g., barley seed chitinase).

[0020] As used herein the term "comprising" or "comprises" is used in reference to compositions, methods, and respective component(s) thereof, that are essential to the invention, yet open to the inclusion of unspecified elements, whether essential or not.

[0021] As used herein the term "consisting essentially of" refers to those elements required for a given embodiment. The term permits the presence of elements that do not materially affect the basic and novel or functional characteristic(s) of that embodiment of the invention.

[0022] The term "consisting of" refers to compositions, methods, and respective components thereof as described herein, which are exclusive of any element not recited in that description of the embodiment.

[0023] The term "conservatively modified variant" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon

in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid that encodes a polypeptide is implicit in each described sequence.

[0024] For polypeptide sequences, "conservatively modified variants" refer to a variant which has conservative amino acid substitutions, amino acid residues replaced with other amino acid residue having a side chain with a similar charge. Families of amino acid residues having side chains with similar charges have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine).

[0025] The term "engineered cell" or "recombinant host cell" (or simply "host cell") refers to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

[0026] The term "isolated" means a reference molecule (e.g., a polypeptide or a polynucleotide) is removed from its natural surroundings. However, some of the components found with it may continue to be with an "isolated" molecule. Thus, an "isolated polypeptide" is not as it appears in nature but may be substantially less than 100% pure protein.

[0027] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same. Two sequences are "substantially identical" if two sequences have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity over a specified region, or, when not specified, over the entire sequence), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of

the following sequence comparison algorithms or by manual alignment and visual inspection. Optionally, the identity exists over a region that is at least about 50 nucleotides (or 10 amino acids) in length, or more preferably over a region that is 100 to 500 or 1000 or more nucleotides (or 20, 50, 200 or more amino acids) in length.

[0028] Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman, Adv. Appl. Math. 2:482c, 1970; by the homology alignment algorithm of Needleman and Wunsch, J. Mol. Biol. 48:443, 1970; by the search for similarity method of Pearson and Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444, 1988; by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, Madison, WI); or by manual alignment and visual inspection (see, e.g., Brent et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (ringbou ed., 2003)). Two examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., Nuc. Acids Res. 25:3389-3402, 1977; and Altschul et al., J. Mol. Biol. 215:403-410, 1990, respectively.

[0029] Other than percentage of sequence identity noted above, another indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

[0030] The term "operably linked" refers to a functional relationship between two or more polynucleotide (e.g., DNA) segments. Typically, it refers to the functional relationship of a transcriptional regulatory sequence to a transcribed sequence. For example, a promoter or enhancer sequence is operably linked to a coding sequence if it stimulates or modulates the transcription of the coding sequence in an appropriate host cell or other expression system. Generally, promoter transcriptional regulatory sequences that are operably linked to a transcribed sequence are physically contiguous to the transcribed sequence, i.e., they are cis-

acting. However, some transcriptional regulatory sequences, such as enhancers, need not be physically contiguous or located in close proximity to the coding sequences whose transcription they enhance.

[0031] A "transgenic animal" refers to any animal, preferably a non-human mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA.

[0032] The term "vector" is intended to refer to a polynucleotide molecule capable of transporting another polynucleotide to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors").

[0033] Expressing a ChBD- or CBD-containing fusion polypeptide on cell surface as selection marker can be carried out in accordance with the exemplification provided herein and other methods well known in the art. Chitin-binding domains and cellulase-binding domains from various sources are known and well characterized in the art. See, e.g., Watanabe et al., J. Bacteriol. 176:4465-72, 1994; Hashimoto et al., J. Bacteriol. 182: 3045-3054, 2000; Svitil et al., Microbiol. 144:1299-1308; 1998; Tjoelker et al., J. Biol. Chem. 275:514-520, 2000; Akagi et al., J. Biochem. 139: 483-493, 2006; Linder et al., Proc. Natl. Acad. Sci. USA 93:12251-5, 1996; and Tomme et al., in "Enzymatic Degradation of Insoluble Carbohydrates", Chapter 10, pp 142–163 (ACS Symposium Series, Vol. 618; May

1996). Vectors for expressing the fusion polypeptide can be generated in accordance with the techniques described herein and methods routinely practiced in the art, e.g., as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, N.Y., (3rd ed., 2000); and Brent et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (ringbou ed., 2003).

[0034] Unless otherwise stated, other procedures or steps required for practicing the present invention can be based on standard procedures as described, e.g., in *Methods in Enzymology*, Volume 289: Solid-Phase Peptide Synthesis, J. N. Abelson, M. I. Simon, G. B. Fields (Editors), Academic Press; 1st edition (1997) (ISBN-13: 978-0121821906); U.S. Pat. Nos. 4,965,343, and 5,849,954; Sambrook et al., *supra*; Brent et al., *supra*; White et al., *PCR Protocols*, Academic Press, Inc. (1990); Freshney, *Culture of Animal Cells*, Wiley-Liss (2005); Murray et al., *Gene Transfer and Expression Protocols*, The Humana Press Inc. (1991); Davis et al., *Basic Methods in Molecular Biology*, Elsevier Science Publishing, Inc., New York, USA (1986); or *Methods in Enzymology: Guide to Molecular Cloning Techniques* Vol. 152, S. L. Berger and A. R. Kimmerl Eds., Academic Press Inc., San Diego, USA (1987); *Current Protocols in Protein Science* (CPPS) (John E. Coligan, et. al., ed., John Wiley and Sons, Inc.), *Current Protocols in Cell Biology* (CPCB) (Juan S. Bonifacino et. al. ed., John Wiley and Sons, Inc.), and *Culture of Animal Cells: A Manual of Basic Technique* by R. Ian Freshney, Publisher: Wiley-Liss; 5th edition (2005), *Animal Cell Culture Methods* (Methods in Cell Biology, Vol. 57, Jennie P. Mather and David Barnes editors, Academic Press, 1st edition, 1998).

[0035] As exemplification, vectors expressing the ChBD linked to EGFP on the cell surface were constructed (Figure 1A). One vector (BH1) was constructed based on the pBudCE4.1 vector (Invitrogen). The ChBD-EGFP cassette was inserted into the cloning site under the control of the CMV promoter. The second cloning site under the control of the EF-1a promoter can be used to express another gene of interest (Figure 1B). The other vector (BH2) was constructed based on the pCEP4 vector (Invitrogen). The CMV-ChBD-EGFP and EF-1a promoter cassettes from BH1 were inserted before the OriP of pCEP4 (Figure 1C).

[0036] To determine whether the transformation with the plasmid encoding the ChBD was efficient and was expressed on the cell surface in a functional form, the transformed cells were tested for their ability to both express GFP and bind to chitin beads. EGFP alone and EGFP linked to the ChBD were efficiently expressed in HEK 293T cells (Figure 4). The ChBD expression was confined to the cell surface (Figure 2A). Cells expressing the ChBD-

EGFP encoded constructs bound efficiently to the chitin beads whereas cells that only expressed EGFP did not (Figure 2B).

[0037] In addition to selection of intact cells that express the ChBD, the methods of the invention are also useful for purification of plasma membranes for biochemical studies. As exemplification, HEK 293T cells that expressed the ChBD on their cell surface were sonicated and the membrane fragments were collected on chitin beads. Gel electrophoresis and western blotting studies showed that this procedure afforded highly purified membrane preparations (Figure 3).

[0038] The methods of the invention allow selection of eukaryotic cells by engineered expression on the cell of a specific ligand binding domain for selectability. Importantly, the binding functionality that is used does not have a counterpart in animal cells thereby endowing the modified cell surface with a property that is orthogonal to animal cells. The methods of the present invention are advantageous over other selection methods currently known in the art, e.g., using antibodies targeting cell surface proteins or the avidin-biotin system. While antibodies might be an alternative, they generally do not have the affinity of the ChBD and are much more expensive to use. The avidin-biotin system is also inferior because avidin is a tetramer and biotin is an essential vitamin such that the ligand is already present in all cells. Thus, even if the avidin tetramer could be expressed in functional form on the cell surface, it would already be saturated with ligand. Indeed, the avidin-biotin system has been tried and failed in eukaryotic cells. Finally, fluorescent activated cell sorting, while useful in some selection settings, is a method that requires expensive instrumentation as opposed to the "Bench Top" procedure described herein.

[0039] In addition to the methods described herein for conferring to cells surface markers for orthogonal selectability, isolated or recombinantly expressed ChBD- or CBD-containing fusion polypeptides are also encompassed by the present invention. The invention additionally provides isolated or substantially purified polynucleotides (DNA or RNA) which encode the fusion polypeptides described herein. Expression vectors and engineered host cells harboring the vectors for expressing polynucleotides encoding the polypeptides are also provided in the invention. The polynucleotide encoding the fusion polypeptide is typically operably linked to a promoter in the expression vectors. As exemplified herein, the expression construct can further encode a signal peptide for translocation of the expressed polypeptide across cell membrane. The expression vectors of the invention are not subject to any particular limitation, and may be, for example, bacteriophages, plasmids, cosmids or

phagemids. Examples of recombinant bacteriophage or phagemid vectors include that based on a filamentous phage such as M13. Plasmid vectors include those based on plasmids from, e.g., E. coli (e.g., pBR322, pBR325, pUC118 and pUC119), plasmids from Bacillus subtilis (e.g., pUB110 and pTP5), and plasmids from yeasts (e.g., YEp13, YEp24 and YCp50). The expression vectors can also include animal viruses such as retroviruses, vaccinia viruses and insect viruses (e.g., baculoviruses). Similarly, the host cells to which the vectors are introduced can be any of a variety of expression host cells well known in the art, e.g., bacteria (e.g., E. coli), yeast cell, or animal cells such as CHO, COS or 293 cells. In some preferred embodiments, the host cells are recombinant animal cells (e.g., mammalian cells) which have integrated a polynucleotide encoding the ChBD- or CBD-containing polypeptide and express the polypeptide on the cell surface.

[0040] Methods of the invention find wide uses not only for selection of cells but also for isolating cells in vitro and in vivo in any situation where one wishes to trace the fate and ultimately recover a minor population of cells. Toward this end, the invention further provides transgenic non-human animals such as mouse strains where the ChBD (or CBD) is either constitutively expressed in all cells or inducible in certain cell lineages. This should allow long-term studies on cell fates, where the likely immunogenicity of the ChBD would otherwise be a problem in cell transfer experiments. Thus, cells that constitutively express the ChBD can be followed in other mouse strains where expression of the ChBD is repressed by, for example, tetracycline inactivation of a sensitive tetracycline-dependent transactivator.

EXAMPLES

[0041] The following examples are provided to further illustrate the invention but not to limit its scope.

Example 1 Materials and methods

[0042] Some specific techniques and protocols for practicing methods of the invention are exemplified below.

[0043] Cell culture: Human embryonic kidney HEK 293T cells were purchased from ATCC, and maintained in DMEM (GIBCO) supplemented with 10% fetal calf serum (GIBCO) and non-essential amino acids (GIBCO) at 37°C in a humidified CO_2 incubator. HEK 293T cells were transfected with X-tremeGENE 9 DNA transfection reagent (Roche).

[0044] Plasmid construction: ChBD-EGFP was cloned using overlapping PCR. First, the fragment extending from IgK leader to the PDGFR TMD was obtained using the following primers:

CMV-CBD-5: ACGCGTCGACATGGAGACAGACACACTCCTGCTATG (SEQ ID NO:1)

CMV-CBD-A:

ATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGGTTCCAGGTTCCA

CTGGTGACTATCCATATGATGTTCCAGATTATGCTG (SEQ ID NO:2)

CMV-CBD-B:

GCTGTGTTGACCTGCCAAGCGGATACACCAGGATTTGTCGTGGCCGGCTGGGCCA

AGGCTCCAGCATAATCTGGAACATCATATGGATAGTC (SEQ ID NO:3)

CMV-CBD-C:

GCTTGGCAGGTCAACACAGCTTATACTGCGGGACAATTGGTCACATATAACGGC

AAGACGTATAAATGTTTGCAGCCTCACACATCATTGGCAG (SEQ ID NO:4)

CMV-CBD-D:

GAGATGAGCTTCTGTTCGAGGCCTCGGGGGCCTTGAAGCTGCCACAAGGCAGGA

ACGTTGGATGGTTCCCATCCTGCCAATGATGTGTGAGGCTG (SEQ ID NO:5)

CMV-CBD-E:

GGCCTCGAACAGAAGCTCATCTCAGAAGAGGATCTGAATGCTGTGGGCCAGGAC

ACGCAGGAGGTCATCGTGGTGCCACACTCCTTGCCCTTTAAGG (SEQ ID NO:6)

CMV-CBD-F:

CATGATGAGGATGATAAGGGAGATGATGGTGAGCACCACCAGGGCCAGGATGGC

TGAGATCACCACCACCTTAAAGGGCAAGGAGTGTGGC (SEQ ID NO:7)

CMV-CBD-3:

CGCGGATCCCTAACGTGGCTTCTTCTGCCAAAGCATGATGAGGATGATAAGGGA

GATGATG (SEQ ID NO:8)

The EGFP fragment was amplified from the pEGFP (Clontech) using the following primers:

EF-GFP-5:

CCGTTTCGAACCAGACGTCTGGGCCGCCACCATGGTGAGCAAGGGCGAGGAG

(SEQ ID NO:9)

EF-GFP-3:

CCGTTTCGAACTCCAGCATGCTGGTCTTGTACAGCTCGTCCATGCCGAG (SEQ ID

NO:10)

To generate vector pBudCE4.1-EGFP, this fragment was inserted into *Bst*B I sites of pBudCE4.1. Thereafter, ChBD and the EGFP genes were fused using the following primers:

CG-5:

ACGCGTCGACATGGAGACAGACACACTCCTGCTATGGGTAC (SEQ ID NO:11)

CG-CP-F:

ATGGAGACAGACACACTCCTGCTATGGGTAC (SEQ ID NO:12)

CG-CP-R:

GCTCCTCGCCCTTGCTCACCATGCTGCCTCCTGCAGCGGCCGCTCCGGAACGTGG

CTTCTTCTGCCAAAGCATG (SEQ ID NO:13)

CG-EGFP-F:

ATGGTGAGCAAGGGCGAGGAGC (SEQ ID NO:14)

CG-EGFP-R:

CTTGTACAGCTCGTCCATGCCGAGAG (SEQ ID NO:15)

CG-3:

CGCGGATCCCTACTTGTACAGCTCGTCCATGCCGAGAG (SEQ ID NO:16)

To generate vector BH1, the ChBD-EGFP fusion was inserted into *Sal* I and *Bam*H I sites of pBudCE4.1 (Invitrogen). Then the *Not*I site in EGFP was deleted using QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent) with the primer CGTTCCGGAGCAGCCGCTGCAGG (SEQ ID NO:17). To generate vector BH2, ChBD-

EGFP and the CMV promoter were amplified from BH1 and inserted into *Bsr*G I and *Bam*H I sites of pCEP4 (Invitrogen). Then the EF-1α cassette was amplified from BH1 and inserted into the *Xba* I and *Pci* I sites.

[0045] Immunofluorescence using confocal microscopy: BH1 and mock transfected HEK 293T cells growing on collagen-coated coverslips were fixed in 10% formaldehyde before staining. The cell membranes were stained with 5 μg/ml Alexa Fluor 555 conjugated wheat germ agglutinin (Invitrogen). The nuclei were stained with 5 μg/ml Hoechst 34580 (Invitrogen). Confocal fluorescent images were obtained using a Bio-Rad (Zeiss) Radiance 2100 Rainbow laser scanning confocal microscope attached to a Nikon TE2000-U microscope with infinity corrected optics with a 60x objective. After sequential excitation, blue, green and red fluorescent images of the same cell were collected. Images were analyzed with ImageJ software. The term co-localization refers to the coincidence of green and red fluorescence, as measured by the confocal microscope.

[0046] Cell binding: BH1 or pBudCE4.1-EGFP transfected HEK 293T cells were mixed with 10-fold non-transfected HEK 293T cells and then incubated with magnetic chitin-coated beads (New England Biolabs) at room temperature for 1 hr. Then the beads were separated with a magnet and washed 5 times with PBS at room temperature. An aliquot of beads was transferred to slides and observed by fluorescent microscopy.

[0047] Cell membrane purification: BH1 or mock transfected HEK 293T cells were harvested, washed 3 times with cold PBS, and then sonicated in 600 μl PBS on ice. The homogenate was incubated with magnetic chitin-coated beads blocked with BSA. The beads were separated with magnet and the unbound fraction (flow-through) was collected. The beads were washed 8 times with PBS at room temperature, and resuspended in 600 μl PBS. The homogenate, flow-through and bead fractions were treated with SDS-PAGE loading buffer, then loaded 10 μl for western blotting. Antibodies against α1 sodium/potassium ATPase (ab7671, Abcam) and pan Cadherin (ab22744, Abcam) were used together with secondary HRP-conjugated goat-anti-mouse IgG (1030-05, Southern Biotech). HRP-conjugated antibody against β Actin was purchased from Cell Signaling.

Example 2 Cell surface expression of ChBD for affinity selection

[0048] One vector (BH1) was constructed based on the pBudCE4.1 vector (Invitrogen). The ChBD-EGFP cassette was inserted into the cloning site under the control of the

Cytomegalovirus (CMV) promoter. The second cloning site that is under the control of the EF-1α promoter can be used to express another gene of interest. The other vector (BH2) was constructed based on the pCEP4 vector (Invitrogen). The CMV-ChBD-EGFP and EF-1α promoter cassettes from BH1 were inserted before the OriP of pCEP4. The construct is anchored to the cell membrane using the PDGF gene product that is a protein that spans the plasma membrane. This membrane-spanning domain sits between the ChBD protein that is expressed on the outer surface of the cell and the EGFP protein that is present on the cytoplasmic side of the membrane. Since the EGFP protein is fluorescent it allows for the rapid determination of the efficiency of a given transformation.

[0049] To determine whether the transformation with the plasmid encoding the ChBD was efficient and was expressed on the cell surface in a functional form, the transformed cells were tested for their ability to both express GFP and bind to chitin beads. EGFP alone and EGFP linked to the ChBD were efficiently expressed. The ChBD expression was confined to the cell surface. When the cells expressing the ChBD domain on their cell surface were mixed with chitin beads for thirty minutes, the cells bound to the beads whereas cells that only expressed EGFP did not. In this experiment, the surface of the chitin beads appears to be saturated with cells that express the ChBD. The absolute number of cells binding to the beads is a function of the ratio of beads to cells. In the pictured experiment, excess cells were used solely for ease of visualization. After washing, cells expressing the ChBD-EGFP encoded constructs retained their affinity for the chitin beads. To determine the enrichment power of the method, different ratios of transformed to non-transformed cells were studied to determine the ability to recover rare cells. Transformed cells could be recovered when they represented only 0.5% of the population. Usually, after transformations a higher percentage of transformed cells are present. Thus, the method should be general and in many cases avoid the, often long, post-transformation enrichment procedures such as antibiotic selections.

[0050] It was essential to determine if the cells remained viable after they bound to the chitin beads. To show this, cells were bound to the chitin beads and then the complex was thoroughly washed and placed back into culture. The cells that bound initially were easily visible. After 4 days in culture new growth of cells was evident. Thus, the cells that bound to the beads were capable of replication that qualitatively appeared unimpeded. After the beads were saturated with cells, the newly replicated cells were no longer attached to the beads. Thus, a culture could be established from the viable progeny of cells that were originally attached to the beads.

[0051] In addition to selection of intact cells that express the ChBD, the method should be useful for purification of plasma membranes for biochemical studies. To test this possibility, HEK 293T cells that expressed the ChBD on their cell surface were sonicated and the membrane fragments were collected on chitin beads. Gel electrophoresis and western blotting studies showed that this procedure afforded highly purified membrane preparations. Beta-actin was used as a control to show that there is no contamination in membranes that were isolated based on their affinity for chitin coated beads. Actin is present in large amounts in the cell cytoplasm and if the purification were not stringent this protein would show up as a contaminant in the purified membranes.

***

[0052] 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.

[0053] 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 engineering a surface selection marker on a eukaryotic cell, comprising expressing on the surface of the cell a polypeptide comprising a chitin-binding domain (ChBD) or cellulase-binding domain (CBD).

2. The method of claim 1, wherein the polypeptide is expressed from a vector introduced into the cell.

3. The method of claim 1, wherein the polypeptide further comprises a eukaryotic transmembrane domain that is fused to the ChBD or CBD.

4. The method of claim 3, wherein the transmembrane domain is from platelet-derived growth factor receptor (PDGFR).

5. The method of claim 1, wherein the polypeptide further comprises an enhanced green fluorescent protein (EGFP) that is fused to the ChBD.

6. The method of claim 1, wherein the ChBD is ChBD from *Bacillus circulans* WL-12 chitinase or ChBD from *Vibrio harveyi* chitinase A.

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

8. An engineered or recombinant eukaryotic cell comprising on its surface a heterologous polypeptide comprising a chitin-binding domain (ChBD) or cellulase-binding domain (CBD).

9. The cell of claim 8, wherein the polypeptide is expressed from a vector introduced into the cell.

10. The cell of claim 8, wherein the polypeptide further comprises a eukaryotic transmembrane domain that is fused to the ChBD or CBD.

11. The cell of claim 10, wherein the transmembrane domain is from platelet-derived growth factor receptor (PDGFR).

12. The cell of claim 8, wherein the polypeptide further comprises an enhanced green fluorescent protein (EGFP) that is fused to the ChBD.

13. The cell of claim 8, wherein the ChBD is ChBD from *Bacillus circulans* WL-12 chitinase or ChBD from *Vibrio harveyi* chitinase A.

14. The cell of claim 8, wherein the cell is a mammalian cell.

15. An isolated or recombinant polypeptide comprising a chitin-binding domain (ChBD) fused to a transmembrane domain of a eukaryotic cell.

16. The polypeptide of claim 15, wherein the ChBD is ChBD from *Bacillus circulans* WL-12 chitinase or ChBD from *Vibrio harveyi* chitinase A.

17. The polypeptide of claim 15, wherein the transmembrane domain is from platelet-derived growth factor receptor (PDGFR).

18. A polynucleotide encoding the polypeptide of claim 15.

19. A vector comprising the polynucleotide of claim 18.

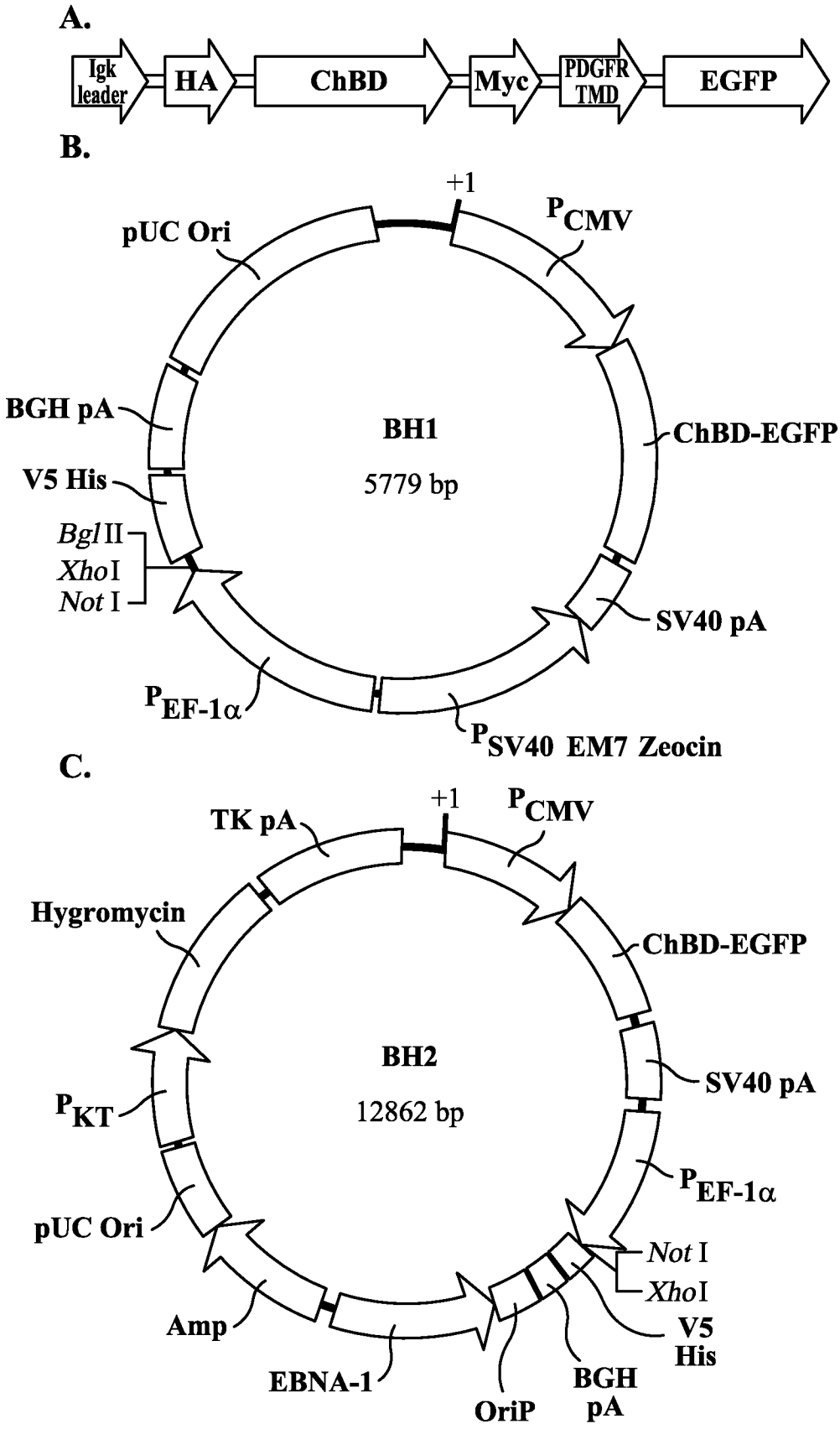
A.
Igk leader
HA
ChBD
Myc
PDGFR TMD
EGFP
B.
+1
pUC Ori
P CMV
BGH pA
BH1
5779 bp
ChBD-EGFP
V5 His
Bgl II
Xho I
Not I
SV40 pA
P EF-1α
P SV40 EM7 Zeocin
C.
+1
P CMV
TK pA
Hygromycin
ChBD-EGFP
BH2
12862 bp
SV40 pA
P KT
P EF-1α
pUC Ori
Not I
Xho I
V5 His
Amp
EBNA-1
OriP
BGH pA

FIG. 1

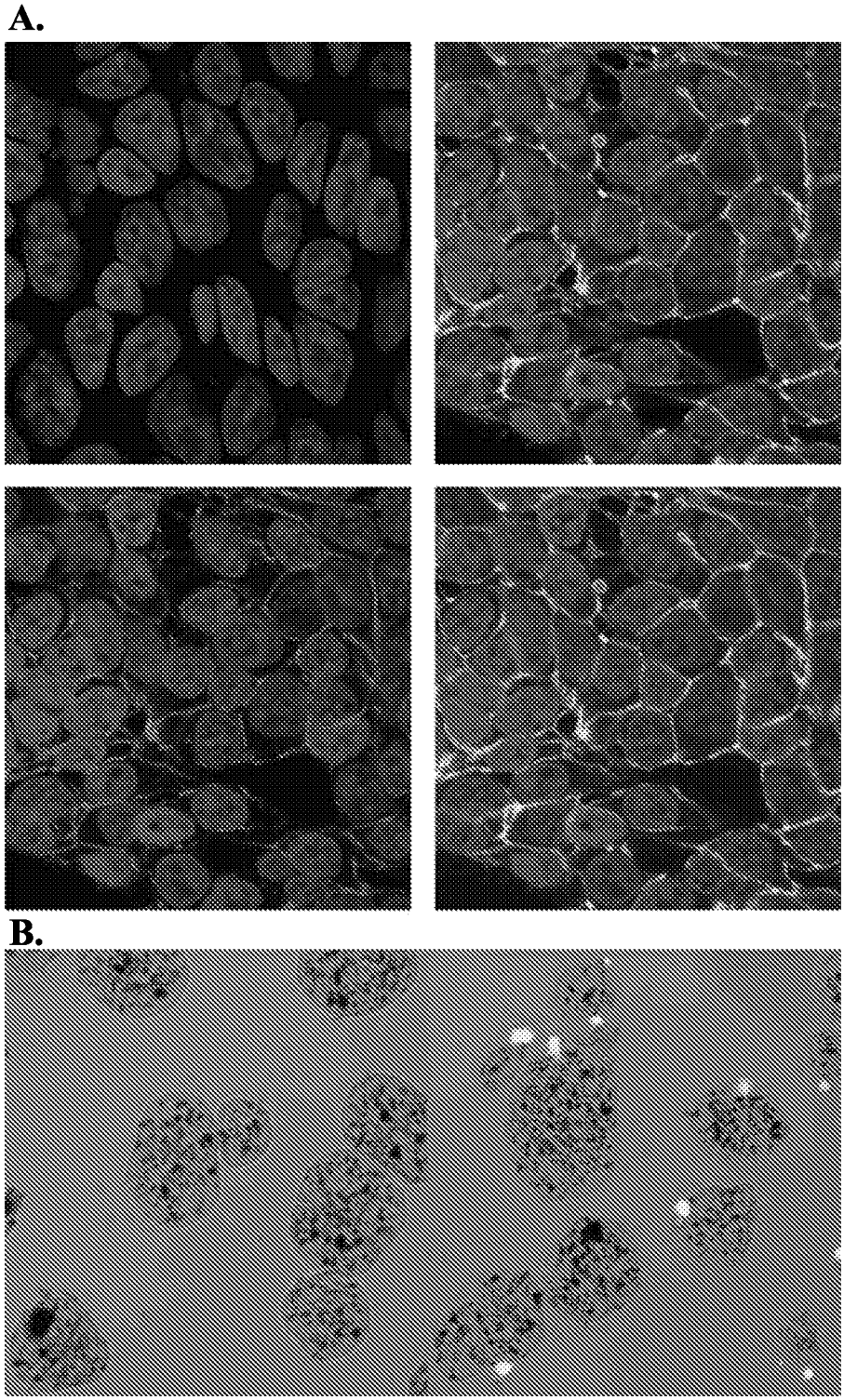

FIG. 2

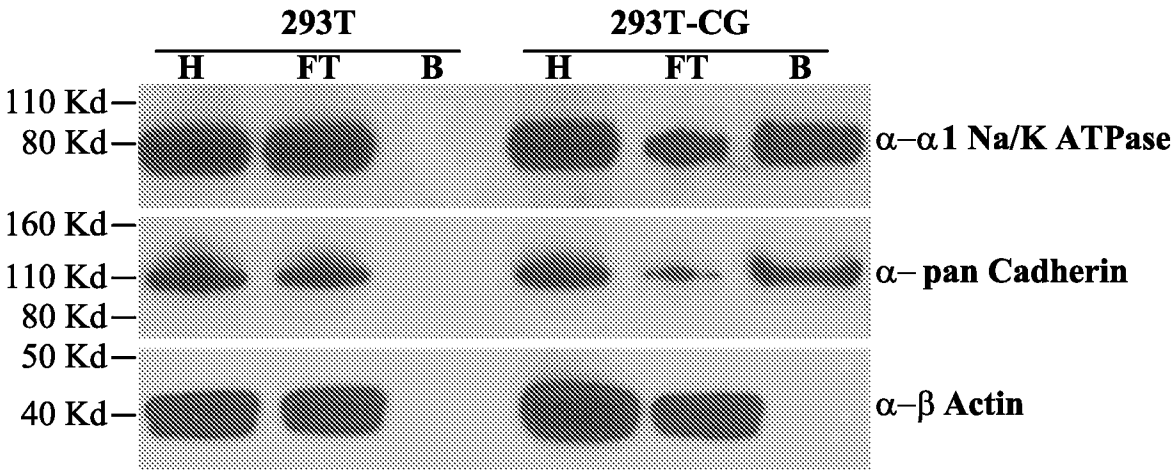
293T
293T-CG
H
FT
B
H
FT
B
110 Kd
80 Kd
α–α1 Na/K ATPase
160 Kd
110 Kd
80 Kd
α– pan Cadherin
50 Kd
40 Kd
α–β Actin

FIG. 3

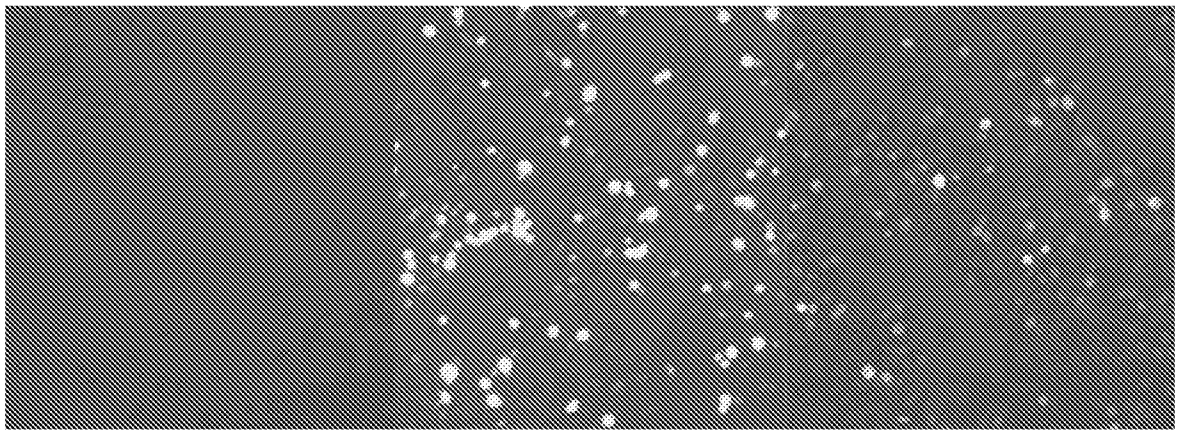

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/051232

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/79(2006.01)i, C12N 15/65(2006.01)i, C12N 15/85(2006.01)i, C07K 19/00(2006.01)i, C12N 15/62(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N 15/79; C12N 5/10; C12N 15/12; C12N 15/65; C12N 15/85; C07K 19/00; C12N 15/62

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: eukaryotic cell, cell surface, selectable marker, chitin-binding domain, chbd, cellulase binding domain, cbd

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WANG, JEN-YOU et al., `Immobilization of cells with surface-displayed chitin-binding domain` Applied and environmental microbiology, January 2006, Vol.72, No. 1, pp.927-931 See abstract; p.927, right column, lines 16-20, 33-36.	1-19
A	WANG, AIJUN A. et al., `Whole-cell immobilization using cell surface-exposed cellulose-binding domain` Biotechnology progress, 07 April 2001, Vol.17, Issue.3, pp.407-411 See abstract; p.407, right column, lines 29-39.	1-19
A	GOTOH, HIDEO et al., `Cell-surface streptavidin fusion protein for rapid selection of transfected mammalian cells` Gene, 03 November 2006, Vol.389, No.2, pp.146-153 See abstract; p.150, right column, [presence of strptavidin fusion protein o s surfaces of mammalian cells].	1-19
A	HAN, HUAMIN et al., `An efficient vector system to modify cells genetically` Plos one, 11 November 2011, Vol.6, Issue.11, p.1-10 See abstract.	1-19

☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 September 2013 (27.09.2013)	**27 September 2013 (27.09.2013)**

Name and mailing address of the ISA/KR
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 302-701, Republic of Korea
Facsimile No. +82-42-472-7140

Authorized officer
HEO Joo Hyung
Telephone No. +82-42-481-8150

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/051232

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WANG, XIULI et al., `A transgene-encoded cell surface polypeptide for selection, in vivo tracking, and ablation of engineered cells` Blood, 07 June 2011, Vol.118, No.5, pp.1255-1263 See abstract.	1-19
A	WO 91-03554 A1 (GENENTECH, INC.) 21 March 1991 See claims 1, 26, 27 and 30-34.	1-19
PX	PENG, YINGJIE et al., `Engineering cell surfaces for orthogonal selectability` Angewandte chemie international edition, 13 December 2012, Vol.52, Issue.1, pp.336-340 See the whole document.	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2013/051232

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 91-03554 A1	21/03/1991	AU 1991-32090 A	08/04/1991