



- (51) International Patent Classification:
C12Q 1/68 (2006.01)
- (21) International Application Number:
PCT/US2013/054250
- (22) International Filing Date:
9 August 2013 (09.08.2013)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/683,690 15 August 2012 (15.08.2012) US
- (71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [—/US]; 17 Quincy Street, Cambridge, 02138 (US).
- (72) Inventors: **CHURCH, George, M.**; 218 Kent Street, Brookline, MA 02446 (US). **MALI, Prashant**; 88 Beacon Street, Apt. 54, Somerville, MA 02143 (US).
- (74) Agent: **IWANICKI, John, P.**; 28 State Street, Suite 1800, Boston, MA 02109 (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, 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, SA, 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:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: POLYNUCLEOTIDE-BINDING DOMAINS AS A MEANS OF CELL LABELING, CELL ORGANIZATION AND POLYMER SEQUENCING

(57) Abstract: The present invention relates to a protein comprising a polynucleotide-binding domain and a transmembrane domain. These proteins can be used in methods of sequencing polynucleotides, methods of labeling polynucleotides, and methods of organizing cells (e.g., to construct tissues).



POLYNUCLEOTIDE-BINDING DOMAINS AS A MEANS OF CELL LABELING, CELL ORGANIZATION AND POLYMER SEQUENCING

Related Application Data

5 This application claims priority to U.S. Provisional Patent Application No. 61/683,690, filed on August 15, 2012 and is hereby incorporated herein by reference in its entirety for all purposes.

Background of the Invention

10 The invention relates to a fusion protein with applications in cell labeling, nanopore sequencing and the organizing of cells.

Our ability to understand and ultimately engineer complex cellular niches relies on methods that enable monitoring of cells and their interactions. Towards this the use of fluorescent proteins has revolutionized probing of biological phenomenon. However using standard 3-color imaging modalities, multiplexed probing is limited to a few distinct cell types corresponding to a combinatorial expression of resolvable fluorescent proteins.

15 A second set of unsolved problems have hampered the development of nanopore sequencing. These critical technical challenges include the need for a means to guide the polymer through the nanopore and a means to regulate the rate at which the polymer passes through.

An additional unsolved problem is how to create artificial tissues.

20 The present invention seeks to address these problems.

Summary of the Invention

The present invention uses polynucleotide-binding domains and complementary polynucleotides to provide solutions to the above-mentioned challenges. DNA binding domains such as zinc finger proteins (ZFs) and more recently transcription activator-like effectors (TALEs) and their ligands provide a unique and versatile tool, since both the receptor (ZF or TALE protein) and the ligand (DNA) are highly programmable and hence the space of orthogonal interactions that can be engineered is potentially infinite.

30 Accordingly, a first aspect, the invention relates to a protein including (a) a polynucleotide-binding domain; and (b) a transmembrane domain. In one embodiment, the polynucleotide-binding domain is a programmable DNA binding domain. In a related aspect, the polynucleotide-binding domain is a zinc finger protein or TAL effector protein. In a related embodiment, the zinc finger domain is a C₂H₂ triple finger. In another related embodiment, the zinc finger is a synthetic zinc finger such as a zinc finger selected from the group consisting of Z1, Z2, Z3, Z4, Z5, Z6, Z7, Z8, Z9, Z10, 35 Z11, Z12, Z13, Z14, Z15, and Z16. In some aspects, the protein includes an Ig κ-chain leader

sequence at its N-terminus. In some embodiments, the protein includes an import sequence at its C-terminus. In some aspects, the transmembrane domain is C-terminal of the polynucleotide-binding domain. In a related embodiment, the protein is a nanopore subunit.

In a related aspect, the invention relates to a polynucleotide encoding the above-mentioned protein. The polynucleotide may be operably linked to a promoter. In one embodiment, the polynucleotide is a free, linear unit. In another aspect, the invention relates to a plasmid including the polynucleotide. In another aspect, the invention relates to a viral vector including the polynucleotide. In another aspect, the invention relates to a cell including the abovementioned polynucleotide, plasmid, or vector.

In some embodiments, the invention relates to a protein complex including one or more proteins of the first aspect. In one embodiment, the complex is a nanopore. In some embodiments, the nanopore is selected from the group consisting of alpha-hemolysin (α HL) and mycobacterium smegmatis porin A (MspA). In a related embodiment, the invention relates to a membrane including one or more variants of the nanopore. In another aspect, the invention relates to a cell including one or more variants of the protein or protein complex, where the polynucleotide-binding domain(s) of the protein(s) or protein complex(es) are on the surface of the cell in a manner capable of interaction with materials external to the cell.

In another aspect, the invention relates to a method of sequencing a polynucleotide, the method including: (a) providing two separate pools of liquid containing electrically conductive medium and a nanopore-perforated interface between the two pools, where the interface contains a nanopore; (b) providing candidate polynucleotide molecules in one of the pools; (c) applying a voltage differential across the pools; and (d) making interface-dependent measurements of ionic current over time as individual nucleotides of a single polynucleotide interact sequentially with the interface, yielding data suitable to determine a nucleotide-dependent characteristic of the polynucleotide. In a related aspect, the interface is a membrane.

In another aspect, the invention relates to a method of tagging a cell including the steps of (a) providing a cell including one or more variants of the protein or protein complex, where the polynucleotide-binding domain(s) of the protein(s) or protein complex(es) are on the surface of the cell in a manner capable of interaction with materials external to the cell; (b) providing a polynucleotide probe including a detectable label and capable of interacting with the surface polynucleotide-binding domain(s); (c) contacting the cell with the polynucleotide probe; and (d) detecting the polynucleotide probe. In certain embodiments, the detectable label is a fluorescent label or a luminescent label, quenching label, biotin label, ligation label, chemically reactive moiety such as azide, chemically reactive moiety that enables click-chemistry, detectable small molecule, or the sequence of polynucleotide probe.

In another aspect, the invention relates to a method of labeling and sequentially relabeling cells, the method including the steps: (a) providing a cell including one or more variants of the protein or protein complex, where the polynucleotide-binding domain(s) of the protein(s) or protein complex(es) are on the surface of the cell in a manner capable of interaction with materials external to the cell; (b) providing a polynucleotide probe including a detectable label that is capable of interacting with the surface polynucleotide-binding domain(s); (c) contacting the cell with the polynucleotide probe; (d) detecting the polynucleotide probe; (e) dissociating the polynucleotide probe from the polynucleotide-binding domain(s); (f) providing a subsequent polynucleotide probe including a detectable label; (g) contacting the cells with the subsequent polynucleotide probe; and (h) detecting the subsequent polynucleotide probe. In certain embodiments, one or more of the detectable label is a fluorescent label or a luminescent label, quenching label, biotin label, ligation label, chemically reactive moiety such as azide, chemically reactive moiety that enables click-chemistry, detectable small molecule, or the sequence of polynucleotide probe.

In another related embodiment, the dissociation is achieved by the addition of a high concentration of unmarked polynucleotide(s) capable of displacing the polynucleotide probe. In another related embodiment, the method further includes repeating one or more times the steps (e) to (h) with an additional polynucleotide probe or probes.

In another aspect, the invention relates to a method of labeling a cell, the method including the steps of (a) providing a cell providing a cell including one or more variants of the protein or protein complex, where the polynucleotide-binding domain(s) of the protein(s) or protein complex(es) are on the surface of the cell in a manner capable of interaction with materials external to the cell; (b) providing a scaffold polynucleotide to mediate the interaction of the surface polynucleotide binding domain(s) with one or more polynucleotide probe(s), the scaffold polynucleotide including (i) a double-stranded domain and (ii) a single-stranded domain including one or more sites capable of hybridization, where the double-stranded domain interacts with the surface polynucleotide-binding domain(s) and the single-stranded domain is available to interact with one or more polynucleotide probes; (c) providing a polynucleotide probe including a detectable label and capable of interacting with the scaffold polynucleotide; (d) contacting the cell with the scaffold polynucleotide and the polynucleotide probe; and (e) detecting the polynucleotide probe. In related embodiments, the detectable label is a fluorescent label or is a luminescent label, quenching label, biotin label, ligation label, chemically reactive moiety such as azide, chemically reactive moiety that enables click-chemistry, detectable small molecule, or the sequence of polynucleotide probe.

In another aspect, the invention relates to a method of labeling and sequentially relabeling cells, the method including the steps of (a) providing a cell providing a cell including one or more variants of the protein or protein complex, where the polynucleotide-binding domain(s) of the protein(s) or protein complex(es) are on the surface of the cell in a manner capable of interaction with

materials external to the cell; (b) providing a scaffold polynucleotide to mediate the interaction of the surface polynucleotide binding domain(s) with one or more polynucleotide probe(s), the scaffold polynucleotide including (i) a double-stranded domain and (ii) a single-stranded domain including one or more sites capable of hybridization, where the double-stranded domain interacts with the surface polynucleotide-binding domain(s) and the single-stranded domain is available to interact with one or more polynucleotide probes; (c) providing a polynucleotide probe including a detectable label and capable of interacting with the scaffold polynucleotide; (d) contacting the cell with the scaffold polynucleotide and the polynucleotide probe; (e) detecting the polynucleotide probe; (f) providing a signal-quenching polynucleotide that includes (i) a polynucleotide sequence complementary to a region sufficiently adjacent to the polynucleotide probe and (ii) a mechanism of quenching the signal of the polynucleotide probe label; (g) contacting the cells with one or more of the quenching polynucleotide probe; (h) providing a subsequent polynucleotide probe including a detectable label and complementary to an unbound segment of the scaffold polynucleotide; (i) contacting the cells with the subsequent polynucleotide probe; and (j) detecting the subsequent polynucleotide probe. In a related embodiment, the labels are fluorescent labels and the mechanism of quenching is a mechanism of quenching fluorescence. In another related embodiment, one or more of the detectable labels is a luminescent label, quenching label, biotin label, ligation label, chemically reactive moiety such as azide, chemically reactive moiety that enables click-chemistry, detectable small molecule, or the sequence of said polynucleotide probe. In another related embodiment, the method further includes the step of (k) repeating one or more times the steps (f) to (j) with additional quenching polynucleotide(s) and polynucleotide probe(s). In some embodiments, the cells are fixed or are growing adherently. In some embodiments, the cells are in solution. In some embodiments, expression of the polynucleotide-binding domain is under control of a promoter and the level of labeling is indicative of promoter activity.

In another aspect, the invention relates to a method for organizing cells, the method including the steps of (a) providing a solid surface to which one or more polynucleotides are affixed in a defined manner; (b) providing a cell or cells including a surface polynucleotide-binding domain capable of directly binding the affixed polynucleotides; and (c) contacting the solid surface with the cells, thereby affixing the cells. In a related embodiment, the affixed cells include an additional surface polynucleotide-binding domain, and the method further includes the steps: (d) providing a new cell or cells including surface polynucleotide-binding domains; (e) providing one or more polynucleotide(s) capable of tethering the affixed and the new cells in a sequence-specific manner; and (f) contacting the tethering polynucleotide(s) with the affixed cells and the new cells. In another related embodiment, the tethering polynucleotide is a double-stranded polynucleotide probe including regions capable of being bound by the surface polynucleotide-binding domains of both the affixed cells and the new cells. In another related embodiment, the tethering is mediated by two polynucleotides, each including (i) a

double-stranded domain capable of interacting with a surface polynucleotide-binding domain and (ii) a single-stranded domain, and where one tethering probe's double-stranded domain interacts directly with the affixed cells, the other tethering probe's double-stranded domain interacts directly with the new cells, and the single-stranded domains of the two tethering probes are complementary so as to tether the cells. In a related embodiment, the method further includes the step: (h) repeating steps (d)-(f) one or more times with additional cells and/or tethering polynucleotides.

In another aspect, the invention relates to a method for organizing cells, the method including the steps of (a) providing a solid surface to which one or more polynucleotides are affixed in a defined manner; (b) providing a cell or cells including a surface polynucleotide-binding domain; (c) providing a scaffolding polynucleotide capable of binding the cell surface polynucleotide-binding domain, the polynucleotide probe including (a) a double-stranded domain and (b) a single-stranded domain, where the double-stranded domain interacts with the surface polynucleotide-binding domain(s) and the single-stranded domain is available to interact with one or more of the affixed polynucleotides; and (d) contacting the solid surface with the cells, thereby affixing the cells. In a related embodiment, the affixed cells include an additional surface polynucleotide-binding domain, and the method further includes the steps of (e) providing a new cell or cells including surface polynucleotide-binding domains; (f) providing one or more polynucleotide(s) capable of tethering the affixed and the new cells in a sequence-specific manner; and (g) contacting the tethering polynucleotide(s) with the affixed cells and the new cells. In another related embodiment, the tethering polynucleotide is a double-stranded polynucleotide probe including regions capable of being bound by the surface polynucleotide-binding domains of both the affixed cells and the new cells. In another related embodiment, the tethering is mediated by two polynucleotides, each including (i) a double-stranded domain capable of interacting with a surface polynucleotide-binding domain and (ii) a single-stranded domain, and where one tethering probe's double-stranded domain interacts directly with the affixed cells, the other tethering probe's double-stranded domain interacts directly with the new cells, and the single-stranded domains of the two tethering probes are complementary so as to tether the cells. In another related embodiment, the method further includes the step of (h) repeating steps (e)-(g) one or more times with additional cells and/or tethering polynucleotides.

In another aspect, the invention relates to a method of tethering free cells, the method including the steps of (a) providing two or more free cells including one or more surface polynucleotide-binding domains; (b) providing one or more tethering polynucleotides capable of tethering two of the free cells; and (c) contacting the one or more polynucleotides with the free cells. In a related embodiment, the tethering polynucleotide is a double-stranded polynucleotide probe including regions capable of being bound by the surface polynucleotide-binding domains of both the affixed cells and new cells. In another related aspect, the tethering is mediated by two polynucleotides, each including (i) a double-stranded domain capable of interacting with a surface polynucleotide-

binding domain of the free cells and (ii) a single-stranded domain, where the single-stranded domains of the two tethering probes are complementary so as to tether the free cells. In another related embodiment, the method further includes the step of (d) Repeating steps (a)-(c) one or more times with additional cells and/or tethering polynucleotides.

5 In another aspect, the invention relates to an artificial tissue produced by the above-mentioned methods. In particular embodiment, the artificial tissue is vascular tissue or an organoid.

“Fusion protein” means a protein having a sequence derived from two or more source proteins.

“Fusion gene” means a gene that encodes a fusion protein.

10 “Polynucleotide-binding domain” means any polypeptide or portion of a polypeptide that binds polynucleotides with specificity toward polypeptides of a particular sequence, sequences, or class of sequences. “DNA binding domain” means a polynucleotide-binding domain that interacts with DNA. “Surface polynucleotide-binding domain” means a polynucleotide-binding domain integrated within the membrane of a cell in a manner that permits the interaction with polynucleotides external to the cell.

15 “Nanopore” means a hole or passage through a membrane formed by a multimeric protein ring. Typically, the passage is 0.2-25 nm wide.

“Nanopore sequencing” means a method of determining the components of a polymer, such as a polynucleotide, based upon interaction of the polymer with the nanopore. Nanopore sequencing may be achieved by measuring a change in the conductance of ions through a nanopore that occurs when
20 the size of the opening is altered by interaction with the polymer.

“Polynucleotide probe” means any polynucleotide that may be appended to a cell in a manner that permits the detection of the cell, assignment of identity or class of the cell, or enables the utilization of the cell for a desired purpose. A probe may include a detectable label. “Tagging” means any process that results in the appending of a probe to a cell.

25 “Scaffold polynucleotide” means a polynucleotide probe that includes a double stranded domain and a single stranded domain, such that the double stranded domain is designed to interact with a surface polynucleotide-binding domain and the single stranded domain is designed to interact with one or more polynucleotide probes to guide the association of these elements.

“Tissue” means any group of cells organized into a defined structure.

30 The present invention provides advantages in cell labeling, nanopore sequencing and the organizing of cells. Other features and advantages of the invention will be apparent from the following Detailed Description, the drawings, and the claims.

Brief Description of the Drawings

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

- 5 The foregoing and other features and advantages of the present invention will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings in which:

Figure 1 is a table listing 16 synthetic zinc fingers with their protein sequences, target sequences, and the vectors constructed.

- 10 Figure 2 is a schematic presenting one non-limiting mechanism by which cells expressing fusion polynucleotide-binding proteins might express them at the cell surface and interact with complementary DNA molecules.

Figure 3 is a series of eight images (2 phase and 6 fluorescence) showing that zinc finger fusion proteins expressed in cells are able to bind complementary probes with specificity.

- 15 Figure 4a is a set of four images showing the interaction of cells expressing a zinc finger fusion protein with a dsDNA or ssDNA probe in the presence or absence of salmon sperm DNA. Figure 4b is a set of 12 images showing the interaction of cells expressing a zinc finger fusion protein with a dsDNA probe in the presence or absence of an ssDNA probe. Figure 4c is 4 graphs displaying a FACS analysis of control cells or cells expressing a zinc finger fusion protein in the presence of
20 complementary probes.

- Figure 5a is a set of 3 images showing cells expressing two zinc finger fusion proteins in the presence of complementary dsDNA probes and a scatter plot showing the overlap characteristics of the two probes. Figure 5b is two images showing the labeling of cells expressing two or three zinc finger fusion proteins and two scatter plots showing overlap when labeling the zinc fingers with
25 complementary probes. Figure 5c is 4 images of cells expressing one, two or three zinc finger fusion proteins in a total of 7 combinations and labeled with each of three probes, with an overlay displayed in the fourth image.

- Figure 6a shows a series of 16 images, each individually showing a cell expressing a zinc finger fusion protein in the presence of complementary dsDNA probe. Figure 6b shows cells
30 expressing each of 16 zinc finger fusion proteins and the interaction of each when probed with 16 probes.

Figure 7 is a schematic workflow of an image analysis strategy that may be employed to analyze the binding of probes to polynucleotide-binding domains.

Figure 8 is a bar graph showing the correlation of fluorescence intensity of identified cell regions.

Figure 9a is a chart showing the results of image analysis, including the counts and fractions of segments and NOVS formed per channel aggregated over all images and all channels. Figure 9b is a chart showing the results of image analysis, including overall segment and NOVS counts and fractions from each channel in each image.

Figure 10 is a pair of box plot charts showing the segment area size distribution for all images analyzed.

Figure 11 is a box plot chart and a bar graph showing segment intensity distribution of all images analyzed.

Figure 12a shows 3 scatter plots and 3 pixel histograms showing, respectively, the mean intensities resulting from image analysis and the pixel counts resulting from image analysis. Figure 12b shows 3 scatter plots and 3 pixel histograms showing, respectively, the mean intensities resulting from image analysis and the pixel counts resulting from image analysis.

Figure 13 shows 5 scatter plots and 15 pixel histograms showing, respectively, the mean intensities resulting from image analysis and the pixel counts resulting from image analysis.

Figure 14a shows 2 sets of 5 images, each set displaying a 48 minute time course in which cells expressing a zinc finger fusion protein in the presence of a complementary dsDNA probe are treated with either salmon sperm DNA or salmon sperm DNA and a high concentration of a non-fluorescent target dsDNA in solution. Figure 14b is a chart displaying the residual mean normalized pixel intensities measured over the 48 minute time course of dissociation.

Figure 15a is a schematic presenting one non-limiting method by which cells expressing fusion polynucleotide-binding proteins might be sequentially labeled and re-labeled through the use of quencher probes. Figure 15b is a series of images showing the labeling, quenching, and subsequent relabeling of cells expressing one of three zinc finger fusion proteins and a schematic showing a labeling scheme for the zinc finger fusion protein-expressing cells utilized in the images, as well as three others. Figure 15c is (a) a schematic presenting one non-limiting method by which cells expressing fusion polynucleotide-binding proteins can be sequentially labeled and re-labeled through the use of quencher probes and (b) 4 images showing mixed cells, each expressing one of 6 zinc finger fusion proteins, that are labeled, quenched, and re-labeled.

Figure 16 is a schematic presenting one non-limiting method by which a zinc finger fusion protein could serve as a reporter of endogenous activity and a series of 12 images showing cells with an inducible zinc finger fusion protein in the presence of a complementary probe under conditions that induce or do not induce expression of the zinc finger fusion protein.

Figure 17 is a set of 4 images showing cells with an inducible zinc finger fusion protein in the presence of a complementary probe under conditions that induce or do not induce expression of the zinc finger fusion protein.

5

Detailed Description

The present invention relates to a fusion protein comprising a polynucleotide-binding domain and a transmembrane domain. In some embodiments, the fusion protein is expressed at the cell surface and is thereby available to bind polynucleotides outside the cell. The applications of this protein include the improvement of nanopore technology for DNA sequencing, advanced methodologies for labeling cells, and construction of artificial tissues.

10

Characteristics of the Fusion Protein

The polynucleotide-binding domain of the fusion protein may be any polypeptide that is known to bind polynucleotides, such as DNA, or that is encompassed by a family of proteins known to bind polynucleotides. Zinc fingers are a well-recognized group of polynucleotide-binding proteins, encompassing various subclasses of fold groups including Cys₂His₂, gag knuckle, treble clef, zinc ribbon, Zn₂/Cys₆, and TAZ2 domain-like factors. Other polynucleotide-binding domains include helix-turn-helix, leucine zipper, winged helix, winged helix-turn-helix, helix-loop-helix, HMG-box, immunoglobulin fold, B3 domain, TAL effector DNA-binding domain, and other domains. These and others known in the art are suitable for use in the current invention. Polynucleotide-binding proteins like zinc fingers can be modified to generate polynucleotide-binding proteins that bind a desired sequence or class of sequences with specificity. The Cys₂His₂ motif is especially well-suited to such modification, providing a scaffold upon which specificity factors can be assembled by methods that will be familiar to those skilled in the art.

15

20

25

30

35

A transmembrane domain is another component of the fusion protein of present invention. The transmembrane domain may be any three-dimensional structure which is thermodynamically stable in a membrane. The transmembrane domain may be a single transmembrane alpha helix. The transmembrane domain can be a helical bundle, seven-transmembrane protein, beta barrels, or any other class of transmembrane domain known in the art, as well as any member of such a class or any polypeptide otherwise known to act as a transmembrane domain. In some embodiments of the fusion protein, the transmembrane domain is C-terminal of the polynucleotide-binding domain. In some embodiments, a transmembrane domain is defined as a polypeptide sequence with high hydrophobicity score based on any of the algorithms known in the art or with a long series of hydrophobic residues. In some embodiments, the transmembrane domain is a platelet derived growth factor (PDGF) transmembrane domain. In other embodiments, the transmembrane domain is selected from the transmembrane regions of transmembrane proteins of known structure or related proteins.

Transmembrane proteins of known structure include ram prostaglandin H₂ synthase-1 (COX-1), squalene-hopene cyclase: *Alicyclobacillus acidocaldarius*, monoamine oxidase B: human mitochondrial outer membrane, fatty acid amide hydrolase: *Rattus norvegicus*, sulfide:quinone oxidoreductase in complex with decylubiquinone: *Aquifex aeolicus*, porin: *Rhodobacter capsulatus*, TolC outer membrane protein: *Escherichia coli*, bacteriorhodopsin (BR): *Halobacterium salinarium*, and many others, as curated in a list available at [http://blanco.biomol.uci.edu/membrane_proteins_xtal.html], the contents of which is herein incorporated by reference. In other embodiments, the transmembrane domain is selected from the transmembrane regions of transmembrane proteins known in humans or related proteins, as curated in a list available at [https://modbase.compbio.ucsf.edu/projects/membrane/] and incorporated herein by reference.

The invention features any genetic material that may be translated into the fusion gene of the invention. The protein may be expressed from a linear segment of DNA, DNA encoded within a plasmid, DNA encoded within a virus, DNA incorporated within a genome, DNA present within a cell, or DNA present within an organism.

Nanopores

The fusion protein may be part of a nanopore or nanopore subunit. Nanopores, as used herein, are transmembrane structures that may permit the passage of molecules through a membrane. A multimeric nanopore may include any number of the fusion nanopore subunits (such as 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 fusion nanopore subunits) and any number of other subunits (such as 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 other nanopore subunits). Examples of nanopores include α -hemolysin (*Staphylococcus aureus*) and MspA (*Mycobacterium smegmatis*). In two embodiments, the nanopore fusion gene comprises or is part of an α -hemolysin or MspA multimer. Other examples of nanopores may be found in the art describing nanopore sequencing or described in the art as pore-forming toxins. In some embodiments, the nanopore fusion gene comprises or is part of a pore-forming toxin, such as the β -PFTs Pantone-Valentine leukocidin S, aerolysin, and Clostridial Epsilon-toxin, the α -PFTs cytolysin A, the binary PFT anthrax toxin, or others such as pneumolysin or gramicidin.

Nanopores have become technologically and economically significant with the advent of nanopore sequencing technology. Methods for nanopore sequencing are known in the art, for example, as described in U.S.P.N. 5,795,782, which is incorporated by reference. Briefly, nanopore sequencing involves a nanopore-perforated membrane immersed in a voltage-conducting fluid. A voltage is applied across the membrane, and an electric current results from the conduction of ions through the nanopore. When the nanopore interacts with polymers, such as DNA, flow through the nanopore is modulated in a monomer-specific manner, resulting in a change in the current that permits identification of the monomer(s).

The development of nanopore sequencing is hampered by critical technical challenges, including the need for a means to guide the polymer through the nanopore and a means to regulate the rate at which the polymer passes through. The present invention provides a solution to these challenges by placing a polynucleotide-binding domain at the nanopore to modulate the movement of the polymer. In some embodiments, the polynucleotide-binding domain of the nanopore fusion gene binds polynucleotides and guides their progress through the nanopore at a moderated rate.

The fusion nanopore can be integrated with other nanopore technologies. In some embodiments, a DNA polymerase is also used to ratchet a DNA substrate through the nanopore. In some embodiments, the polymerase is phi29 DNA polymerase, T7 DNA polymerase, Klenow fragment of DNA polymerase 1, or other DNA polymerase. In other embodiments, passive approaches can be used to slow the movement of DNA. These approaches may include nucleotide labeling, end termination of ssDNA with DNA hairpins, the use of positively charged residues in the nanopore as molecular brakes, and modification of pore shape to optimize processivity.

In some embodiments, the polynucleotide-binding domain of the fusion nanopore is non-sequence-specific and interacts with polynucleotides with minimal or no sequence discrimination. In some embodiments, a non-specific polynucleotide-binding domain is used for sequencing.

In some embodiments, the polynucleotide-binding domain of the fusion nanopore is specific to particular polynucleotide sequences or classes of sequences. In some embodiments, a specific polynucleotide-binding domain is used, for instance, for use in polymorphism analysis, analysis of particularly selected loci, medical diagnostic applications, forensics, or other purposes.

In some embodiments, the membrane of the nanopore sequencing technique is lipid bilayer. In some embodiments, the membrane of the nanopore sequencing technique is an artificial membrane, composed of a material such as Al₂O₃, TiO₂, HfO₂, SiO₂, SiN, or grapheme.

The various embodiments of the fusion nanopore may be employed in the sequencing of single-stranded or double-stranded DNA, cDNA, RNA, mRNA, tRNA, rRNA, microRNA, siRNA, or any polynucleotide, as well as other polymers including but not limited to polypeptides.

Surface Polynucleotide-binding Domains

In additional embodiments, the fusion protein is expressed within or on the surface of a cell. In some embodiments, the fusion protein is associated with the membrane of a cell in a manner that permits the polynucleotide-binding domain to interact with polynucleotides external to the cell. In some embodiments, the fusion protein includes a secretion signal to direct the fusion protein to the cell surface. One such secretion signal is the Igk-chain leader sequence. In some embodiments, an Igk-chain leader sequence is at the N-terminus of the fusion protein. In some embodiments, the fusion protein includes an endoplasmic reticulum import sequence to direct the fusion protein to the cell

surface. An endoplasmic reticulum import sequence may be based upon the serotonin receptor 5HT3A.

Labeling Strategies

5 In some embodiments, a surface polynucleotide-binding protein is used as part of a cell-labeling strategy. A shortfall of fluorescent labeling, a primary labeling method in the art, is that it relies upon the detection of a limited number of colors. In the present invention, surface polynucleotide-binding proteins enable a barcoding strategy capable of distinguishing a much greater number of cells or cell types.

Direct Labeling Strategies

10 In one embodiment of this labeling strategy, each cell or cell type of interest expresses or is engineered to express one or more distinct surface polynucleotide-binding domain(s) with specificity to a unique polynucleotide sequence. The cells may then be contacted with polynucleotide probes, some or all of which are complementary to the polynucleotide-binding domain(s) expressed by the population of cells. The probes can interact with the polynucleotide-binding domains, tagging the cells.

15 In some embodiments, the probes may be labeled. For instance, in some embodiments, the probes are labeled with fluorescent moieties. In some embodiments, each probe is labeled with a distinct fluorescent moiety. Thus, the probes tag the cells expressing the complementary polynucleotide-binding domain(s), and cells expressing each probed polynucleotide-binding domain are then recognizable by fluorescent microscopy. However, because the number of distinguishable fluorescent moieties is limited, the number of species distinguishable by this method alone is limited.

20 A much greater number of species can be detected if the initial probes are removed or quenched, and the cells subsequently re-probed with a second probe set. In some embodiments, the initial set of probes is dissociated. This dissociation can be achieved by the addition of a high concentration of unlabeled probes targeted to the same polynucleotide-binding domain(s) as the initial probes, displacing the initial probes. A subsequent, distinct probe set may then be introduced, relabeling the cells. In certain embodiments, the subsequent probe set targets a distinct, previously untargeted set of polynucleotide-binding domain(s) expressed by the same set of cells. In one embodiment, each of the subsequent probes is labeled with a distinct fluorescent moiety. The probes then tag the cells expressing the complementary polynucleotide-binding domain(s), and the identity of cells expressing each probed polynucleotide-binding domain is recognizable by fluorescent microscopy. Because the color coding of the second probe set is independent of the first, the number of cells or cell types that can be distinguished increases exponentially with each subsequent round of labeling. The relabeling of the cells may be repeated indefinitely. In one embodiment, relabeling is

carried out until all cell types have been distinguished or the availability of unique polynucleotide-binding domain(s) has been exhausted. The series of labels associated with each cell over the course of one or more rounds of labeling serves as a barcode for the identity of that cell, as defined by the surface polynucleotide-binding domains it expresses.

5 In certain embodiments of direct barcode labeling, cells are fixed or growing adherently. In other embodiments, the cells are in solution, such as liquid media. In some embodiments, one or more polynucleotide probes are prehybridized with an oligonucleotide. In some embodiments, labels or methods of probe detection may include, but need not be limited to, fluorescence, quenching, luminescence, ligation, PCR, chemically re-active moieties such as azides that enable click-chemistry, 10 biotinylation and others.

Indirect Labeling Strategies

In alternative embodiments, probes associate indirectly with the surface polynucleotide-binding domains. A scaffold polynucleotide binds the surface polynucleotide-binding domain, and the 15 probes then bind the scaffold. Because a single scaffold can bind multiple probes, this method allows barcoding without dissociation steps and requires only one surface polynucleotide-binding domain per cell type.

In some embodiments, a cell or cell type of interest expresses or is engineered to express one or more distinct surface polynucleotide-binding domain(s) with specificity to a unique polynucleotide 20 sequence. The cells may then be contacted with one or more scaffold polynucleotides. Each scaffold polynucleotide includes both a double-stranded domain and a single-stranded domain. The double-stranded domain interacts with the surface polynucleotide-binding domain(s) to which it is complementary, while the single-stranded domain remains available for interaction with additional reagents, such as polynucleotide probes. In some embodiments, each distinct scaffold polynucleotide 25 bears a distinct single-stranded domain. In some embodiments, cells are then contacted with polynucleotide probes complementary to the available single-stranded domains of the bound scaffold polynucleotides. The length of the single-stranded domain of the scaffold polynucleotide may be greater than the length of the polynucleotide probe, such that a single scaffold single-stranded domain could accommodate multiple polynucleotide probes.

30 In some embodiments, the probes may be labeled. For instance, in some embodiments, the probes will be labeled with fluorescent moieties. In some embodiments, each probe will be labeled with a distinct fluorescent moiety. The probes can tag the cells associated with the complementary single-stranded domain(s), and surface polynucleotide-binding domains of each cell are then visualized by fluorescent microscopy. However, because the number of distinguishable fluorescent moieties is 35 limited, the number of species distinguishable by this method alone is limited.

In certain embodiments, the label of the initial probe may be quenched by a signal-quenching polynucleotide. The signal-quenching probe is a polynucleotide that is complementary to a region adjacent to that bound by the initial probe and that bears a signal-quenching mechanism. In some embodiments, the label is a fluorescent label and the signal-quenching mechanism is a fluorescence-quenching mechanism. In some embodiments, cells are contacted with these signal-quenching probes and the signal generated by the label of the initial probe is extinguished. The cells are thereby primed for subsequent re-labeling.

In certain embodiments, cells are subsequently probed with a second set of labeled polynucleotide probes. In further embodiments, these probes are designed to target a segment of the single-stranded domain of the scaffold polynucleotide that remains unbound by either the initial probe or the signal-quenching probe. In additional embodiments, the process of labeling, quenching, and re-labeling may followed by a second quenching and, in some embodiments, repeated in sequence until all cells are distinguished or all probe binding sites are exhausted. In some embodiments, the labels are fluorescent labels and the signal-quenching mechanism is a fluorescence-quenching mechanism.

In certain embodiments of indirect barcode labeling, cells are fixed or growing adherently. In other embodiments, the cells are in solution, such as liquid media. In some embodiments, one or more polynucleotide probes are prehybridized with an oligonucleotide. In some embodiments, labels or methods of probe detection may include, but need not be limited to, fluorescence, quenching, luminescence, ligation, PCR, chemically re-active moieties such as azides that enable click-chemistry, biotinylation and others.

Gene Expression

The present invention provides tools for the determination of gene expression. In some embodiments, the expression of a fusion gene of the present invention is driven by the promoter of a gene of interest. In this way, expression of the fusion gene is indicative the transcriptional activity of the gene of interest. In some embodiments, the promoter of the gene of interest drives expression of the fusion gene. In some embodiments, the fusion gene is expressed in tandem with the entire gene of interest, or a portion thereof. In some embodiments, the protein translated from the expressed fusion gene associates with the cell membrane and presents a surface polynucleotide-binding domain. The polynucleotide-binding domain may be labeled by any of the direct or indirect labeling methods described above, and the level of labeling is indicative of the activity of the utilized promoter.

Artificial Tissues

The ability to generate artificial tissues is a subject of enormous medical and scientific interest. The present invention provides a means of organizing cells, and thereby generating artificial tissues.

In some embodiments, polynucleotide probes are affixed to a solid surface in a defined manner. These polynucleotide probes are complementary to one or more surface polynucleotide-binding domains present on a provided set of cells. Upon contacting the cells with the affixed polynucleotides, the surface polynucleotide binding domains bind the probes to which they are complementary, thereby forming a layer of cells organized according to the interaction of the fusion gene and the probes. In some embodiments, the cells possess unbound surface polynucleotide-binding domains available to interact with additional polynucleotides.

In further embodiments, a second set of cells is organized upon the first. In some embodiments, the surface-affixed cells are contacted with a second set of cells with surface polynucleotide-binding domains and a polynucleotide tether. In some embodiments, the polynucleotide tether is a double-stranded polynucleotide with ends respectively complementary to surface polynucleotide-binding domains present on the first or second set of cells. In other embodiments, the tether may composed of two scaffold polynucleotides, with double-stranded regions complementary to the surface polynucleotide-binding domains of the initial or new cells, respectively, and single-stranded domains complementary to each other. The specificity of the surface polynucleotide-binding domains defines the manner in which the second set of cells is arranged upon the first.

In additional embodiments of the above method, the initial set of cells may be tethered to the surface-affixed probes by a scaffolding polynucleotide having a double-stranded domain complementary to one or more surface polynucleotide-binding domains on the cells and a single-stranded domain complementary to the surface-affixed probes.

In additional embodiments, similar strategies may be employed to tether cells in solution, such as liquid media. Two or more cells, each expressing one or more surface polynucleotide-binding domains, could be tethered directly by a double-stranded polynucleotide tether or indirectly by a pair of scaffolding polynucleotides with complementary single-stranded domains. In additional embodiments, subsequent cells and polynucleotides could be added to generate tissues of increasing complexity.

In some embodiments, one or more polynucleotide probes, scaffolds or tethers are prehybridized with an oligonucleotide.

The number of distinct probes, scaffolds, tethers and cell types added in each step may be limited (such as 1, 2, or 3) or extensive (such as 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, or 1000). The subsequent tethering of additional sets of cells may be carried out repeatedly as described, generating tissues of increasing complexity.

Examples of tissues that may be constructed, or of which models or equivalents may be constructed, in the described fashion include, but are not limited to organoids, vascular networks, muscle tissue and any tissue found in the human body.

Zinc finger proteins expressed on the cell membrane can bind their cognate target dsDNA specifically in both simplex and multiplex settings and can thus serve as faithful cellular barcodes. Furthermore, like fluorescent proteins they can also serve as reporters of endogenous activity such as transcriptional activation. This feature opens the possibility to doing highly-multiplexed tracking of endogenous genes and pathways. Finally, by exploiting the potential for storing additional sequence information in the sZF DNA probes and reading this using a novel tagging scheme, an almost arbitrary number of thus barcoded cells in complex mixtures can be imaged.

Because sZFs are expressed on the cell surface they are physically accessible and hence offer the ability to not only label cells using DNA probes (as outlined above), but they also provide convenient handles for DNA mediated cell capture, and targeting of barcoded cells. Thus as compared to fluorescent proteins or antibodies, sZFs provide more versatile barcodes for cells.

The examples below are intended to illustrate, rather than limit, the invention.

Examples

Example 1. Design and construction of synthetic polynucleotide-binding domain/polynucleotide pairs

DNA binding domains such as zinc finger proteins (ZFs), and more recently transcription activator-like effectors (TALEs), have found numerous applications as synthetic transcription factors and editing tools such as site-specific nucleases, recombinases, and methylases. As a receptor-ligand pair, polynucleotide-binding domains and polynucleotides provide a unique and versatile combination since both the receptor (ZF or TALE protein) and the ligand (DNA) are highly programmable and hence the space of orthogonal interactions that can be engineered is potentially infinite. A total of 16 zinc finger proteins were designed and generated, providing concrete examples of systematically programmed polynucleotide-binding domain/polynucleotide pairs. The protein sequences and target dsDNA sequences are provided in Fig. 1.

Example 2. Expressing zinc-finger DNA binding domains on the cell surface

To express zinc-finger (sZF) DNA binding domains on the cell surface, we fused each domain's N terminus to an Ig κ -chain leader sequence and each C-terminus to the platelet derived growth factor (PDGF) transmembrane domain.

To determine whether cells expressing these fusion proteins would express them at the cell surface and, as a result, be able to interact with complementary DNA molecules, the cells were exposed to fluorophore-tagged DNA molecules. A schematic presenting one non-limiting mechanism by which this interaction may occur is shown in Fig. 2.

Cells expressing the sZF constructs strongly bound the DNA while control cells exhibited very low binding signals, implying functional zinc-finger proteins were successfully expressed on the cell surface (Fig. 3).

Two aspects of this sZF-DNA interaction were of note: first, sZFs were observed to bind to both single and double stranded DNA molecules (Fig. 4a); however, the former interaction was abrogated in the presence of competitor dsDNA (here Salmon Sperm DNA). Second, sZFs also non-specifically bound to most dsDNA, but again in the presence of competitor dsDNA, binding to only their cognate target dsDNA was retained (Fig. 4b). Similar results were obtained using FACS based assays (Fig. 4c). Thus fusion gene-expressing cells specifically bind target dsDNA probes (Fig. 3).

Additional endoplasmic reticulum import sequences based on the serotonin receptor 5HT3A, and transmembrane domains from Neurokinin-1 receptor (NK1R) or beta-2 adrenergic receptor were used in additional constructs and were met with similar success.

Example 3. Verification of predicted polynucleotide-binding domain/polynucleotide pairs in cells

To verify the 16 zinc finger proteins and their dsDNA target sequences would interact as predicted, and further that they would do so when the polynucleotide-binding domain was expressed at the cell surface, the zinc fingers were modified for expression in cells, expressed in cells, and tested for binding of complementary dsDNA probes as described in Example 2 (Fig. 5a-c).

From these experiments, different sZFs were observed to have different binding affinities for their target dsDNA (Fig. 6a). Specifically, as assayed by both fluorescence intensity and duration of binding, some sZFs bound their targets strongly (ZFs 1, 3, 8, 12, 13, 15, 16), some moderately strongly (ZFs 2, 4, 5, 6, 7, 10, 14), and others only weakly (ZFs 9, 11). Weak binders required low concentrations of dsDNA probes in the solution to prevent loss of fluorescence over time. Furthermore, cells expressing moderate to strong binders show little loss of fluorescence intensity over long durations of time, a feature that greatly facilitates ease of imaging. Next we tested each sZF for its ability to bind its own target dsDNA sequence and also target dsDNA sequences corresponding to the other zinc finger proteins, i.e., a total of 16x16 interactions were probed to generate a complete cross reactivity profile (Fig. 6b). We found that, while most zinc fingers bound their target dsDNA specifically, some showed a significant degree of cross-reactivity (ZFs 1, 8, 13). The strong ZF binders were particularly susceptible to this phenomenon. Interestingly, almost all the zinc fingers were observed to bind the ZF16 target dsDNA, likely in part to the high poly-G rich content of this sequence. Based on the above ZFs 2, 3, 4, 5, 6, 7, 10, 12, 14, 15 were found to be orthogonal to each other and were moderate to strong binders and thus good candidates for barcoding cells. ZFs 9, 11 while also orthogonal to others were weak binders.

Example 4. Use of zinc finger binding domains to barcode cells with a single tag

If sZFs are to serve as faithful barcodes they must also enable differential labeling of cells in complex mixtures. To investigate this, we designed experiments to image cell populations comprising different batches of cells transfected with plasmids such that each expressed a sZF targeted to bind to a different dsDNA probe. Solutions of fluorescently-labeled dsDNA probes corresponding to the different ZFs were then applied to the live cells along with calibrated amounts of blocking DNA, followed by imaging for each fluorophore. Specifically, cells expressing either of sZF1, sZF2, sZF3, or sZF4 were mixed in pairs (sZF1+sZF2; and sZF3+sZF4) or in a pool of three (sZF1+sZF2+sZF3), and were probed using appropriate combinations of Alexa488, Alexa546, and Alexa647 labeled target dsDNA molecules. Qualitative inspection of the images shows that these cells bind one labeled oligonucleotide probe and not the others (Fig. 5a-c), confirming the sZF-dsDNA interactions are sequence specific. Because three colors can be used to label up to seven distinct cell types, we also performed an experiment where cells bearing all possible combinations of sZF1, sZF3 and sZF4 were mixed and stained. Singly, doubly, and triply stained cells can be visually distinguished easily in the resulting images showing that seven distinct cell types can be successfully labeled in this manner (Fig. 5c).

Next, to obtain quantitative measures of the specificity of binding of sZFs to their corresponding oligos in the cell mixture experiments described above, we analyzed images computationally to identify regions occupied by cells and computed measures that assessed the exclusivity of the fluorescence signals in these cells associated with each sZF / oligo pair. The analysis was incorporated into a set of three interactive MatLab (The MathWorks, Natick, MA) applications. Briefly (workflow in Fig. 7), using the MatLab applications, we assembled the scans for each fluorophore into a normalized RGB image, masked out image regions containing dead cells and debris, and set intensity thresholds in each channel to segment the image into regions identifiable as living cells. Because this method of identifying cell regions results in intensities that will be high in at least one channel, which could potentially bias measures of exclusivity based on direct comparison of intensities across channels, we instead examined correlations between intensities in cell regions and also analyzed the extent to which cell segments in one channel overlapped cell segments in another. Pearson correlation coefficients between intensities of pixels in cell regions derived from pairs of channels were negative with $P < .001$ in 20 out of 21 comparisons, indicating that within cells, signals from the two sZF / oligo pair were strongly exclusive. By contrast, when other areas of the image were included, correlations were strongly positive, indicating that oligo signals were non-exclusive and, indeed, positively associated in the image field as a whole (Fig. 8). Regarding segment overlaps, an average of 47.1 segments were found per channel per image, and 38.1 (32.3) of these, or 81% (65%), contained no pixels from segments of another channel in two-channel RG images (three-channel RGB images) (Figs. 9a-b, 10, 11). At an average size of ~250 pixels each, the likelihood of finding even one

segment region completely without pixels reaching the intensity threshold of the other channel is ~ 0.055 , so that the probability of finding ~ 30 of them is $\sim 1.4e-38$. When we inspect two or three dimensional scatterplots of mean intensities of R, G, and B segment regions, where segments in one channel with overlaps with segments in others (“ovlp”) are distinguished from segments in a channel with no such overlaps (“non-ovlp”) (Fig. 5a, 12a-b). Segments (especially non-ovlp segments) were observed to have high intensity in the channel from which they were analyzed and low intensities in the other channels, confirming again that oligos bind strongly to cells expressing their corresponding ZF but not to other cells. Finally, two dimensional pixel intensity histograms that distinguish non-ovlp regions from ovlp-regions in the image show that individual pixels as well as mean pixel intensities in non-ovlp regions are intense in the channel from which they were segmented and of very low intensity in the other, while pixels in overlap regions show no mutual intensity relationship (Fig. 12a-b, 13). Together these results establish that sZFs accurately barcode cells in both simplex and multiplex settings.

Example 5. Use of zinc finger binding domains to barcode cells with multiple tags applied sequentially

For certain applications, the ability to re-probe a cell with different labels or functional tags in a sequential manner (which is not feasible using fluorescent proteins) is also desired. Because the zinc-finger dsDNA interaction is a non-covalent interaction it should be feasible to displace the latter using a competing dsDNA ligand. Towards this, we examined the dsDNA dissociation kinetics from sZFs. In general assayed sZFs were observed to have a high affinity for their target dsDNA thus demonstrating low rates of dissociation that enabled long-term visualization of the tagged cells. To promote dissociation of bound dsDNA we used high concentrations of a non-fluorescent target dsDNA in solution that resulted in a rapidly diminishing fluorescence signal over time (Fig. 14). Thus while the high affinity of sZFs to their target dsDNA greatly facilitates ease of imaging, bound probes can also be actively displaced and hence sZFs enable dynamic re-probing of cells.

In this study we have investigated 16 sZFs, of which up to 10 are orthogonal zinc fingers suitable for barcoding cells. However using standard labeling techniques and 3-color imaging microscopes it is not feasible to identify thus barcoded cells in mixtures comprising >3 distinct cells types. To address this issue we devised a novel sequential imaging approach. In summary, each sZF has a corresponding probe comprising two parts: a dsDNA portion that specifically binds the zinc finger protein, and a single-stranded portion that is designed to include several hybridization sites (Fig. 15a-b). These hybridization sites provide a unique sequence code for the sZF, which is decoded by probing the sites sequentially as follows: in step 1, a fluorophore tagged complementary oligonucleotide is hybridized to its target site enabling a first fluorescence readout; in step 2, two adjacent complementary oligonucleotides are annealed, the first bearing a quencher that suppresses the

step 1 fluorescence signal, and the second bearing another fluorophore that enables a second fluorescent readout and so on. Thus each sZF is encoded by a sequence of fluorescent states that are progressively read by turning each state on and then off. Extending this scheme to n steps enables barcoding of 3^n cell types using just 3 fluorophores. This approach is fast, does not use enzymes or chemical reactions, and is compatible with use on live cells. A basic demonstration of the scheme in a simplex setting is provided in Fig. 15b where sZF expressing cells are sequentially probed - each sZF identity here is encoded by two colors, for instance sZF2 by green in step 1 and red in step 2, sZF3 by red in step 1 and blue in step 2 and similarly for sZFs 6, 12, 14 and 15. Fig. 15c provides a demonstration of this labeling approach in a multiplex setting where the above six sZFs are individually expressed in cells that are subsequently mixed and then imaged. In these experiments, in addition to the step specific quencher and fluorescent oligos that hybridize to the single-stranded portions of the bound sZF probes, we also freshly re-probed the sZFs at each step. This compensated for loss of fluorescence signal due to dissociation of dsDNA probes from the zinc fingers in the time interval between imaging steps (see also Fig. 14a-b). Our results in Fig. 15c confirm that such a sequential tagging scheme can successfully identify the various constituent cells in complex mixtures of >3 barcoded cell types using just 3 resolvable fluorophores.

Example 6. Use of zinc finger binding domains as reporters of endogenous cellular activity

We next investigated whether sZFs could also serve as surrogate reporters of endogenous cellular activity. To this end, lentiviral vectors with small molecule (tetracycline and cumate) inducible promoters to drive sZF expression were constructed. Stable transductions of 293T and HeLa cells were performed, and upon small molecule induction sZF expression could be readily detected by the ability of the cells to bind dsDNA molecules (Fig. 16). Expression of sZFs from the tet responsive promoters was observed to be higher than from the cumate inducible promoters (Fig. 17), but both inducible systems showed robust induction and can be used as versatile tools for barcoding cells. Together we conclude that like fluorescent proteins, sZFs in addition to labeling cells can also serve as reporters of transcriptional activity in multiple cell types.

All patents, patent applications, and publications, including [http://blanco.biomol.uci.edu/membrane_proteins_xtal.html] and [https://modbase.compbio.ucsf.edu/projects/membrane/], mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent, patent application, or publication was specifically and individually indicated to be incorporated by reference.

What is claimed is:

1. A protein comprising (a) a polynucleotide-binding domain; and (b) a transmembrane domain.
2. The protein of claim 1, wherein said polynucleotide-binding domain is a programmable DNA
5 binding domain.
3. The protein of claim 1 or 2, wherein said polynucleotide-binding domain is a zinc finger protein or TAL effector protein.
- 10 4. The protein of claim 2 or 3, wherein said polynucleotide-binding domain is a C₂H₂ triple finger.
5. The protein of claim 2 or 3, wherein said zinc finger is a synthetic zinc finger such as a zinc finger selected from the group consisting of Z1, Z2, Z3, Z4, Z5, Z6, Z7, Z8, Z9, Z10, Z11, Z12, Z13,
15 Z14, Z15, and Z16.
6. The protein of any of claims 1-5, wherein said protein comprises an Ig κ-chain leader sequence at its N-terminus.
- 20 7. The protein of any of claims 1-6, wherein said protein comprises an import sequence at its C-terminus.
8. The protein of any of claims 1-7, wherein said transmembrane domain is C-terminal of the polynucleotide-binding domain.
25
9. A protein of any of claims 1-8, wherein said protein is a nanopore subunit.
10. A protein complex comprising one or more proteins of any of claims 1-9.
- 30 11. The complex of claim 10, wherein said complex is a nanopore.
12. The nanopore of claim 11 selected from the group consisting of alpha-hemolysin (αHL) and mycobacterium smegmatis porin A (MspA).
- 35 13. A membrane comprising one or more variants of the nanopore of claim 11 or 12.

14. A cell comprising one or more variants of the protein or protein complex of any of claims 1-10, wherein the polynucleotide-binding domain(s) of said protein(s) or protein complex(es) are on the surface of said cell in a manner capable of interaction with materials external to said cell.

15. A method of sequencing a polynucleotide, said method comprising:

(a) providing two separate pools of liquid containing electrically conductive medium and a nanopore-perforated interface between said two pools, wherein said interface contains a nanopore of claim 11 or 12;

(b) providing candidate polynucleotide molecules in one of said pools;

(c) applying a voltage differential across said pools; and

(d) making interface-dependent measurements of ionic current over time as individual nucleotides of a single polynucleotide interact sequentially with the interface, yielding data suitable to determine a nucleotide-dependent characteristic of the polynucleotide.

16. The method of claim 15, wherein said interface is a membrane.

17. A method of tagging a cell comprising the steps:

(a) providing a cell of claim 14;

(b) providing a polynucleotide probe comprising a detectable label and capable of interacting with said surface polynucleotide-binding domain(s);

(c) contacting said cell with said polynucleotide probe; and

(d) detecting said polynucleotide probe.

18. The method of claim 17, wherein said detectable label is a fluorescent label.

19. The method of claim 17, wherein said detectable label is a luminescent label, quenching label, biotin label, ligation label, chemically reactive moiety such as azide, chemically reactive moiety that enables click-chemistry, detectable small molecule, or the sequence of said polynucleotide probe.

20. A method of labeling and sequentially relabeling cells, said method comprising the steps:

(a) providing a cell of claim 14;

(b) providing a polynucleotide probe comprising a detectable label that is capable of interacting with said surface polynucleotide-binding domain(s);

(c) contacting said cell with said polynucleotide probe;

(d) detecting said polynucleotide probe;

(e) dissociating said polynucleotide probe from said polynucleotide-binding domain(s);

- (f) providing a subsequent polynucleotide probe comprising a detectable label;
- (g) contacting said cells with said subsequent polynucleotide probe; and
- (h) detecting said subsequent polynucleotide probe.

5 21. The method of claim 20, wherein one or more of said detectable labels is a fluorescent label.

22. The method of claim 20, wherein one or more of said detectable labels is a luminescent label, quenching label, biotin label, ligation label, chemically reactive moiety such as azide, chemically reactive moiety that enables click-chemistry, detectable small molecule, or the sequence of said
10 polynucleotide probe.

23. The method of any of claims 20-22, wherein said dissociation is achieved by the addition of a high concentration of unmarked polynucleotide(s) capable of displacing said polynucleotide probe.

15 24. The method of any of claims 20-23, further comprising repeating one or more times the steps (e) to (h) with an additional polynucleotide probe or probes.

25. A method of labeling a cell, said method comprising the steps:

- (a) providing a cell of claim 14;
- 20 (b) providing a scaffold polynucleotide to mediate the interaction of said surface polynucleotide binding domain(s) with one or more polynucleotide probe(s), said scaffold polynucleotide comprising (i) a double-stranded domain and (ii) a single-stranded domain comprising one or more sites capable of hybridization, wherein said double-stranded domain interacts with said surface polynucleotide-binding domain(s) and said single-stranded domain is available to interact with one or more polynucleotide
25 probes;
- (c) providing a polynucleotide probe comprising a detectable label and capable of interacting with said scaffold polynucleotide;
- (d) contacting said cell with said scaffold polynucleotide and said polynucleotide probe;
- (e) detecting said polynucleotide probe.

30

26. The method of claim 25, wherein said detectable label is a fluorescent label.

27. The method of claim 25, wherein said detectable label is a luminescent label, quenching label, biotin label, ligation label, chemically reactive moiety such as azide, chemically reactive moiety that
35 enables click-chemistry, detectable small molecule, or the sequence of said polynucleotide probe.

28. A method of labeling and sequentially relabeling cells, said method comprising the steps:

(a) providing a cell of claim 14;

(b) providing a scaffold polynucleotide to mediate the interaction of said surface polynucleotide binding domain(s) with one or more polynucleotide probe(s), said scaffold polynucleotide comprising (i) a double-stranded domain and (ii) a single-stranded domain comprising one or more sites capable of hybridization, wherein said double-stranded domain interacts with said surface polynucleotide-binding domain(s) and said single-stranded domain is available to interact with one or more polynucleotide probes;

(c) providing a polynucleotide probe comprising a detectable label and capable of interacting with said scaffold polynucleotide;

(d) contacting said cell with said scaffold polynucleotide and said polynucleotide probe;

(e) detecting said polynucleotide probe;

(f) providing a signal-quenching polynucleotide that comprises (i) a polynucleotide sequence complementary to a region sufficiently adjacent to said polynucleotide probe and (ii) a mechanism of quenching the signal of said polynucleotide probe label;

(g) contacting said cells with one or more of said quenching polynucleotide probe;

(h) providing a subsequent polynucleotide probe comprising a detectable label and complementary to an unbound segment of said scaffold polynucleotide;

(i) contacting said cells with said subsequent polynucleotide probe; and

(j) detecting said subsequent polynucleotide probe.

29. The method of claim 28, wherein said labels are fluorescent labels and said mechanism of quenching is a mechanism of quenching fluorescence.

30. The method of claim 28, wherein one or more of said detectable labels is a luminescent label, quenching label, biotin label, ligation label, chemically reactive moiety such as azide, chemically reactive moiety that enables click-chemistry, detectable small molecule, or the sequence of said polynucleotide probe.

31. The method of claim 28-30, further comprising the step:

(k) repeating one or more times the steps (f) to (j) with additional quenching polynucleotide(s) and polynucleotide probe(s).

32. The method of claim 17-31 wherein said cells are fixed or are growing adherently.

33. The method of claim 17-31 wherein said cells are in solution.

34. The method of any of claims 17-33, wherein expression of said polynucleotide-binding domain is under control of a promoter and the level of labeling is indicative of promoter activity.

35. A method for organizing cells, said method comprising the steps:

- (a) providing a solid surface to which one or more polynucleotides are affixed in a defined manner;
- (b) providing a cell or cells comprising a surface polynucleotide-binding domain capable of directly binding said affixed polynucleotides; and
- (c) contacting said solid surface with said cells, thereby affixing said cells.

36. The method of claim 35, wherein said affixed cells comprise an additional surface polynucleotide-binding domain, and said method further comprises the steps:

- (d) providing a new cell or cells comprising surface polynucleotide-binding domains;
- (e) providing one or more polynucleotide(s) capable of tethering said affixed and said new cells in a sequence-specific manner; and
- (f) contacting said tethering polynucleotide(s) with said affixed cells and said new cells.

37. The method of claim 36, wherein said tethering polynucleotide is a double-stranded polynucleotide probe comprising regions capable of being bound by said surface polynucleotide-binding domains of both said affixed cells and said new cells.

38. The method of claim 36, wherein said tethering is mediated by two polynucleotides, each comprising (i) a double-stranded domain capable of interacting with a surface polynucleotide-binding domain and (ii) a single-stranded domain, and wherein one tethering probe's said double-stranded domain interacts directly with said affixed cells, the other tethering probe's said double-stranded domain interacts directly with said new cells, and the single-stranded domains of the two tethering probes are complementary so as to tether the cells.

39. The method of any of claims 35-38, further comprising the step:

- (h) repeating steps (d)-(f) one or more times with additional cells and/or tethering polynucleotides.

40. A method for organizing cells, said method comprising the steps:

- (a) providing a solid surface to which one or more polynucleotides are affixed in a defined manner;
- (b) providing a cell or cells comprising a surface polynucleotide-binding domain;
- (c) providing a scaffolding polynucleotide capable of binding said cell surface polynucleotide-binding domain, said polynucleotide probe comprising (i) a double-stranded domain and (ii) a single-

stranded domain, wherein said double-stranded domain interacts with said surface polynucleotide-binding domain(s) and said single-stranded domain is available to interact with one or more of said affixed polynucleotides; and

(d) contacting said solid surface with said cells, thereby affixing said cells.

5

41. The method of claim 40, wherein said affixed cells comprise an additional surface polynucleotide-binding domain, and said method further comprises the steps:

(e) providing a new cell or cells comprising surface polynucleotide-binding domains;

(f) providing one or more polynucleotide(s) capable of tethering said affixed and said new cells in a sequence-specific manner; and

10

(g) contacting said tethering polynucleotide(s) with said affixed cells and said new cells.

42. The method of claim 41 wherein said tethering polynucleotide is a double-stranded polynucleotide probe comprising regions capable of being bound by said surface polynucleotide-binding domains of both said affixed cells and said new cells.

15

43. The method of claim 41, wherein said tethering is mediated by two polynucleotides, each comprising (i) a double-stranded domain capable of interacting with a surface polynucleotide-binding domain and (ii) a single-stranded domain, and wherein one tethering probe's said double-stranded domain interacts directly with said affixed cells, the other tethering probe's said double-stranded domain interacts directly with said new cells, and the single-stranded domains of the two tethering probes are complementary so as to tether the cells.

20

44. The method of any of claims 41-43, further comprising the step:

25

(h) Repeating steps (e)-(g) one or more times with additional cells and/or tethering polynucleotides.

45. A method of tethering free cells, said method comprising the steps:

(a) providing two or more free cells comprising one or more surface polynucleotide-binding domains;

(b) providing one or more tethering polynucleotides capable of tethering two of said free cells; and

30

(c) contacting said one or more polynucleotides with said free cells.

46. The method of claim 45, wherein said tethering polynucleotide is a double-stranded polynucleotide probe comprising regions capable of being bound by said surface polynucleotide-binding domains of both said affixed cells and new cells.

35

47. The method of claim 45, wherein said tethering is mediated by two polynucleotides, each comprising (i) a double-stranded domain capable of interacting with a surface polynucleotide-binding domain of said free cells and (ii) a single-stranded domain, wherein the single-stranded domains of the two tethering probes are complementary so as to tether said free cells.

5

48. The method of any of claims 45-47, further comprising the step:
(d) Repeating steps (a)-(c) one or more times with additional cells and/or tethering polynucleotides.

49. An artificial tissue produced by the method of any of claims 35-48.

10

50. The artificial tissue of claim 48, wherein said tissue is vascular tissue or an organoid.

51. A polynucleotide encoding the protein of any of claims 1-9.

15

52. The polynucleotide of claim 51, operably linked to a promoter.

53. The polynucleotide of claim 51 or 52, wherein said polynucleotide is a free, linear unit.

54. A plasmid comprising the polynucleotide of claim 51 or 52.

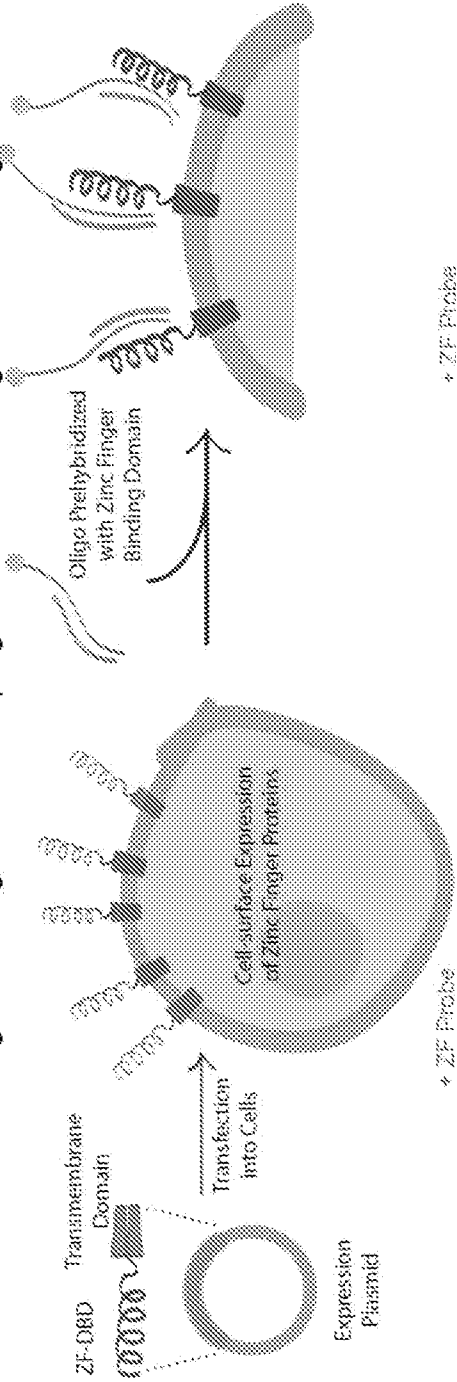
20

55. A viral vector comprising the polynucleotide of claim 51 or 52.

56. A cell comprising the polynucleotide of claim 53, the plasmid of claim 54, or the vector of claim 55.

25

Fig. 2 Scheme for barcoding cells using cell surface programmable zinc-finger DNA binding domains



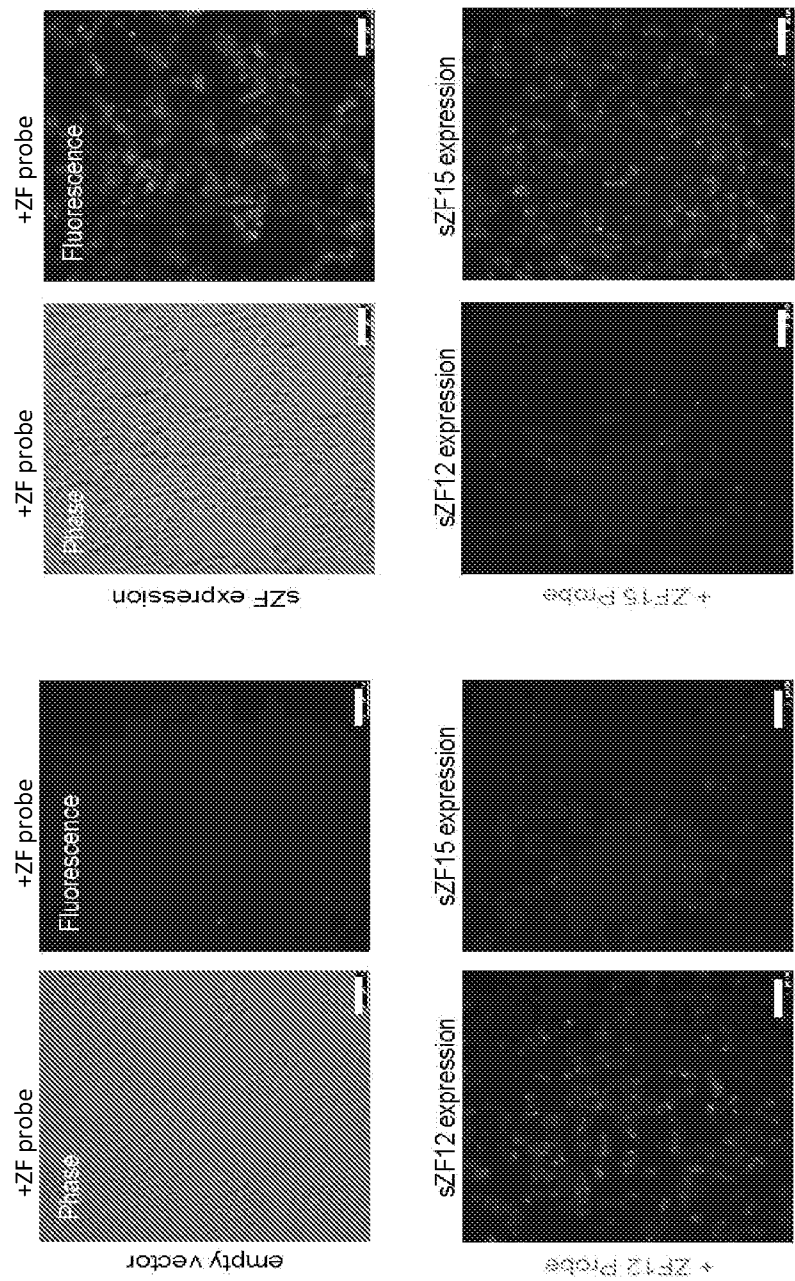


Fig. 3

Fig. 4

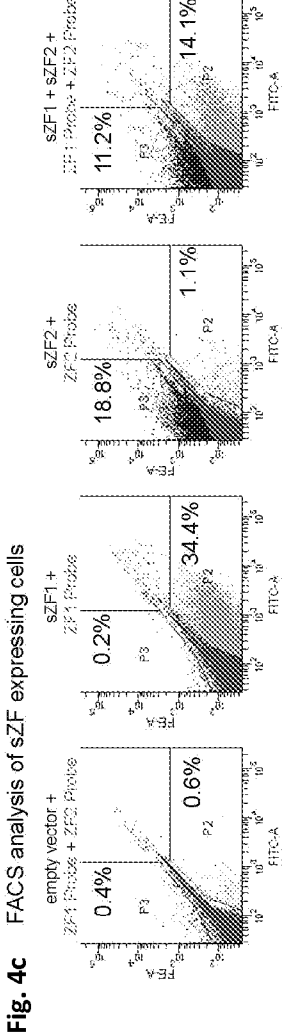
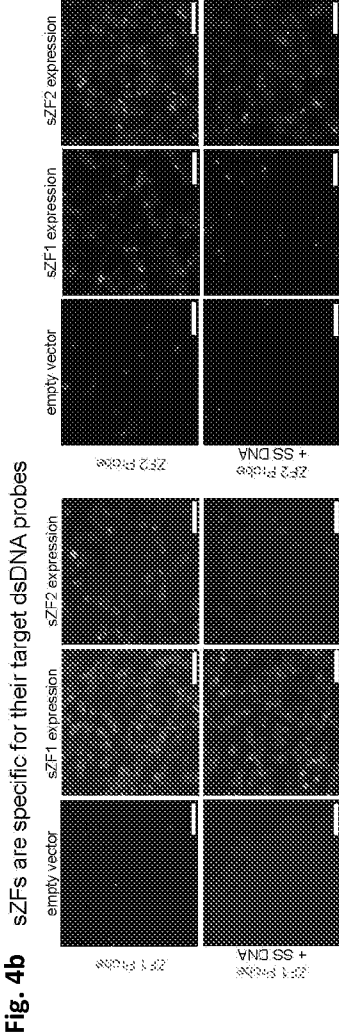
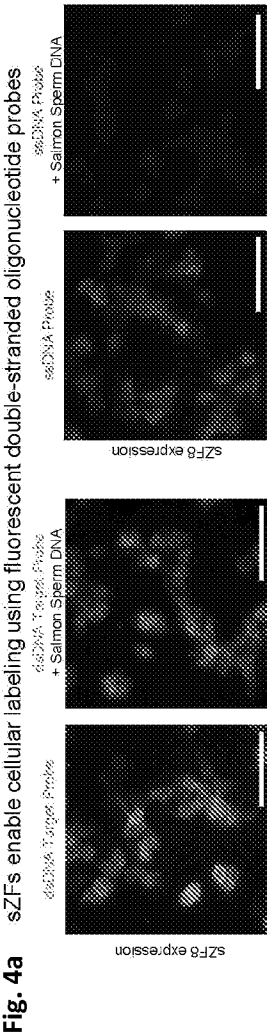


Fig. 5a sZF specific multiplex labeling using dsDNA probes
sZF1 + sZF2 + ZF1 Probe + ZF2 Probe

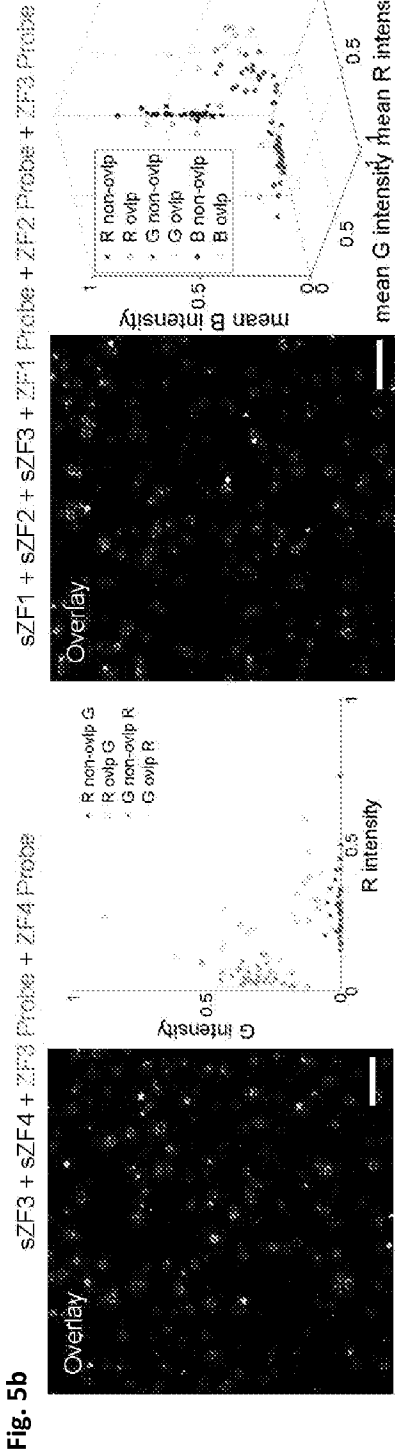
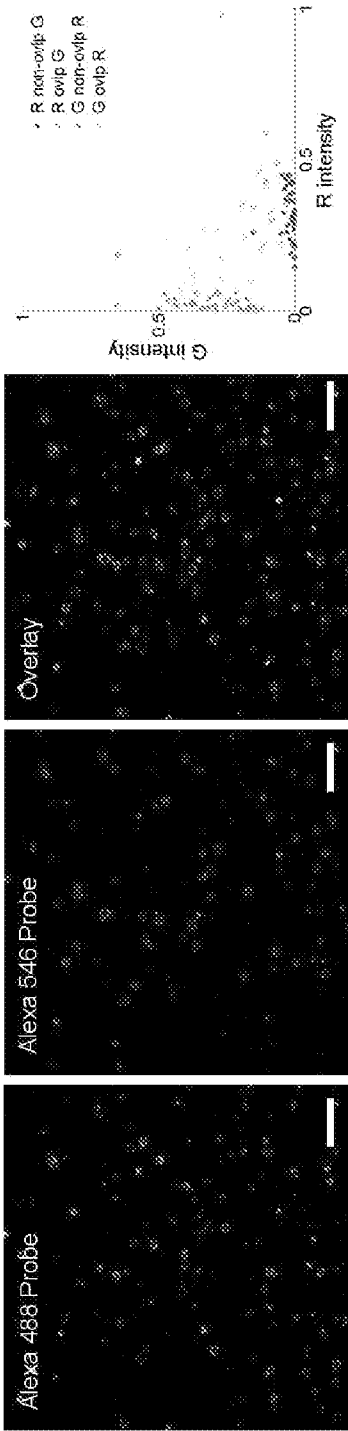


Fig. 5c Multiplex multicolor imaging using combinatorial expression of sZFs

sZF(1, 3, 4, 1+3, 3+4, 4+1, 1+3+4) + ZF1 Probe + ZF3 Probe + ZF4 Probe

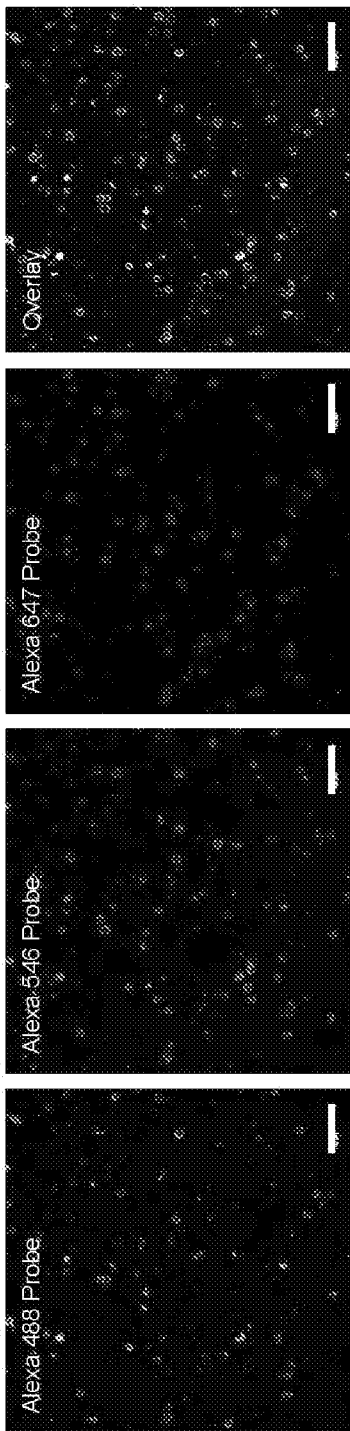


Fig. 6

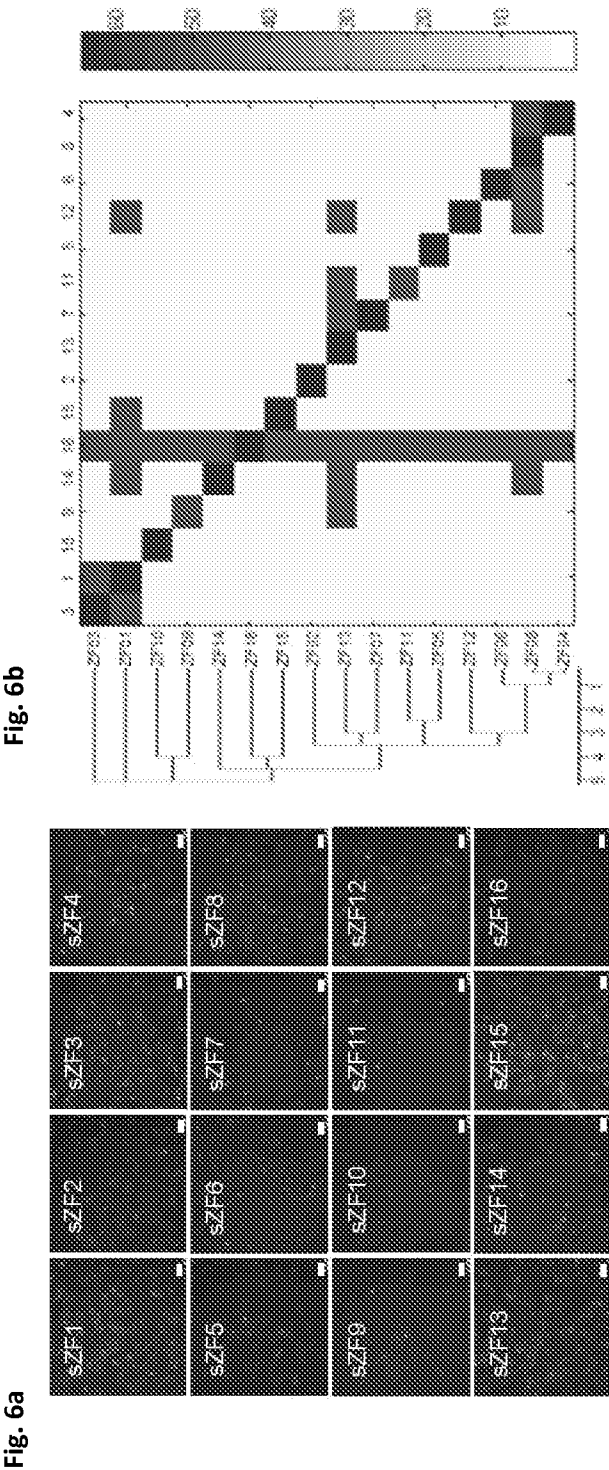


Fig. 7

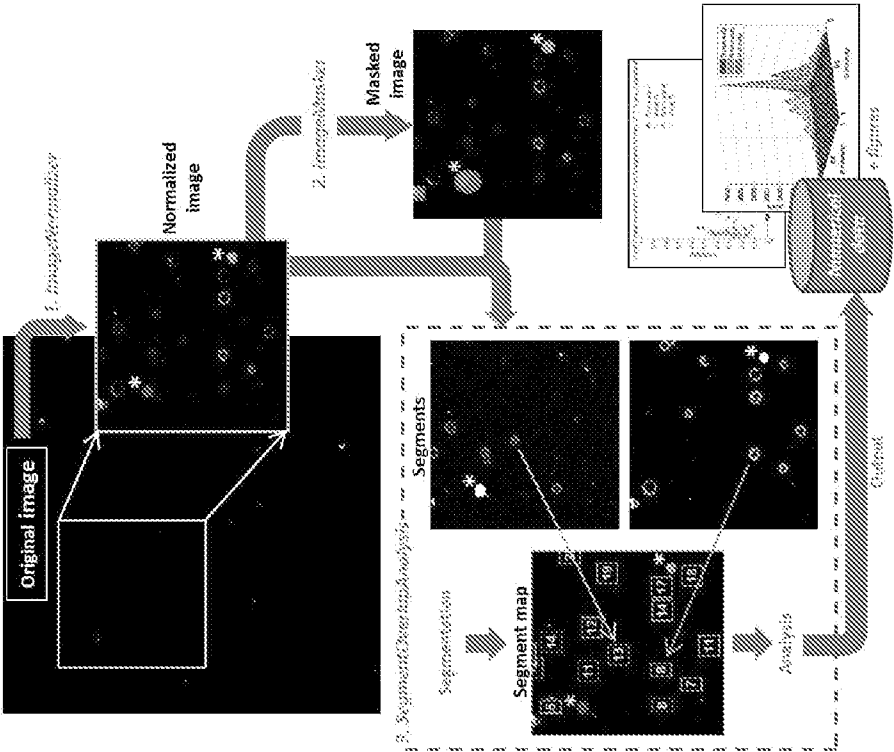


Fig. 8

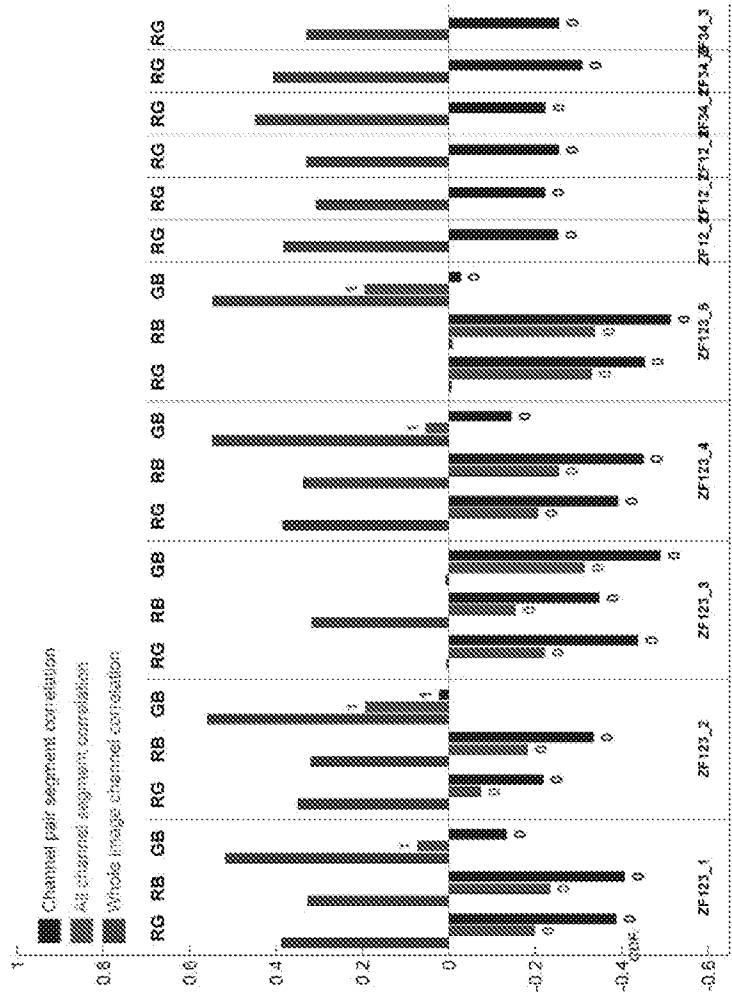


Fig. 9

Fig. 9a Image analysis results: Counts and fractions of segments and NOVS formed per channel aggregated over all images and channels.

	mean	{std}	min	max
Number segments	47.1	(9.8)	33	71
Number NOVS (1 channel)	38.2	(7.9)	27	57
Fraction NOVS (1 channel)	0.81	(0.08)	0.67	0.94
Number NOVS (2 channel)	32.3	(6.9)	24	48
Fraction NOVS (2 channel)	0.65	(0.09)	0.61	0.71

Fig. 9b Image analysis results: Overall segment and NOVS counts and fractions from each channel in each image.

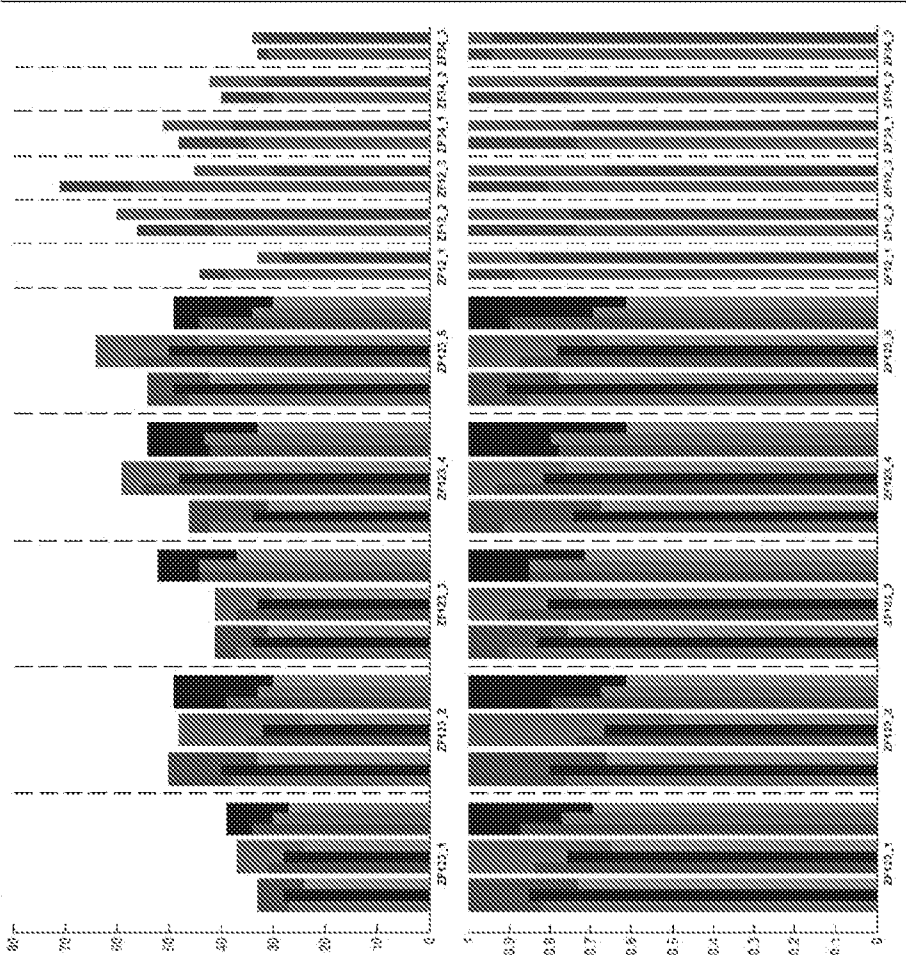


Fig. 10

Image analysis results: Segment area size distributions for all images, presented as box plots.

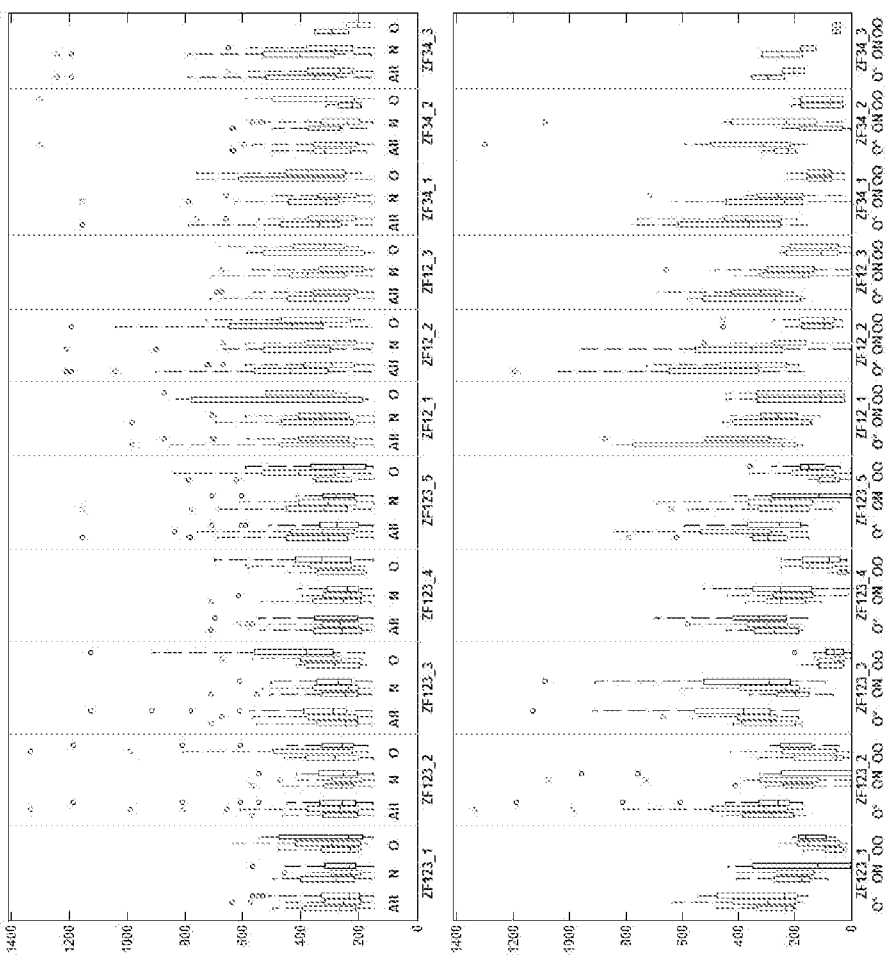


Fig. 11

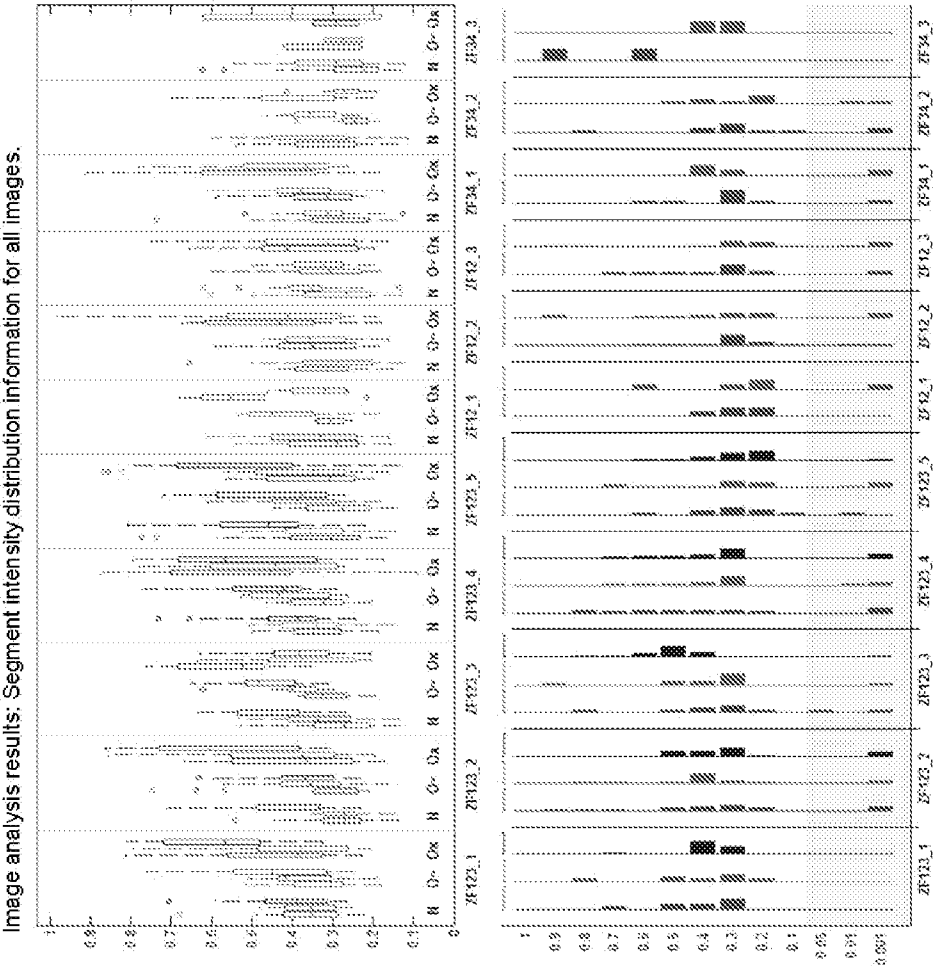


Fig. 12

Fig. 12a Image analysis results: scatterplots of mean intensities, and pixel histograms

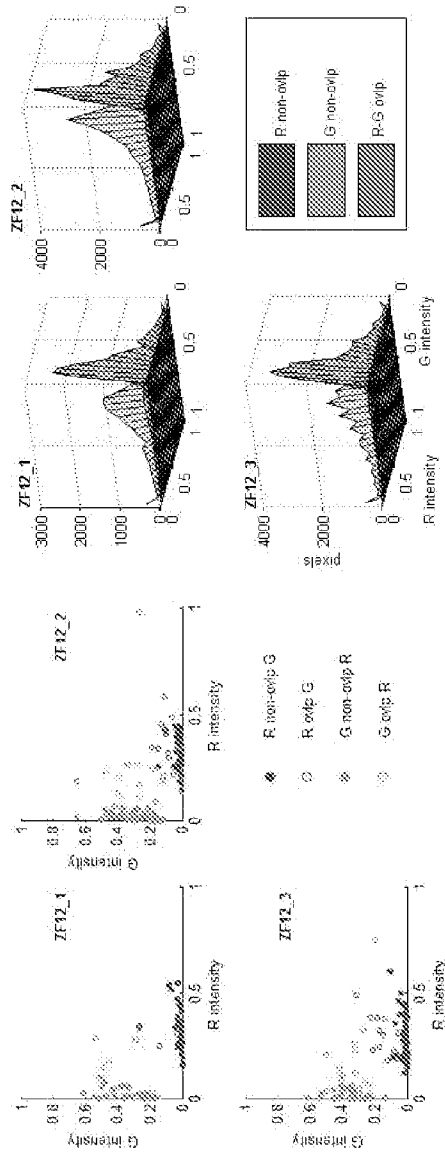


Fig. 12b Image analysis results: scatterplots of mean intensities, and pixel histograms

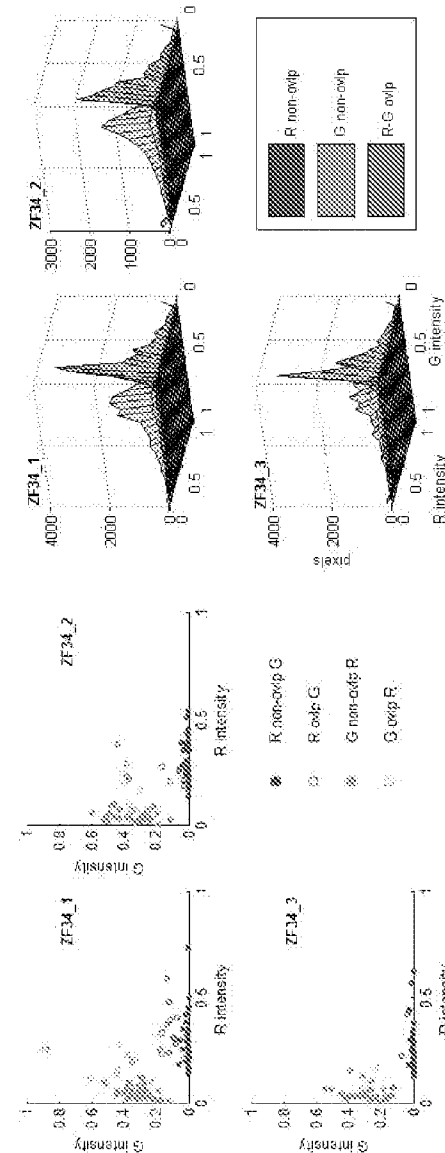


Fig. 13

Image analysis results: scatterplots of mean intensities, and pixel histograms

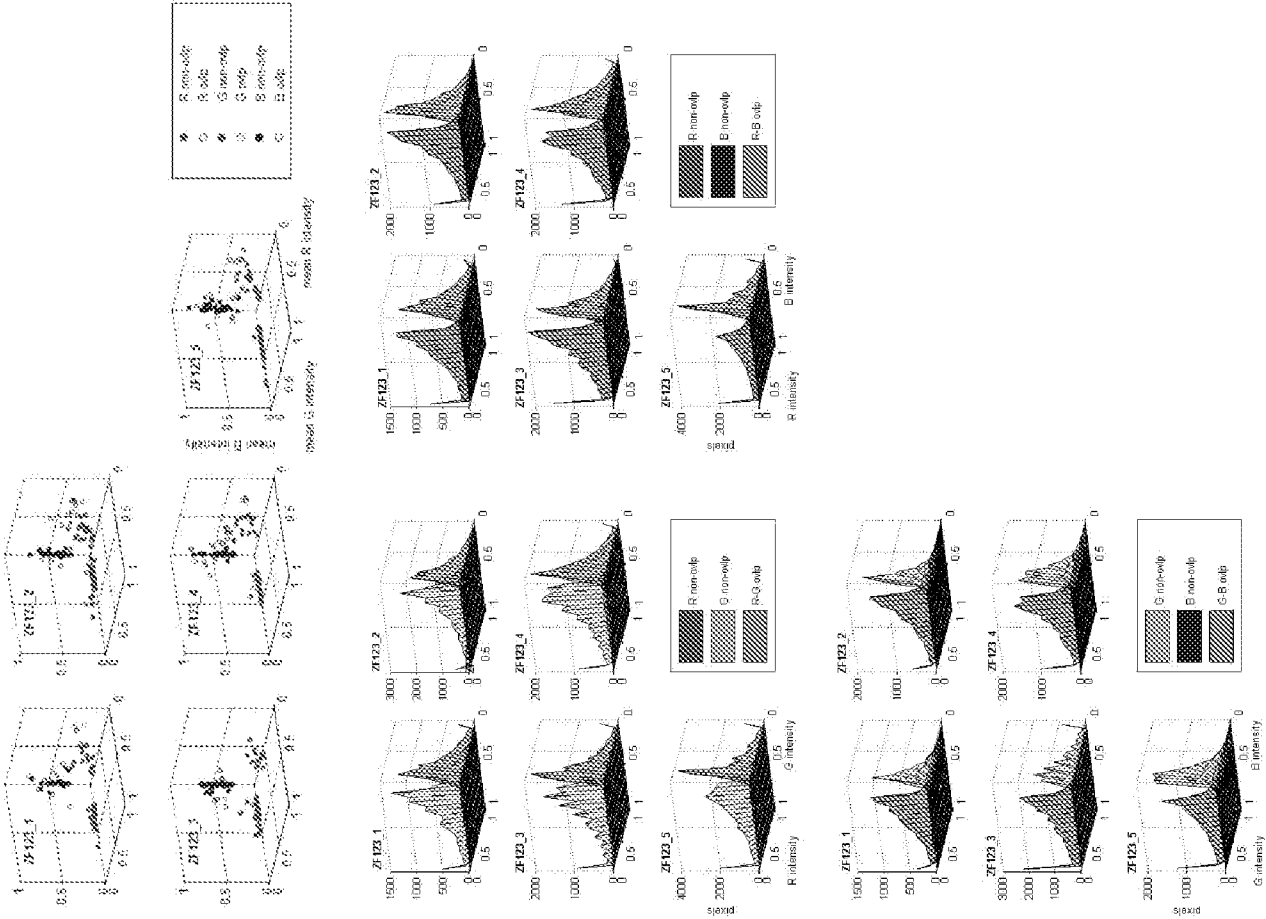


Fig. 14

Fig. 14a sZF target dsDNA displacement kinetics

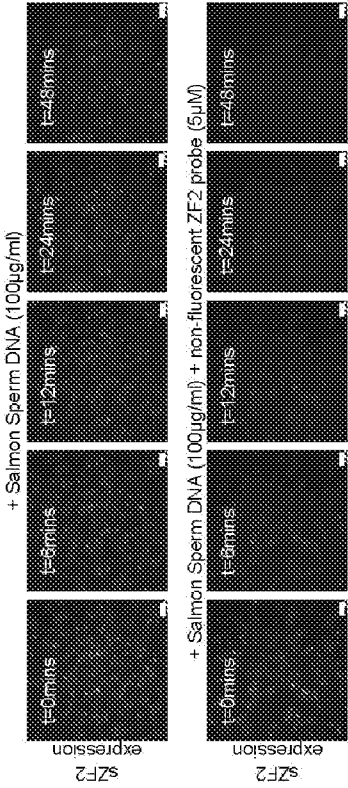


Fig. 14b

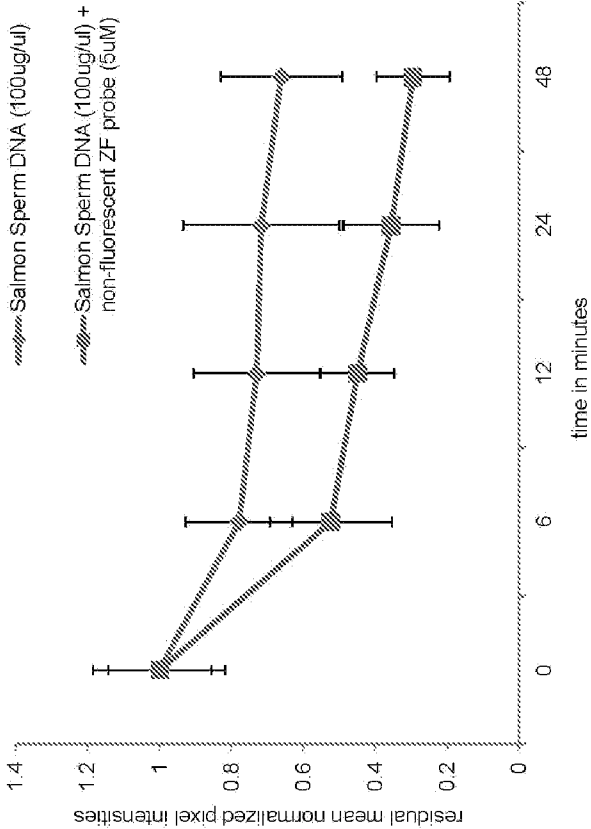


Fig. 15

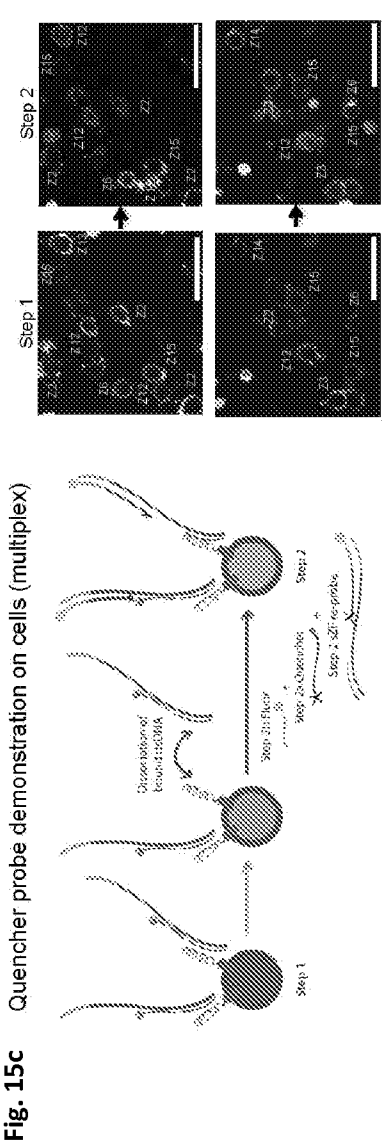
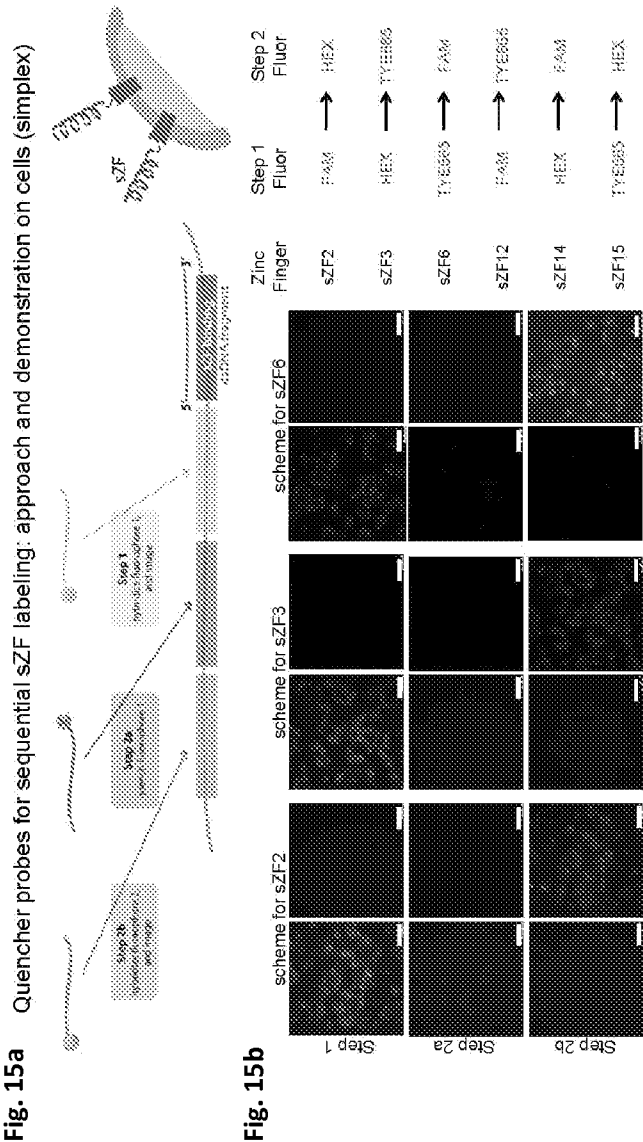


Fig. 16

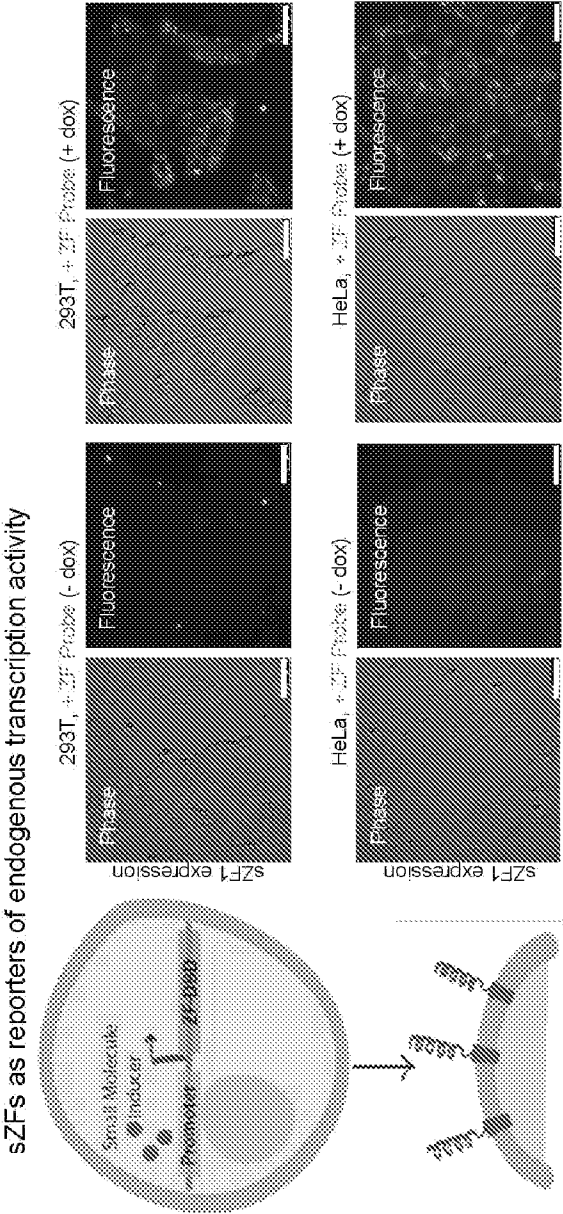


Fig. 17

. Inducible sZF expression

