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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
04 January 2024 (04.01.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/006972 A1

(51) International Patent Classification:
A61K 31/395 (2006.01) *A61K 31/015* (2006.01)
A61K 31/025 (2006.01) *A61K 31/33* (2006.01)

(21) International Application Number:
PCT/US2023/069476

(22) International Filing Date:
30 June 2023 (30.06.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/357,614 30 June 2022 (30.06.2022) US

(71) Applicant: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US).

(72) Inventors: **GOLUB, Todd**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **BONDESON, Daniel**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US).

(74) Agent: **SCHER, Michael B.** et al.; Johnson, Marcou, Isaacs & Nix, LLC, Customer No. 95280, PO Box 691, Hoschton, Georgia 30548 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE,

(54) **Title:** THERAPEUTIC TARGETING OF INOSITOL PYROPHOSPHATE SYNTHESIS IN CANCER

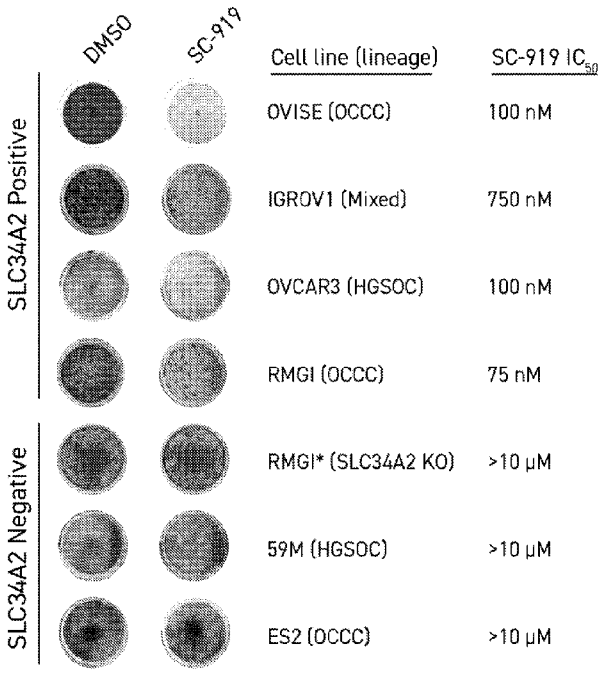


FIG. 8

(57) **Abstract:** The subject matter disclosed herein is generally directed to treating cancers sensitive to phosphate dysregulation with inhibitors of inositol pyrophosphate (PP-InsP) synthesis, in particular, inhibitors of inositol hexakisphosphate kinases IP6Ks.

WO 2024/006972 A1

[Continued on next page]

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))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

THERAPEUTIC TARGETING OF INOSITOL PYROPHOSPHATE SYNTHESIS IN CANCER

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 63/357,614 filed June 30, 2022. The entire contents of the above-identified application are hereby fully incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant Nos. CA242457 and CA212229 awarded by the National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0003] This application contains a sequence listing filed in electronic form as an xml file entitled BROD-5630WP_ST26.xml, created on June 28, 2023, and having a size of 17,921 bytes. The contents of the electronic sequence listing is herein incorporated by reference in its entirety.

TECHNICAL FIELD

[0004] The subject matter disclosed herein is generally directed to treating cancers that are sensitive to phosphate dysregulation with inhibitors of inositol pyrophosphate (PP-InsP) synthesis, in particular, inhibitors of inositol hexakisphosphate kinases IP6Ks.

BACKGROUND

[0005] Ovarian and uterine cancers are among the deadliest cancers that affect women, and while platinum-based chemotherapies and recently approved PARP inhibitors show efficacy for some patients (Lord, C. J. & Ashworth, A. PARP inhibitors: Synthetic lethality in the clinic. *Science* 355, 1152–1158 (2017); Fong, P. C. *et al.* Poly(ADP)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J. Clin. Oncol.* 28, 2512–2519 (2010); and Ledermann, J. *et al.* Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. *N. Engl. J. Med.* 366, 1382–1392 (2012)), outcomes in these cancers have not improved greatly in the past twenty years

(Lheureux, S., Braunstein, M. & Oza, A. M. Epithelial ovarian cancer: Evolution of management in the era of precision medicine. *CA Cancer J. Clin.* 69, 280–304 (2019); and Vaughan, S. *et al.* Rethinking ovarian cancer: recommendations for improving outcomes. *Nat. Rev. Cancer* 11, 719–725 (2011)). Thus, new therapeutic options are needed.

[0006] Recently, Bondeson DP, Paolella BR, Asfaw A, et al. showed that phosphate dysregulation via the XPR1-KIDINS220 protein complex is a therapeutic vulnerability in ovarian cancer (Nat Cancer. 2022;10.1038/s43018-022-00360-7). Phosphate uptake is highly regulated and involves members of the SLC34 and SLC20 gene families (Virkki, et al. Phosphate transporters: a tale of two solute carrier families. *Am. J. Physiol. Renal Physiol.* 293, F643–54 (2007); Blaine, et al. Renal control of calcium, phosphate, and magnesium homeostasis. *Clin. J. Am. Soc. Nephrol.* 10, 1257–1272 (2015); and Sabbagh, *et al.* Intestinal npt2b plays a major role in phosphate absorption and homeostasis. *J. Am. Soc. Nephrol.* 20, 2348–2358 (2009)). Phosphate efflux is achieved via the only annotated human exporter - the Xenotropic and Polytropic Receptor 1 (XPR1) (Giovannini, D., Touhami, J., Charnet, P., Sitbon, M. & Battini, J.-L. Inorganic phosphate export by the retrovirus receptor XPR1 in metazoans. *Cell Rep.* 3, 1866–1873 (2013)). XPR1 is involved in phosphate homeostasis of various tissues (Ansermet, C. *et al.* Renal Fanconi Syndrome and Hypophosphatemic Rickets in the Absence of Xenotropic and Polytropic Retroviral Receptor in the Nephron. *J. Am. Soc. Nephrol.* 28, 1073–1078 (2017); Xu, X. *et al.* Murine Placental-Fetal Phosphate Dyshomeostasis Caused by an Xpr1 Deficiency Accelerates Placental Calcification and Restricts Fetal Growth in Late Gestation. *J. Bone Miner. Res.* 35, 116–129 (2020); and Legati, A. *et al.* Mutations in XPR1 cause primary familial brain calcification associated with altered phosphate export. *Nat. Genet.* 47, 579–581 (2015)). XPR1 is regulated by inositol pyrophosphates (InsP7/InspP8), which are synthesized by PPIP5K and IP6K enzymes (see, e.g., Shears SB. Inositol pyrophosphates: why so many phosphates?. Adv Biol Regul. 2015;57:203-216; Wilson MS, Jessen HJ, Saiardi A. The inositol hexakisphosphate kinases IP6K1 and -2 regulate human cellular phosphate homeostasis, including XPR1-mediated phosphate export. J Biol Chem. 2019;294(30):11597-11608; Wild R, Gerasimaite R, Jung JY, et al. Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains. Science. 2016;352(6288):986-990; and Li X, Gu C, Hostachy S, et al. Control of XPR1-dependent cellular phosphate efflux by InsP8 is an exemplar for functionally-exclusive inositol pyrophosphate signaling. Proc Natl Acad Sci USA. 2020;117(7):3568-3574).

[0007] IP6K inhibitors, through IP6K inhibitor-mediated suppression of cellular phosphate export, have been proposed to treat hyperphosphataemia, chronic kidney disease (CKD), cardiac failure, diabetes and the like (see, e.g., Moritoh Y, Abe SI, Akiyama H, et al. The enzymatic activity of inositol hexakisphosphate kinase controls circulating phosphate in mammals. Nat Commun. 2021;12(1):4847. Published 2021 Aug 11. doi:10.1038/s41467-021-24934-8; and International Patent Application publication number WO2018182051A1). It is unknown whether other diseases, such as cancer, can be treated by inhibiting inositol pyrophosphate synthesis or using IP6K inhibitors.

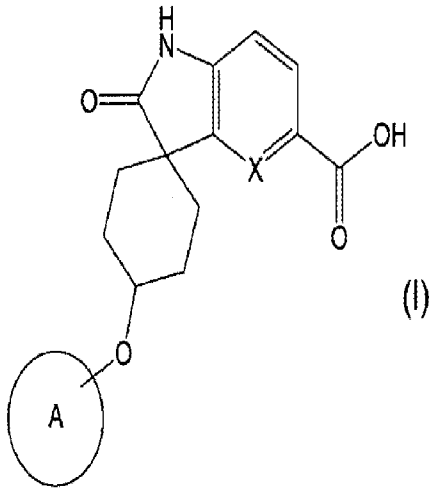
[0008] Citation or identification of any document in this application is not an admission that such a document is available as prior art to the present invention.

SUMMARY

[0009] In one aspect, the present invention provides for a method of treating cancer in a subject in need thereof comprising administering to the subject one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis. In certain embodiments, the cancer is characterized by high expression of SLC34A2 in tumor tissue as compared to expression in normal tissue.

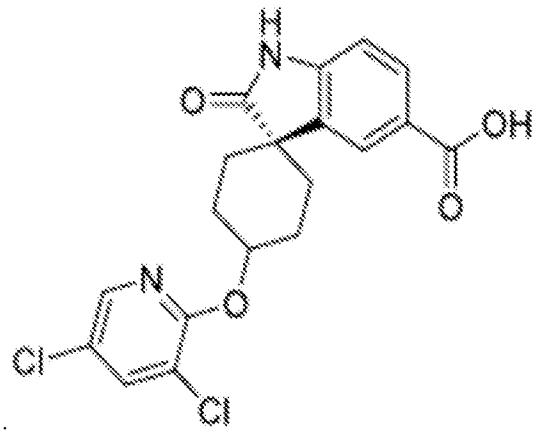
[0010] In another aspect, the present invention provides for a method of treating cancer in a subject in need thereof comprising: detecting tumors sensitive to phosphate dysregulation by detecting in the subject increased expression of SLC34A2 relative to a control, wherein if the subject has a tumor sensitive to phosphate dysregulation, including administration of one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis; if the subject does not have a tumor sensitive to phosphate dysregulation, administering a standard of care treatment that does not include administration of one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis.

[0011] In certain embodiments, the one or more therapeutic agents is capable of inhibiting one or more inositol hexakisphosphate kinases (IP6Ks) selected from the group consisting of IP6K1, IP6K2, and IP6K3. In certain embodiments, the one or more agents are capable of inhibiting enzymatic activity of IP6K1, IP6K2, and/or IP6K3 with an *in vitro* IC_{50} of less than 10 nM, preferably about 5 nM. In certain embodiments, the inhibitor is represented by the



(I)

[0001] formula: or a salt thereof, wherein ring A is an optionally substituted aromatic ring; X is CH or N. In certain embodiments, the inhibitor is



represented by the formula: or a salt thereof.

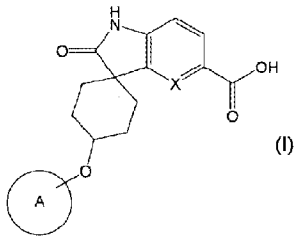
[0002] In certain embodiments, the one or more therapeutic agents is capable of inhibiting one or more diphosphoinositol pentakisphosphate kinases (PPIP5Ks) selected from the group consisting of PPIP5K1 and PPIP5K2.

[0003] In certain embodiments, the cancer is selected from the group consisting of ovarian cancer, uterine cancer, breast cancer, bile duct cancer, liver and lung cancer.

[0004] In certain embodiments, the method further comprises administering to the subject one or more therapeutic agents capable of inhibiting the suppression of SLC34A2.

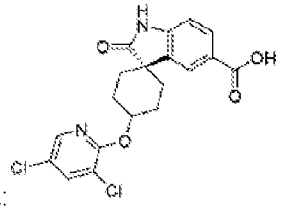
[0005] In certain embodiments, the one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis are co-administered within a standard of care treatment regimen. In certain embodiments, the standard of care treatment regimen comprises surgery and chemotherapy. In certain embodiments, the standard of care treatment regimen comprises administration of an immunotherapy or a PARP inhibitor. In certain embodiments, the immunotherapy is a checkpoint blockade therapy.

[0006] In another aspect, the present invention provides for a method for identifying a cancer sensitive to phosphate dysregulation, comprising: applying an IP6K inhibitor to a cancer cell or cell population; and detecting the inositol pyrophosphate level in the cell or cell population, wherein the cancer is sensitive if the inositol pyrophosphate level is decreased as compared to a control cell or population not treated with the inhibitor; or detecting the phosphate concentration in the cell or cell population, wherein the cancer is sensitive if the phosphate concentration is increased as compared to a control cell or population not treated with the inhibitor. In certain embodiments, the cancer cell or population is obtained or derived from a subject in need thereof.



(I)

formula: or a salt thereof, wherein ring A is an optionally substituted aromatic ring; X is CH or N. In certain embodiments, the inhibitor is represented by the formula:



or a salt thereof.

[0012] In certain embodiments, the one or more therapeutic agents is capable of inhibiting one or more diphosphoinositol pentakisphosphate kinases (PPIP5Ks) selected from the group consisting of PPIP5K1 and PPIP5K2.

[0013] In certain embodiments, the cancer is selected from the group consisting of ovarian cancer, uterine cancer, breast cancer, bile duct cancer, liver and lung cancer.

[0014] In certain embodiments, the method further comprises administering to the subject one or more therapeutic agents capable of inhibiting the suppression of SLC34A2.

[0015] In certain embodiments, the one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis are co-administered within a standard of care treatment regimen. In certain embodiments, the standard of care treatment regimen comprises surgery and chemotherapy. In certain embodiments, the standard of care treatment regimen comprises administration of an immunotherapy or a PARP inhibitor. In certain embodiments, the immunotherapy is a checkpoint blockade therapy.

[0016] In another aspect, the present invention provides for a method for identifying a cancer sensitive to phosphate dysregulation, comprising: applying an IP6K inhibitor to a cancer cell or cell population; and detecting the inositol pyrophosphate level in the cell or cell population, wherein the cancer is sensitive if the inositol pyrophosphate level is decreased as compared to a control cell or population not treated with the inhibitor; or detecting the phosphate concentration in the cell or cell population, wherein the cancer is sensitive if the phosphate concentration is increased as compared to a control cell or population not treated with the inhibitor. In certain embodiments, the cancer cell or population is obtained or derived from a subject in need thereof.

[0017] In another aspect, the present invention provides for a method for identifying an agent capable of inhibiting XPR1-mediated phosphate export, comprising: applying a candidate agent derived from an IP6K inhibitor to a cancer cell or cell population; and detecting modulation of phosphate efflux in the cell or cell population by the candidate agent, thereby identifying the agent.

[0018] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0020] **FIG. 1 – The SPX domain of XPR1 is a therapeutic target.** (top) Illustration showing that the SPX domain is de-repressed by binding to Inositol Pyrophospates. (bottom) Diagrams showing synthesis of Inositol Pyrophospates (InsP7/InspP8).

[0021] **FIG. 2 – Crystal structure of SPX domain from the alpha fold model for XPR1.** The SPX domain (bottom) has been crystalized. The transmembrane (TM) domains are modeled.

[0022] **FIG. 3 – DepMap analysis of InsPP synthesizing enzymes.** Graph showing expression of the indicated genes in DepMap cell lines.

[0023] **FIG. 4 – DepMap analysis of InsPP synthesizing enzymes.** Graph showing chronos score for the indicated genes in DepMap cell lines. See, e.g., Dempster JM, Boyle I, Vazquez F, et al. Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. Genome Biol. 2021;22(1):343.

[0024] **FIG. 5 – DepMap analysis of phosphate transporters and InsPP synthesizing enzymes.** Heatmap showing correlation between the indicated genes in DepMap (i.e., correlation between dependencies).

[0025] **FIG. 6A-FIG. 6B – IP6K inhibitors. FIG. 6A.** Graph showing phosphate efflux in an ovarian cancer cell line treated with TNP or RBD_00. **FIG. 6B.** Graph showing cell

viability in an ovarian cancer cell line treated with TNP or RBD_00 and the indicated CRISPR guide sequences.

[0026] **FIG. 7A-FIG. 7C – SC-919 inhibits XPR1-dependent cell lines. FIG. 7A.** Graph showing phosphate efflux inhibition ovarian cancer cell lines. **FIG. 7B.** Cell viability assay in ovarian cancer cell lines treated with SC-919. **FIG. 7C.** Cell viability assay in wild type and SLC34A2 inactivated RMGI ovarian clear cell carcinoma cell line. Cells were treated with vehicle, SC-919, or XRBD.

[0027] **FIG. 8 – SC-919 inhibits XPR1-dependent cell lines.** Cell viability assay on a panel of ovarian cancer cell lines treated with SC-919. SLC34A2-HIGH cell lines are OVISE, IGROV1, OVCAR3, and RMGI. SLC34A2-LOW cell lines are 59M, ES2, and RMGI after SLC34A2 inactivation.

[0028] **FIG. 9 – Confirmation of cell death after treatment with SC-919.** The OVISE cell line was treated with DMSO (vehicle), bortezomib (positive control for apoptosis), Dox (induces shRNA that inhibits XPR1), and SC-919. Cell death was measured by flow cytometric analyses of Annexin V staining and DAPI uptake.

[0029] **FIG. 10 – Evaluation of SC-919 in PRISM.** Barcoded cancer cell lines were pooled together, treated with increasing concentrations of SC-919, and barcode abundance was quantified using next generation sequencing.

[0030] **FIG. 11 – Evaluation of SC-919 in PRISM.** The 400 cancer cell lines observed in FIG. 10 were treated with 10 μM SC-919. (top) graph showing the highest correlations of SC-919 treatment and cell lines sensitive to CRISPR knockout of genes in the cell lines. (bottom) graph showing the highest correlations of SC-919 treatment and the expression profile of different genes across the 400 cancer cell lines.

[0031] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0032] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and

Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[0033] As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[0034] The term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0035] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0036] The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[0037] As used herein, a "biological sample" may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a "bodily fluid". The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal

fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[0038] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

[0039] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to "one embodiment", "an embodiment," "an example embodiment," means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0040] Reference is made to International Patent Application PCT/US2021/045227, filed August 9, 2021. Reference is also made to Bondeson DP, Paolella BR, Asfaw A, et al. Phosphate dysregulation via the XPR1-KIDINS220 protein complex is a therapeutic vulnerability in ovarian cancer [published online ahead of print, 2022 Apr 18]. Nat Cancer.

2022;10.1038/s43018-022-00360-7. doi:10.1038/s43018-022-00360-7. All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

OVERVIEW

[0041] Exploiting cancer-specific metabolic states is an attractive strategy to kill cancer cells while sparing normal tissues (Faubert, B., Solmonson, A. & DeBerardinis, R. J. Metabolic reprogramming and cancer progression. *Science* 368, (2020)). Embodiments disclosed herein provide for treating cancers sensitive to phosphate dysregulation with specific inhibitors of inositol pyrophosphate (PP-InsP) synthesis, in particular, specific inhibitors of inositol hexakisphosphate kinases (IP6Ks). Applicants show for the first time that cancers sensitive to phosphate dysregulation are vulnerable to inhibitors of inositol hexakisphosphate kinases (IP6Ks). Thus, the present invention provides for treating cancers using inhibitors of inositol pyrophosphate (PP-InsP) synthesis, including inhibitors of inositol hexakisphosphate kinases (IP6Ks) and diphosphoinositol pentakisphosphate kinases (PPIP5Ks). Sensitive tumors have increased expression of the phosphate importer SLC34A2 and do not sufficiently suppress phosphate import in response to XPR1/phosphate efflux inhibition to prevent cell growth inhibition. Tumors can be made sensitive to inhibitors of inositol pyrophosphate (PP-InsP) synthesis by inhibiting suppression of SLC34A2 phosphate import that can occur in non-sensitive cancers in response to inhibiting phosphate export.

[0042] Embodiments disclosed herein also provide for identifying cancers sensitive to phosphate dysregulation using inhibitors of inositol pyrophosphate (PP-InsP) synthesis and detecting the inositol pyrophosphate level in the cell or cell population or detecting the phosphate concentration in the cell or cell population. Subject specific cancers can be analyzed to determine personalized treatments.

[0043] Embodiments disclosed herein are also directed to a method of treating cancer based on determining if the subject has a tumor sensitive to phosphate dysregulation and administering one or more inhibitors of inositol pyrophosphate (PP-InsP) synthesis if the subject has a tumor sensitive to phosphate dysregulation. In another aspect, embodiments disclosed herein provide methods of determining whether a subject has a tumor sensitive to

phosphate dysregulation by detecting expression of SLC34A2 in a tumor sample or detecting amplification in XPR1 copy number in a tumor sample.

Exemplary genes

[0044] In example embodiments, all gene name symbols refer to the gene as commonly known in the art. The examples described herein that refer to the human gene names are to be understood to also encompasses genes in any other organism, for example mouse genes or any other gene used in a model of disease (e.g., homologous, orthologous genes). Any reference to the gene symbol is a reference made to the entire gene or variants of the gene. Any reference to the gene symbol is also a reference made to the gene product (e.g., protein). Gene symbols may be those referred to by the HUGO Gene Nomenclature Committee (HGNC) or National Center for Biotechnology Information (NCBI). A signature as described herein may encompass any of the genes described herein.

[0045] As used herein, IP6K1 refers to inositol hexakisphosphate kinase 1 (Also known as: IHPK1, PiUS). Exemplary sequences include the following NCBI accession numbers: NM_153273.4, NM_001006115.3, NM_001242829.2, NP_695005.1, NP_001006115.1, and NP_001229758.1. This gene encodes a member of the inositol phosphokinase family. The encoded protein converts inositol hexakisphosphate (InsP6) to diphosphoinositol pentakisphosphate (InsP7/PP-InsP5). It also converts 1,3,4,5,6-pentakisphosphate (InsP5) to PP-InsP4. Alternatively spliced transcript variants have been described. Diseases associated with IP6K1 include Type 2 Diabetes Mellitus.

[0046] As used herein, IP6K2 refers to inositol hexakisphosphate kinase 2 (Also known as: PIUS, IHPK2, InsP6K2). Exemplary sequences include the following NCBI accession numbers: NM_001005909.3, NM_001005910.3, NM_001005911.3, NM_001146178.3, NM_001146179.3, NM_001190316.2, NM_001190317.2, NM_016291.4, NP_001005909.1, NP_001005910.1, NP_001005911.1, NP_001139650.1, NP_001139651.1, NP_001177245.1, NP_001177246.1, and NP_057375.2. This gene encodes a member of the inositol phosphokinase family. This protein is likely responsible for the conversion of inositol hexakisphosphate (InsP6) to diphosphoinositol pentakisphosphate (InsP7/PP-InsP5). It may also convert 1,3,4,5,6-pentakisphosphate (InsP5) to PP-InsP4 and affect the growth suppressive and apoptotic activities of interferon-beta in some ovarian cancers. Alternative splicing results in multiple transcript variants encoding different isoforms.

[0047] As used herein, IP6K3 refers to inositol hexakisphosphate kinase 3 (Also known as: IHPK3, INSP6K3). Exemplary sequences include the following NCBI accession numbers: NM_054111.5, NM_001142883.2, NP_473452.2, and NP_001136355.1. This gene encodes a protein that belongs to the inositol phosphokinase (IPK) family. This protein is likely responsible for the conversion of inositol hexakisphosphate (InsP6) to diphosphoinositol pentakisphosphate (InsP7/PP-InsP5). It may also convert 1,3,4,5,6-pentakisphosphate (InsP5) to PP-InsP4. Alternative splicing results in multiple transcript variants encoding the same protein.

[0048] As used herein, PPIP5K1 refers to diphosphoinositol pentakisphosphate kinase 1 (Also known as: HISPPD2A, IP6K, IPS1, VIP1, hsVIP1). Exemplary sequences include the following NCBI accession numbers: NM_001130858.4, NP_001124330.1, NM_001130859.3, NP_001124331.1, NM_001190214.2, NP_001177143.1, NM_001354382.3, NP_001341311.1, NM_001354383.2, NP_001341312.1, NM_001354384.2, NP_001341313.1, NM_001354385.2, NP_001341314.1, NM_001354386.2, NP_001341315.1, NM_001354387.2, NP_001341316.1, NM_001354388.2, NP_001341317.1, NM_001354389.2, NP_001341318.1, NM_001354390.2, NP_001341319.1, NM_001354391.2, NP_001341320.1, NM_001354392.2, NP_001341321.1, NM_001354393.2, NP_001341322.1, NM_001354394.2, NP_001341323.1, NM_001354395.2, NP_001341324.1, NM_001354396.2, NP_001341325.1, NM_001354397.2, NP_001341326.1, NM_001354398.2, NP_001341327.1, NM_001354399.2, NP_001341328.1, NM_001354400.2, NP_001341329.1, NM_001354401.2, NP_001341330.1, NM_001354402.2, NP_001341331.1, NM_001393969.1, NP_001380898.1, NM_001393970.1, NP_001380899.1, NM_001393971.1, NP_001380900.1, NM_001394395.1, NP_001381324.1, NM_014659.6, and NP_055474.3. This gene encodes a protein that belongs to the inositol phosphokinase (IPK) family. Bifunctional inositol kinase that acts in concert with the IP6K kinases IP6K1, IP6K2 and IP6K3 to synthesize the diphosphate group-containing inositol pyrophosphates diphosphoinositol pentakisphosphate, PP-InsP5, and bis-diphosphoinositol tetrakisphosphate, (PP)2-InsP4. PP-InsP5 and (PP)2-InsP4, also respectively called InsP7 and InsP8, regulate a variety of cellular processes, including apoptosis, vesicle trafficking, cytoskeletal dynamics, exocytosis, insulin signaling and neutrophil activation. Phosphorylates inositol hexakisphosphate (InsP6) at positions 1 or 5 to

produce PP-InsP5 which is in turn phosphorylated by IP6Ks to produce (PP)2-InsP4. Alternatively, phosphorylates at position 1 or 5 PP-InsP5, produced by IP6Ks from InsP6, to produce (PP)2-InsP4. Activated when cells are exposed to hyperosmotic stress. Diseases associated with PPIP5K1 include Glaucomatocyclitic Crisis and Gastric Leiomyoma.

[0049] As used herein, PPIP5K2 refers to diphosphoinositol pentakisphosphate kinase 2 (Also known as: CFAP160, DFNB100, HISPPD1, IP7K2, VIP2). Exemplary sequences include the following NCBI accession numbers: NM_001276277.3, NP_001263206.1, NM_001281471.3, NP_001268400.1, NM_001345871.2, NP_001332800.1, NM_001345872.2, NP_001332801.1, NM_001345873.2, NP_001332802.1, NM_001345874.2, NP_001332803.1, NM_001345875.2, NP_001332804.1, NM_001345876.2, NP_001332805.1, NM_001345877.2, NP_001332806.1, NM_001345878.2, NP_001332807.1, NM_015216.5, and NP_056031.2. Despite containing a histidine acid phosphatase domain, the encoded protein functions as an inositol pyrophosphate kinase, and is thought to lack phosphatase activity. This kinase activity is the mechanism by which the encoded protein synthesizes high-energy inositol pyrophosphates, which act as signaling molecules that regulate cellular homeostasis and other processes. This gene may be associated with autism spectrum disorder in human patients. Bifunctional inositol kinase that acts in concert with the IP6K kinases IP6K1, IP6K2 and IP6K3 to synthesize the diphosphate group-containing inositol pyrophosphates diphosphoinositol pentakisphosphate, PP-InsP5, and bis-diphosphoinositol tetrakisphosphate, (PP)2-InsP4. PP-InsP5 and (PP)2-InsP4, also respectively called InsP7 and InsP8, regulate a variety of cellular processes, including apoptosis, vesicle trafficking, cytoskeletal dynamics, exocytosis, insulin signaling and neutrophil activation. Phosphorylates inositol hexakisphosphate (InsP6) at positions 1 or 3 to produce PP-InsP5 which is in turn phosphorylated by IP6Ks to produce (PP)2-InsP4. Alternatively, phosphorylates at position 1 or 3 PP-InsP5, produced by IP6Ks from InsP6, to produce (PP)2-InsP4. Required for normal hearing. Diseases associated with PPIP5K2 include Deafness, Autosomal Recessive 100 and Autosomal Recessive Non-Syndromic Sensorineural Deafness Type Dfnb.

[0050] As used herein, XPR1 refers to xenotropic and polytropic retrovirus receptor 1 (Also known as: IBGC6, SLC53A1, SYG1, X3). Exemplary sequences include the following NCBI accession numbers: NM_004736.4, NM_001135669.2, NM_001328662.2, NP_004727.2, NP_001129141.1 and NP_001315591.1. XPR1 includes a SPX domain (N-

terminal) and EXS domain (C-terminal) that can be targeted, in addition to targeting the entire protein, by a therapeutic agent (e.g., small molecules, antibodies) or agent for detecting expression (e.g., antibodies). XPR1 is a phosphate exporter in metazoans, a function that does not require the SPX domain (see, e.g., Giovannini, et al., Inorganic phosphate export by the retrovirus receptor XPR1 in metazoans. Cell Rep. 2013;3(6):1866-1873; Legati, et al. Mutations in XPR1 cause primary familial brain calcification associated with altered phosphate export. Nat Genet. 2015;47(6):579-581; Ansermet, et al., Renal Fanconi Syndrome and Hypophosphatemic Rickets in the Absence of Xenotropic and Polytropic Retroviral Receptor in the Nephron. J Am Soc Nephrol. 2017;28(4):1073-1078).

[0051] As used herein, KIDINS220 refers to kinase D interacting substrate 220 (Also known as: ARMS, SINO). Exemplary sequences include the following NCBI accession numbers: NM_020738.4, NM_001348729.2, NM_001348731.2, NM_001348732.2, NM_001348734.2, NM_001348735.2, NM_001348736.2, NM_001348738.2, NM_001348739.2, NM_001348740.2, NM_001348741.2, NM_001348742.2, NM_001348743.2, NM_001348745.2, NP_065789.1, NP_001335658.1, NP_001335660.1, NP_001335661.1, NP_001335663.1, NP_001335664.1, NP_001335665.1, NP_001335667.1, NP_001335668.1, NP_001335669.1, NP_001335670.1, NP_001335671.1, NP_001335672.1, and NP_001335674.1. KIDINS220 includes an Ankyrin repeat-containing domain (N-terminal) and KAP family P-loop domain that can be targeted by a therapeutic agent (e.g., small molecules, antibodies) or agent for detecting expression (e.g., antibodies). The P-loop domain is characterized by two conserved motifs, termed the Walker A and B motifs. The Walker A motif, also known as the Walker loop, or P-loop, or phosphate-binding loop, is a motif in proteins that is associated with phosphate binding. The Walker B motif is a motif in most P-loop proteins situated well downstream of the A-motif.

[0052] As used herein, SLC34A2 refers to solute carrier family 34 member 2 (Also known as: NAPI-3B, NAPI-IIb, NPTIIb, PULAM). Exemplary sequences include the following NCBI accession numbers: NM_006424.3, NM_001177998.2, NM_001177999.1, NP_006415.3, NP_001171469.2, and NP_001171470.1. Synthetic peptides derived from a complementarity determining region hypervariable domain amino acid sequence of a humanized monoclonal antibody to NaPi2B transporter has been described for inhibiting tumor growth or treating cancer (US 9,193,797 B2).

[0053] As used herein, SLC20A1 refers to solute carrier family 20 member 1 (Also known as: GLVR1, Glvr-1, PIT1, PiT-1). Exemplary sequences include the following NCBI accession numbers: NM_005415.5 and NP_005406.3.

[0054] As used herein, PAX8 refers to paired box 8. Exemplary sequences include the following NCBI accession numbers: NM_003466.4, NM_013952.4, NM_013953.4, NM_013992.4, NP_003457.1, NP_039246.1, NP_039247.1, and NP_054698.1.

[0055] As used herein, FGF23 refers to fibroblast growth factor 23 (Also known as: ADHR, FGFN, HFTC2, HPDR2, HYPF, PHPTC). Exemplary sequences include the following NCBI accession numbers: NM_020638.3 and NP_065689.1.

[0056] As used herein "RBD", "RBD_00", and "XRBD" are used interchangeably and refer to a receptor binding protein, which is an inhibitor of XPR1 (see, e.g., International Patent Application PCT/US2021/045227). In example embodiments, the RBD protein is derived from an enveloped virus glycoprotein and capable of interacting with the XPR1 membrane receptor. In example embodiments, RBD is an about 238 residue fragment of the envelope glycoprotein for a retrovirus, such as X-MLV, and the fragment inhibits XPR1 phosphate efflux (see, e.g., Giovannini, et al., Inorganic phosphate export by the retrovirus receptor XPR1 in metazoans. Cell Rep. 2013;3(6):1866-1873).

THERAPEUTIC METHODS TARGETING CANCER DEPENDENCIES

[0057] In one aspect, the present invention provides for treating cancer in a subject in need thereof comprising administering to the subject one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis. The InsP family of molecules consists of monophosphorylated inositol (InsP) to the fully phosphorylated $InsP_6$ (inositol hexakisphosphate), and are generated by InsP kinases that reversibly add phosphates to specific positions of the 6-carbon inositol ring backbone. InsPs can be further phosphorylated to produce the high-energy inositol pyrophosphates (PP-InsPs), which while only found in low concentrations in the cell, have been implicated in an array of developmental, metabolic and signaling processes. As used herein, "inositol pyrophosphate (PP-InsP) synthesis" refers to any further phosphorylation of a phosphate on any inositol phosphate. In preferred embodiments, phosphorylation of $InsP_5$, $InsP_6$, or $InsP_7$ is inhibited (see, e.g., Fig. 1). As used herein, "treatment" or "treating," or "palliating" or "ameliorating" are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any

therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested. As used herein "treating" includes ameliorating, curing, preventing it from becoming worse, slowing the rate of progression, or preventing the disorder from re-occurring (i.e., to prevent a relapse).

[0058] In example embodiments, a therapeutically effective amount of one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis is administered. The term “effective amount” or “therapeutically effective amount” refers to the amount of an agent that is sufficient to effect beneficial or desired results. The therapeutically effective amount may vary depending upon one or more of: the subject and disease condition being treated, the weight and age of the subject, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The term also applies to a dose that will provide an image for detection by any one of the imaging methods described herein. The specific dose may vary depending on one or more of: the particular agent chosen, the dosing regimen to be followed, whether it is administered in combination with other compounds, timing of administration, the tissue to be imaged, and the physical delivery system in which it is carried.

[0059] For example, in methods for treating cancer in a subject, an effective amount of a combination of inhibitors is any amount that provides an anti-cancer effect, such as reduces or prevents proliferation of a cancer cell or is cytotoxic towards a cancer cell. In example embodiments, the effective amount of an inhibitor is reduced when an inhibitor is administered concomitantly or in combination with one or more additional inhibitors as compared to the effective amount of the inhibitor when administered in the absence of one or more additional inhibitors.

Exemplary cancers

[0060] The present invention may be useful for the treatment of any cancer sensitive to phosphate dysregulation, such as cancers dependent on XPR1:KIDINS220-mediated phosphate export. In other example embodiments, the cancer has increased expression of SLC34A2 as compared to normal tissue (e.g., ovarian cancer, uterine cancer, breast cancer,

bile duct cancer, liver and lung cancer). In preferred embodiments, the cancer is ovarian or uterine cancer. Detection of SLC34A2 expression is further described herein.

[0061] Exemplary cancers that may benefit from treatment with one or more inhibitors of inositol pyrophosphate (PP-InsP) synthesis include ovarian cancer, uterine cancer, breast cancer, bile duct cancer, liver and lung cancer. Exemplary cancers that may benefit from treatment with one or more inhibitors of inositol pyrophosphate (PP-InsP) synthesis also include liquid tumors such as leukemia (e.g., acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, acute myeloblastic leukemia, acute promyelocytic leukemia, acute myelomonocytic leukemia, acute monocytic leukemia, acute erythroleukemia, chronic leukemia, chronic myelocytic leukemia, chronic lymphocytic leukemia), polycythemia vera, lymphoma (e.g., Hodgkin's disease, non-Hodgkin's disease), Waldenstrom's macroglobulinemia, heavy chain disease, or multiple myeloma. The cancer may include solid tumors such as sarcomas and carcinomas. Examples of solid tumors include, but are not limited to, fibrosarcoma, myxosarcoma, liposarcoma, chondrosarcoma, osteogenic sarcoma, chordoma, angiosarcoma, endotheliosarcoma, lymphangiosarcoma, lymphangioendotheliosarcoma, synovioma, mesothelioma, Ewing's tumor, leiomyosarcoma, rhabdomyosarcoma, squamous cell carcinoma, basal cell carcinoma, adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma, papillary carcinoma, papillary adenocarcinomas, cystadenocarcinoma, medullary carcinoma, epithelial carcinoma, bronchogenic carcinoma, hepatoma, colorectal cancer (e.g., colon cancer, rectal cancer), anal cancer, pancreatic cancer (e.g., pancreatic adenocarcinoma, islet cell carcinoma, neuroendocrine tumors), breast cancer (e.g., ductal carcinoma, lobular carcinoma, inflammatory breast cancer, clear cell carcinoma, mucinous carcinoma), ovarian carcinoma (e.g., ovarian epithelial carcinoma or surface epithelial-stromal tumour including serous tumour, endometrioid tumor and mucinous cystadenocarcinoma, sex-cord-stromal tumor), prostate cancer, liver and bile duct carcinoma (e.g., hepatocelluar carcinoma, cholangiocarcinoma, hemangioma), choriocarcinoma, seminoma, embryonal carcinoma, kidney cancer (e.g., renal cell carcinoma, clear cell carcinoma, Wilm's tumor, nephroblastoma), cervical cancer, uterine cancer (e.g., endometrial adenocarcinoma, uterine papillary serous carcinoma, uterine clear-cell carcinoma, uterine sarcomas and leiomyosarcomas, mixed mullerian tumors), testicular cancer, germ cell tumor, lung cancer (e.g., lung adenocarcinoma, squamous cell carcinoma, large cell carcinoma, bronchioloalveolar carcinoma, non-small-cell carcinoma, small cell carcinoma,

mesothelioma), bladder carcinoma, signet ring cell carcinoma, cancer of the head and neck (e.g., squamous cell carcinomas), esophageal carcinoma (e.g., esophageal adenocarcinoma), tumors of the brain (e.g., glioma, glioblastoma, astrocytoma, medulloblastoma, craniopharyngioma, ependymoma, pinealoma, hemangioblastoma, acoustic neuroma, oligodenroglioma, schwannoma, meningioma), neuroblastoma, retinoblastoma, neuroendocrine tumor, melanoma, cancer of the stomach (e.g., stomach adenocarcinoma, gastrointestinal stromal tumor), or carcinoids. Lymphoproliferative disorders are also considered to be proliferative diseases. In certain example embodiments, the cancer is ovarian cancer. In certain other example embodiments, the cancer is uterine cancer.

Therapeutic Agents

[0062] In example embodiments, the present invention provides for one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis. In example embodiments, the therapeutic agents are capable of inhibiting one or more inositol hexakisphosphate kinases (IP6Ks). In example embodiments, the therapeutic agents are capable of inhibiting one or more diphosphoinositol pentakisphosphate kinases (PPIP5Ks). The terms "therapeutic agent", "therapeutic capable agent" or "treatment agent" are used interchangeably and refer to a molecule or compound, or combination of molecules or compounds, that confers some beneficial effect upon administration to a subject. The beneficial effect includes enablement of diagnostic determinations; amelioration of a disease, symptom, disorder, or pathological condition; reducing or preventing the onset of a disease, symptom, disorder or condition; and generally counteracting a disease, symptom, disorder or pathological condition.

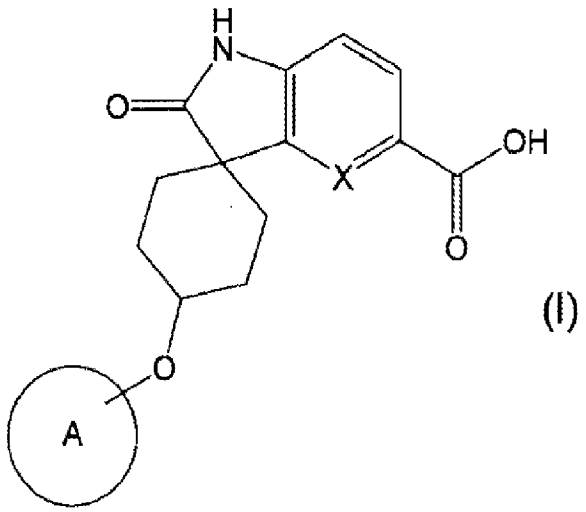
Small Molecules

[0063] In example embodiments, the one or more therapeutic agents comprise a small molecule that inhibits expression or activity of IP6Ks and/or PPIP5Ks. The term "small molecule" refers to compounds, preferably organic compounds, with a size comparable to those organic molecules generally used in pharmaceuticals. The term excludes biological macromolecules (e.g., proteins, peptides, nucleic acids, etc.). Preferred small organic molecules range in size up to about 5000 Da, e.g., up to about 4000, preferably up to 3000 Da, more preferably up to 2000 Da, even more preferably up to about 1000 Da, e.g., up to about 900, 800, 700, 600 or up to about 500 Da. In example embodiments, the small molecule may block an enzyme active site.

Inositol Hexakisphosphate Kinase Inhibitors

[0001] In example embodiments, the one or more therapeutic agents are inhibitors of one or more inositol hexakisphosphate kinases (IP6Ks). Mammals have three IP6K subtypes of IP6K1, IP6K2, and IP6K3. The one or more IP6Ks can be IP6K1, IP6K2, or IP6K3. The one or more agents may be capable of inhibiting enzymatic activity of IP6K1, IP6K2, and/or IP6K3 with an *in vitro* IC_{50} of less than 10 nM, preferably about 5 nM (see, e.g., Moritoh Y, Abe SI, Akiyama H, et al. The enzymatic activity of inositol hexakisphosphate kinase controls circulating phosphate in mammals. Nat Commun. 2021;12(1):4847). The one or more agents targeting IP6K1, IP6K2, and/or IP6K3 may be capable of inhibiting phosphate export in a cellular assay with an IC_{50} of less than 100 nM, preferably about 50 nM, more preferably in ovarian cancer cells (see, e.g., **Fig. 7A**).

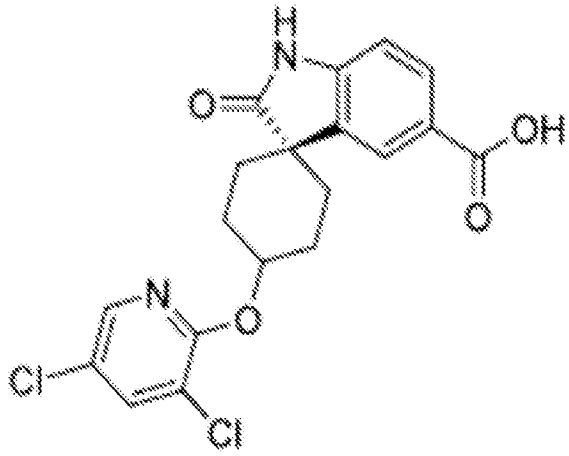
[0002] Non-limiting examples of IP6K inhibitors applicable to the present invention are provided in International Patent Application publication number WO2018182051A1. In example embodiments, the inhibitor is represented by the formula:



(I)

[0003]

[0004] or a salt thereof, wherein ring A is an optionally substituted aromatic ring; X is CH or N. In example embodiments, the inhibitor is represented by the formula:



[0005] or a salt thereof. The IUPAC name is (ls,4s)-4-((3,5-dichloropyridin-2-yl)oxy)-2'-oxo-1',2'-dihydrospiro[cyclohexane-1,3'- indole]-5'-carboxylic acid.

[0006] In example embodiments, the IP6K inhibitor N2-(m-trifluorobenzyl), N6-(p-nitrobenzyl) purine (TNP) is not applicable to the present invention because it is not specific to cancers sensitive to phosphate dysregulation (see, e.g., Li, et al., Proc Natl Acad Sci USA. 2020;117(7):3568-3574).

Agents capable of inhibiting the suppression of SLC34A2

[0070] Phosphate-sensing mechanisms are present in cells which counteract increased intracellular phosphate caused by blocking phosphate efflux (e.g., by inhibiting XPR1), possibly due to feedback mechanisms which suppress phosphate import upon increased intracellular phosphate. However, in cancers sensitive to phosphate dysregulation the mechanisms cannot overcome the increased phosphate caused by blocking phosphate efflux (see, e.g., Bondeson, et al., 2022 and PCT/US2021/045227). In example embodiments, the method further comprises administering to the subject one or more therapeutic agents capable of inhibiting the suppression of SLC34A2 phosphate import. Applicants previously observed a strong and highly correlated transcriptional response among the most XPR1-dependent cell lines. First, Applicants observed the up-regulation of FGF23, a critical phosphate homeostatic hormone typically expressed by osteogenic bone cells in response to elevated serum phosphate (Minisola, S. *et al.* Tumour-induced osteomalacia. *Nat Rev Dis Primers* **3**, 17044 (2017)). Tumors associated with tumor-induced osteomalacia (TIO) release a peptide hormone-like substance known as fibroblast growth factor 23 (FGF23) that lowers phosphate levels. In example embodiments, one or more therapeutic agents capable of inhibiting the suppression of SLC34A2 is an FGF23 antagonist capable of blocking FGF23 from reducing phosphate uptake. In example embodiments, the agent is an anti-FGF23 antibody, such as, but not limited to Burosumab. Burosumab, sold under the brand name Crysvita, is a human monoclonal antibody medication for the treatment of X-linked hypophosphatemia and tumor-induced osteomalacia. Applicants also observed the downregulation of phosphate import. Two phosphate importer genes, SLC20A1 and SLC34A2, were significantly decreased after XPR1 knockout. In example embodiments, the agent increases expression or activity of SLC20A1 and/or SLC34A2 in tumor cells. In example embodiments, SLC20A1 and/or SLC34A2 are overexpressed in tumor cells (e.g., using a tumor specific vector or directly administering a vector to the tumor).

Standard of Care

[0071] Aspects of the invention involve modifying the therapy within a standard of care based on the detection of any of the biomarkers as described herein. In one embodiment, therapy comprising an agent is administered within a standard of care where addition of the agent is synergistic within the steps of the standard of care. In one embodiment, the agent targets and/or shifts a tumor to be more vulnerable to a therapeutic agent targeting

XPR1:KIDINS220-mediated phosphate export (e.g., an inhibitor of PP-InsP synthesis). In one embodiment, the agent inhibits expression or activity of one or more genes involved in phosphate homeostasis. The term "standard of care" as used herein refers to the current treatment that is accepted by medical experts as a proper treatment for a certain type of disease and that is widely used by healthcare professionals. Standard of care is also called best practice, standard medical care, and standard therapy. Standards of care for cancer generally include surgery, lymph node removal, radiation, chemotherapy, targeted therapies, antibodies targeting the tumor, and immunotherapy. Immunotherapy can include checkpoint blockers (CBP), chimeric antigen receptors (CARs), and adoptive T-cell therapy. The standards of care for the most common cancers can be found on the website of National Cancer Institute (www.cancer.gov/cancertopics). A treatment clinical trial is a research study meant to help improve current treatments or obtain information on new treatments for patients with cancer. When clinical trials show that a new treatment is better than the standard treatment, the new treatment may be considered the new standard treatment.

[0072] In example embodiments, the present invention provides for one or more therapeutic agents (e.g., inhibitors of PP-InsP synthesis) that can be used in combination with the standard of care for the cancer. Targeting XPR1:KIDINS220-mediated phosphate export in combination within a standard of care may provide for enhanced or otherwise previously unknown activity in the treatment of disease.

[0073] In example embodiments, the present invention provides for a combination therapy comprising a treatment described herein with a treatment that is part of the standard of care for a cancer (i.e., a therapeutic regime). In example embodiments, the standard of care for treating ovarian cancer comprises surgery, chemotherapy, and targeted therapy (see, e.g., Lheureux *et al.*, Epithelial ovarian cancer: Evolution of management in the era of precision medicine. CA Cancer J Clin. 2019 Jul;69(4):280-304). Ovarian cancer is a cancer that forms in or on an ovary. Symptoms may include bloating, pelvic pain, abdominal swelling, and loss of appetite, among others. Common areas to which the cancer may spread include the lining of the abdomen, lymph nodes, lungs, and liver. The most common type of ovarian cancer is ovarian carcinoma (>95% of all cases). There are five main subtypes of ovarian carcinoma, of which high-grade serous carcinoma is the most common. These tumors are believed to start in the cells covering the ovaries, though some may form at the Fallopian tubes. Less common types of ovarian

cancer include germ cell tumors and sex cord stromal tumors. A diagnosis of ovarian cancer is confirmed through a biopsy of tissue, usually removed during surgery.

[0074] If caught and treated in an early stage, ovarian cancer is often curable. Treatment usually includes some combination of surgery, radiation therapy, and chemotherapy. Outcomes depend on the extent of the disease, the subtype of cancer present, and other medical conditions. The overall five-year survival rate in the United States is 45%.

[0075] If ovarian cancer recurs, it is considered partially platinum-sensitive or platinum-resistant, based on the time since the last recurrence treated with platins: partially platinum-sensitive cancers recurred 6-12 months after last treatment, and platinum-resistant cancers have an interval of less than 6 months.

[0076] For platinum-sensitive tumors, platins are utilized for second-line chemotherapy, often in combination with other cytotoxic agents. Regimens include carboplatin combined with pegylated liposomal doxorubicin, gemcitabine, or paclitaxel. If the tumor is determined to be platinum-resistant, vincristine, dactinomycin, and cyclophosphamide (VAC) or some combination of paclitaxel, gemcitabine, and oxaliplatin can be used as a second-line therapy.

[0077] Systemic therapy can include single to combination chemotherapy approaches alone or in combination with targeted therapy. In example embodiments, surgery includes surgery for accurate surgical staging, primary debulking surgery, interval debulking surgery, and secondary debulking surgery. In example embodiments, chemotherapy includes carboplatin, cisplatin and paclitaxel. In example embodiments, targeted therapy includes Bevacizumab, which is a humanized monoclonal antibody against vascular endothelial growth factor (VEGF), and poly (ADP-ribose) polymerase (PARP) inhibitors (e.g., Olaparib, Niraparib, and Rucaparib). Other therapies that may be used in combination with the present invention include agents targeting the folate receptor (e.g., mirvetuximab soravtansine (IMGN853), which is an ADC consisting of an anti-FRα antibody linked to the tubulin-disrupting maytansinoid DM4 drug, a potent antimitotic agent). In example embodiments, checkpoint blockade therapy is used in a combination therapy. As used herein, checkpoint blockade therapy (CPB) refers to antibodies that block the activity of checkpoint receptors, including CTLA-4, PD-1, Tim-3, Lag-3, and TIGIT, either alone or in combination. The checkpoint blockade therapy may comprise anti-TIM3, anti-CTLA4, anti-PD-L1, anti-PD1, anti-TIGIT, anti-LAG3, or combinations thereof. Anti-PD1 antibodies are disclosed in U.S. Pat. No. 8,735,553. Antibodies to LAG-3 are disclosed in U.S. Pat. No. 9,132,281. Anti-

CTLA4 antibodies are disclosed in U.S. Pat. No. 9,327,014; U.S. Pat. No. 9,320,811; and U.S. Pat. No. 9,062,111. Specific check point inhibitors include, but are not limited to, anti-CTLA4 antibodies (e.g., Ipilimumab and Tremelimumab), anti-PD-1 antibodies (e.g., Nivolumab, Pembrolizumab, Dostarlimab), and anti-PD-L1 antibodies (e.g., Atezolizumab). In example embodiments, chemotherapy in combination with immunotherapy is used in the treatment of ovarian cancer. In example embodiments, the combination therapy comprises paclitaxel plus pembrolizumab, preferably in patients with platinum-resistant ovarian cancer. In example embodiments, the combination therapy comprises immunotherapy combined with PARP inhibitors.

[0078] In example embodiments, inhibitors of PP-InsP synthesis may be administered in combination with the current standard of care and may provide for improved treatment and/or less toxicity.

Detection Methods

[0079] In example embodiments, the therapeutic methods include determining whether a subject is a candidate for a treatment inhibiting PP-InsP synthesis. For example, the subject has a cancer that is sensitive to phosphate dysregulation. In example embodiments, to determine the efficacy of treatment, phosphate efflux, phosphate regulation (e.g., inositol pyrophosphate levels), or morphological changes are detected.

Detection of SLC34A2

[0080] In example embodiments, detection of increased expression of SLC34A2 and/or covarying genes indicates that a tumor is sensitive to inhibition of XPR1:KIDINS220-mediated phosphate export. SLC34A2 expression can be determined by detection of the protein or RNA transcripts using any method described further herein. Antibodies capable of detecting SLC34A2 have been developed (see, e.g., MX35: Yin BW, et al., Monoclonal antibody MX35 detects the membrane transporter NaPi2b (SLC34A2) in human carcinomas. Cancer Immun. 2008;8:3; Levan K, et al., Immunohistochemical evaluation of epithelial ovarian carcinomas identifies three different expression patterns of the MX35 antigen, NaPi2b. BMC Cancer. 2017;17(1):303; and RebMab200: Lopes dos Santos, et al., Rebmab200, a Humanized Monoclonal Antibody Targeting the Sodium Phosphate Transporter NaPi2b Displays Strong Immune Mediated Cytotoxicity against Cancer: A Novel Reagent for Targeted Antibody Therapy of Cancer. PLoS One. 2013; 8(7): e70332) and are applicable to the present invention. Detection of SLC34A2 with an anti-NaPi2b antibody has been described for determining

whether a cancer is responsive to a NaPi2b-targeted antibody drug conjugate (see, US 2019/0160181 A1). Any future antibodies developed are also applicable to the present invention.

[0081] In example embodiments, detecting comprises one or more of immunohistochemistry (IHC), in situ RNA-seq (Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods 10, 857–860 (2013)), quantitative PCR, RNA-seq, single cell or single nuclei RNA-seq (see, e.g. Picelli, S. et al., 2014, "Full-length RNA-seq from single cells using Smart-seq2" Nature protocols 9, 171-181, doi:10.1038/nprot.2014.006; Macosko et al., 2015, "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets" Cell 161, 1202–1214; Klein et al., 2015, "Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells" Cell 161, 1187–1201; Zheng, et al., 2017, "Massively parallel digital transcriptional profiling of single cells" Nat. Commun. 8, 14049 doi: 10.1038/ncomms14049; Hughes, et al., "Highly Efficient, Massively-Parallel Single-Cell RNA-Seq Reveals Cellular States and Molecular Features of Human Skin Pathology" bioRxiv 689273; doi: doi.org/10.1101/689273; and Drokhlyansky E, Smillie CS, Van Wittenberghe N, et al. The Human and Mouse Enteric Nervous System at Single-Cell Resolution. Cell. 2020;182(6):1606-1622.e23), CITE-seq (Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods 14, 865–868 (2017)), western blot, ELISA, Fluorescence In Situ Hybridization (FISH), MERFISH (Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, (2015)), RNA-FISH, mass spectrometry, or FACS.

Copy Number Variation

[0082] In example embodiments, copy number variations (CNV) are detected in a tumor (e.g., XPR1) (see, e.g., Carter SL, et al., Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012 May; 30(5):413-21; Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352, 189–196 (2016); Sathirapongsasuti, J. F. et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics 27, 2648–2654 (2011); Krumm, N. et al. Copy number variation detection and genotyping from exome sequence data. Genome Res. 22, 1525–1532 (2012); de Araújo Lima, L. & Wang, K. PennCNV in whole-genome sequencing data. BMC Bioinformatics 18, 383 (2017); Fan, J. et al. Linking transcriptional and

genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. Genome Res. 28, 1217–1227 (2018); Campbell, K. R. et al. Clonealign: statistical integration of independent single-cell RNA and DNA sequencing data from human cancers. Genome Biol. 20, 54 (2019); Chen, M., Gunel, M. & Zhao, H. SomatiCA: Identifying, characterizing and quantifying somatic copy number aberrations from cancer genome sequencing data. PLoS ONE 8, e78143 (2013); Serin Harmanci, A., et al., CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. Nat Commun 11, 89 (2020); and Oh, et al., Reliable Analysis of Clinical Tumor-Only Whole-Exome Sequencing Data. JCO Clin Cancer Inform. 2020; 4). In example embodiments, amplifications are detected in XPR1 to identify tumors that are sensitive to inhibition of XPR1:KIDINS220-mediated phosphate export (e.g., inhibitors of PP-InsP synthesis). In example embodiments, FISH is used to detect CNVs. In example embodiments, CNVs are detected by whole-exome sequencing (WES) or targeted panel sequencing. In example embodiments, CNVs are detected by inference from a target sequencing panel. In example embodiments, CNVs are determined using RNA-seq.

Detecting morphological changes

[0083] In example embodiments, morphological changes can be used to determine whether a treatment is effective. In example embodiments, treatment of a subject with an inhibitor of XPR1:KIDINS220-mediated phosphate export (e.g., PP-InsP synthesis inhibitor) results in an increase in vacuole-like structures in tumor cells (see, e.g., Bondeson, et al., 2022 and PCT/US2021/045227). In example embodiments, the morphological changes are detected by microscopy. Microscopy is the technical field of using microscopes to view objects and areas of objects that cannot be seen with the naked eye (objects that are not within the resolution range of the normal eye) (see, e.g., Mualla et al., editors. In: Medical Imaging Systems: An Introductory Guide [Internet]. Cham (CH): Springer; 2018. Chapter 5. 2018 Aug 3. DOI: 10.1007/978-3-319-96520-8_5). Any method of microscopy may be used in the present invention (e.g., optical, electron, and scanning probe microscopy, or X-ray microscopy). In preferred embodiments, phase contrast, fluorescence or confocal microscopy is used.

Other therapeutic agents

[0084] In example embodiments, the one or more therapeutic agents can be used to modify protein or transcript levels of IP6Ks and/or PPIP5Ks. In example embodiments, the one or more therapeutic agents can be used to modify enzymatic activity of IP6Ks and/or PPIP5Ks. In example embodiments, the one or more therapeutic agents can be used to decrease

expression or activity of FGF23. In example embodiments, the one or more therapeutic agents can be used to increase expression or activity of SLC20A1 and/or SLC34A2.

Degrader Molecules

[0085] In one embodiment, the agent is a degrader molecule (see, e.g., Ding, et al., Emerging New Concepts of Degrader Technologies, Trends Pharmacol Sci. 2020 Jul;41(7):464-474). The terms "degrader" and "degrader molecule" refer to all compounds capable of specifically targeting a protein for degradation (e.g., ATTEC, AUTAC, LYTAC, or PROTAC, reviewed in Ding, et al. 2020). Proteolysis Targeting Chimera (PROTAC) technology is a rapidly emerging alternative therapeutic strategy with the potential to address many of the challenges currently faced in modern drug development programs. PROTAC technology employs small molecules that recruit target proteins for ubiquitination and removal by the proteasome (see, e.g., Zhou et al., Discovery of a Small-Molecule Degrader of Bromodomain and Extra- Terminal (BET) Proteins with Picomolar Cellular Potencies and Capable of Achieving Tumor Regression. J. Med. Chem. 2018, 61, 462−481; Bondeson and Crews, Targeted Protein Degradation by Small Molecules, Annu Rev Pharmacol Toxicol. 2017 Jan 6; 57: 107–123; and Lai et al., Modular PROTAC Design for the Degradation of Oncogenic BCR-ABL Angew Chem Int Ed Engl. 2016 Jan 11; 55(2): 807–810).

Genetic Modifying Agents

[0086] In example embodiments, a genetic modifying agent can be used to decrease expression or activity of IP6Ks and/or PPIP5Ks, or FGF23. In example embodiments, a genetic modifying agent can be used to increase expression or activity of SLC20A1 and/or SLC34A2. The genetic modifying agent may comprise a programmable nuclease, such as, a CRISPR system, a zinc finger nuclease system, a TALEN, or a meganuclease. In some embodiments, a polynucleotide of the present invention described elsewhere herein can be modified using a genetic modifying agent.

CRISPR-Cas

[0087] In one example embodiment, the genetic modifying agent is a CRISPR-Cas system. CRISPR-Cas systems comprise a Cas polypeptide and a guide sequence, wherein the guide sequence is capable of forming a CRISPR-Cas complex with the Cas polypeptide and directing site-specific binding of the CRISPR-Cas sequence to a target sequence. The Cas polypeptide may induce a double- or single-stranded break at a designated site in the target sequence. The site of CRISPR-Cas cleavage, for most CRISPR-Cas systems, is dictated by distance from a

protospacer-adjacent motif (PAM), discussed in further detail below. Accordingly, a guide sequence may be selected to direct the CRISPR-Cas system to a desired target site at or near the one or more target genes. Additionally, CRISPR systems can be used *in vivo* (see, e.g., Chen H, Shi M, Gilam A, et al. Hemophilia A ameliorated in mice by CRISPR-based in vivo genome editing of human Factor VIII. Sci Rep. 2019;9(1):16838; Hana S, Peterson M, McLaughlin H, et al. Highly efficient neuronal gene knockout in vivo by CRISPR-Cas9 via neonatal intracerebroventricular injection of AAV in mice. Gene Ther. 2021;28(10-11):646-658; and Rosenblum D, Gutkin A, Kedmi R, et al. CRISPR-Cas9 genome editing using targeted lipid nanoparticles for cancer therapy. Sci Adv. 2020;6(47):eabc9450).

[0088] In general, a CRISPR-Cas or CRISPR system as used in herein and in documents, such as International Patent Publication No. WO 2014/093622 (PCT/US2013/074667), refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). See, e.g, Shmakov et al. (2015) "Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems", Molecular Cell, DOI: dx.doi.org/10.1016/j.molcel.2015.10.008.

[0089] CRISPR-Cas systems can generally fall into two classes based on their architectures of their effector molecules, which are each further subdivided by type and subtype. The two class are Class 1 and Class 2. Class 1 CRISPR-Cas systems have effector modules composed of multiple Cas proteins, some of which form crRNA-binding complexes, while Class 2 CRISPR-Cas systems include a single, multi-domain crRNA-binding protein.

[0090] In some embodiments, the CRISPR-Cas system that can be used to modify a polynucleotide of the present invention described herein can be a Class 1 CRISPR-Cas system.

In some embodiments, the CRISPR-Cas system that can be used to modify a polynucleotide of the present invention described herein can be a Class 2 CRISPR-Cas system.

Class 1 CRISPR-Cas Systems

[0091] In some embodiments, the CRISPR-Cas system that can be used to modify a polynucleotide of the present invention described herein can be a Class 1 CRISPR-Cas system. Class 1 CRISPR-Cas systems are divided into types I, II, and IV. Makarova et al. 2020. Nat. Rev. 18: 67-83., particularly as described in Figure 1. Type I CRISPR-Cas systems are divided into 9 subtypes (I-A, I-B, I-C, I-D, I-E, I-F1, I-F2, I-F3, and IG). Makarova *et al.*, 2020. Class 1, Type I CRISPR-Cas systems can contain a Cas3 protein that can have helicase activity. Type III CRISPR-Cas systems are divided into 6 subtypes (III-A, III-B, III-C, III-D, III-E, and III-F). Type III CRISPR-Cas systems can contain a Cas10 that can include an RNA recognition motif called Palm and a cyclase domain that can cleave polynucleotides. Makarova *et al.*, 2020. Type IV CRISPR-Cas systems are divided into 3 subtypes. (IV-A, IV-B, and IV-C). Makarova *et al.*, 2020. Class 1 systems also include CRISPR-Cas variants, including Type I-A, I-B, I-E, I-F and I-U variants, which can include variants carried by transposons and plasmids, including versions of subtype I-F encoded by a large family of Tn7-like transposon and smaller groups of Tn7-like transposons that encode similarly degraded subtype I-B systems. Peters *et al.*, PNAS 114 (35) (2017); DOI: 10.1073/pnas.1709035114; *see also,* Makarova et al. 2018. The CRISPR Journal, v. 1, n5, Figure 5.

[0092] The Class 1 systems typically comprise a multi-protein effector complex, which can, in some embodiments, include ancillary proteins, such as one or more proteins in a complex referred to as a CRISPR-associated complex for antiviral defense (Cascade), one or more adaptation proteins (e.g., Cas1, Cas2, RNA nuclease), and/or one or more accessory proteins (e.g., Cas 4, DNA nuclease), CRISPR associated Rossman fold (CARF) domain containing proteins, and/or RNA transcriptase.

[0093] The backbone of the Class 1 CRISPR-Cas system effector complexes can be formed by RNA recognition motif domain-containing protein(s) of the repeat-associated mysterious proteins (RAMPs) family subunits (e.g., Cas 5, Cas6, and/or Cas7). RAMP proteins are characterized by having one or more RNA recognition motif domains. In some embodiments, multiple copies of RAMPs can be present. In some embodiments, the Class I CRISPR-Cas system can include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more Cas5, Cas6, and/or Cas 7 proteins. In some embodiments, the Cas6 protein is an RNAse, which can be responsible for pre-crRNA

processing. When present in a Class 1 CRISPR-Cas system, Cas6 can be optionally physically associated with the effector complex.

[0094] Class 1 CRISPR-Cas system effector complexes can, in some embodiments, also include a large subunit. The large subunit can be composed of or include a Cas8 and/or Cas10 protein. *See*, e.g., Figures 1 and 2. Koonin EV, Makarova KS. 2019. Phil. Trans. R. Soc. B 374: 20180087, DOI: 10.1098/rstb.2018.0087 and Makarova et al. 2020.

[0095] Class 1 CRISPR-Cas system effector complexes can, in some embodiments, include a small subunit (for example, Cas11). *See*, e.g., Figures 1 and 2. Koonin EV, Makarova KS. 2019 Origins and Evolution of CRISPR-Cas systems. Phil. Trans. R. Soc. B 374: 20180087, DOI: 10.1098/rstb.2018.0087.

[0096] In some embodiments, the Class 1 CRISPR-Cas system can be a Type I CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-A CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-B CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-C CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-D CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-E CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-F1 CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-F2 CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-F3 CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-G CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a CRISPR Cas variant, such as a Type I-A, I-B, I-E, I-F and I-U variants, which can include variants carried by transposons and plasmids, including versions of subtype I-F encoded by a large family of Tn7-like transposon and smaller groups of Tn7-like transposons that encode similarly degraded subtype I-B systems as previously described.

[0097] In some embodiments, the Class 1 CRISPR-Cas system can be a Type III CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-A CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-B CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-C CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-D CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas

system can be a subtype III-E CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-F CRISPR-Cas system.

[0098] In some embodiments, the Class 1 CRISPR-Cas system can be a Type IV CRISPR-Cas-system. In some embodiments, the Type IV CRISPR-Cas system can be a subtype IV-A CRISPR-Cas system. In some embodiments, the Type IV CRISPR-Cas system can be a subtype IV-B CRISPR-Cas system. In some embodiments, the Type IV CRISPR-Cas system can be a subtype IV-C CRISPR-Cas system.

[0099] The effector complex of a Class 1 CRISPR-Cas system can, in some embodiments, include a Cas3 protein that is optionally fused to a Cas2 protein, a Cas4, a Cas5, a Cas6, a Cas7, a Cas8, a Cas10, a Cas11, or a combination thereof. In some embodiments, the effector complex of a Class 1 CRISPR-Cas system can have multiple copies, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14, of any one or more Cas proteins.

Class 2 CRISPR-Cas Systems

[0100] The compositions, systems, and methods described in greater detail elsewhere herein can be designed and adapted for use with Class 2 CRISPR-Cas systems. Thus, in some embodiments, the CRISPR-Cas system is a Class 2 CRISPR-Cas system. Class 2 systems are distinguished from Class 1 systems in that they have a single, large, multi-domain effector protein. In certain example embodiments, the Class 2 system can be a Type II, Type V, or Type VI system, which are described in Makarova et al. "Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants" Nature Reviews Microbiology, 18:67-81 (Feb 2020), incorporated herein by reference. Each type of Class 2 system is further divided into subtypes. *See* Markova et al. 2020, particularly at Figure. 2. Class 2, Type II systems can be divided into 4 subtypes: II-A, II-B, II-C1, and II-C2. Class 2, Type V systems can be divided into 17 subtypes: V-A, V-B1, V-B2, V-C, V-D, V-E, V-F1, V-F1(V-U3), V-F2, V-F3, V-G, V-H, V-I, V-K (V-U5), V-U1, V-U2, and V-U4. Class 2, Type IV systems can be divided into 5 subtypes: VI-A, VI-B1, VI-B2, VI-C, and VI-D.

[0101] The distinguishing feature of these types is that their effector complexes consist of a single, large, multi-domain protein. Type V systems differ from Type II effectors (e.g., Cas9), which contain two nuclear domains that are each responsible for the cleavage of one strand of the target DNA, with the HNH nuclease inserted inside the Ruv-C like nuclease domain sequence. The Type V systems (e.g., Cas12) only contain a RuvC-like nuclease domain that cleaves both strands. Type VI (Cas13) are unrelated to the effectors of Type II and V systems

and contain two HEPN domains and target RNA. Cas13 proteins also display collateral activity that is triggered by target recognition. Some Type V systems have also been found to possess this collateral activity with two single-stranded DNA in *in vitro* contexts.

[0102] In some embodiments, the Class 2 system is a Type II system. In some embodiments, the Type II CRISPR-Cas system is a II-A CRISPR-Cas system. In some embodiments, the Type II CRISPR-Cas system is a II-B CRISPR-Cas system. In some embodiments, the Type II CRISPR-Cas system is a II-C1 CRISPR-Cas system. In some embodiments, the Type II CRISPR-Cas system is a II-C2 CRISPR-Cas system. In some embodiments, the Type II system is a Cas9 system. In some embodiments, the Type II system includes a Cas9.

[0103] In some embodiments, the Class 2 system is a Type V system. In some embodiments, the Type V CRISPR-Cas system is a V-A CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-B1 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-B2 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-C CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-D CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-E CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F1 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F1 (V-U3) CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F2 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F3 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-G CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-H CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-I CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-K (V-U5) CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-U1 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-U2 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-U4 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system includes a Cas12a (Cpf1), Cas12b (C2c1), Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas14, and/or CasΦ.

[0104] In some embodiments, the Class 2 system is a Type VI system. In some embodiments, the Type VI CRISPR-Cas system is a VI-A CRISPR-Cas system. In some

embodiments, the Type VI CRISPR-Cas system is a VI-B1 CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-B2 CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-C CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-D CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system includes a Cas13a (C2c2), Cas13b (Group 29/30), Cas13c, and/or Cas13d.

Guide Molecules

[0105] The following include general design principles that may be applied to the guide molecule. The terms guide molecule, guide sequence and guide polynucleotide refer to polynucleotides capable of guiding Cas to a target genomic locus and are used interchangeably as in foregoing cited documents such as International Patent Publication No. WO 2014/093622 (PCT/US2013/074667). In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. The guide molecule can be a polynucleotide.

[0106] The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay (Qui et al. 2004. BioTechniques. 36(4)702-707). Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible and will occur to those skilled in the art.

[0107] In some embodiments, the guide molecule is an RNA. The guide molecule(s) (also referred to interchangeably herein as guide polynucleotide and guide sequence) that are included in the CRISPR-Cas or Cas based system can be any polynucleotide sequence having

sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting examples of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net).

[0108] A guide sequence, and hence a nucleic acid-targeting guide, may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[0109] In some embodiments, a nucleic acid-targeting guide is selected to reduce the degree secondary structure within the nucleic acid-targeting guide. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold,

developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (*see* e.g., A.R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0110] In one example embodiment, a guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In another example embodiment, the guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In another example embodiment, the direct repeat sequence may be located upstream (i.e., 5') from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3') from the guide sequence or spacer sequence.

[0111] In one example embodiment, the crRNA comprises a stem loop, preferably a single stem loop. In one example embodiment, the direct repeat sequence forms a stem loop, preferably a single stem loop.

[0112] In one example embodiment, the spacer length of the guide RNA is from 15 to 35 nt. In another example embodiment, the spacer length of the guide RNA is at least 15 nucleotides. In another example embodiment, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27 to 30 nt, e.g., 27, 28, 29, or 30 nt, from 30 to 35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer.

[0113] The "tracrRNA" sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In some embodiments, the degree of complementarity between the tracrRNA sequence and crRNA sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and crRNA sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin.

[0114] In general, degree of complementarity is with reference to the optimal alignment of the sca sequence and tracr sequence, along the length of the shorter of the two sequences.

Optimal alignment may be determined by any suitable alignment algorithm and may further account for secondary structures, such as self-complementarity within either the sca sequence or tracr sequence. In some embodiments, the degree of complementarity between the tracr sequence and sca sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher.

[0115] In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and tracr RNA can be 30 or 50 nucleotides in length. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82% or 81% or 80% complementarity between the sequence and the guide, with it being advantageous that off target is 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

[0116] In some embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All of (1) to (3) may reside in a single RNA, i.e., an sgRNA (arranged in a 5' to 3' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence. Where the tracr RNA is on a different RNA than the RNA containing the guide and tracr sequence, the length of each RNA may be optimized to be shortened from their respective native lengths, and each may be independently chemically modified to protect from degradation by cellular RNase or otherwise increase stability.

[0117] Many modifications to guide sequences are known in the art and are further contemplated within the context of this invention. Various modifications may be used to increase the specificity of binding to the target sequence and/or increase the activity of the Cas protein and/or reduce off-target effects. Example guide sequence modifications are described in International Patent Application No. PCT US2019/045582, specifically paragraphs [0178]-[0333]. which is incorporated herein by reference.

Target Sequences, PAMs, and PFSs

[0118] In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. In other words, the target polynucleotide can be a polynucleotide or a part of a polynucleotide to which a part of the guide sequence is designed to have complementarity with and to which the effector function mediated by the complex comprising the CRISPR effector protein and a guide molecule is to be directed. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

[0119] PAM elements are sequences that can be recognized and bound by Cas proteins. Cas proteins/effector complexes can then unwind the dsDNA at a position adjacent to the PAM element. It will be appreciated that Cas proteins and systems target RNA do not require PAM sequences (Marraffini et al. 2010. Nature. 463:568-571). Instead, many rely on PFSs, which are discussed elsewhere herein. In one example embodiment, the target sequence should be associated with a PAM (protospacer adjacent motif) or PFS (protospacer flanking sequence or site), that is, a short sequence recognized by the CRISPR complex. Depending on the nature of the CRISPR-Cas protein, the target sequence should be selected, such that its complementary sequence in the DNA duplex (also referred to herein as the non-target sequence) is upstream or downstream of the PAM. In the embodiments, the complementary sequence of the target sequence is downstream or 3' of the PAM or upstream or 5' of the PAM. The precise sequence and length requirements for the PAM differ depending on the Cas protein used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). Examples of the natural PAM sequences for different Cas proteins are provided herein below and the skilled person will be able to identify further PAM sequences for use with a given Cas protein.

[0120] The ability to recognize different PAM sequences depends on the Cas polypeptide(s) included in the system. *See* e.g., Gleditzsch et al. 2019. RNA Biology. 16(4):504-517. **Table A** (from Gleditzsch et al. 2019) below shows several Cas polypeptides and the PAM sequence they recognize.

Table A – Example PAM Sequences	
Cas Protein	PAM Sequence
SpCas9	NGG/NRG
SaCas9	NGRRT or NGRRN
NmeCas9	NNNNGATT
CjCas9	NNNNRYAC
StCas9	NNAGAAW
Cas12a (Cpf1) (including LbCpf1 and AsCpf1)	TTTV
Cas12b (C2c1)	TTT, TTA, and TTC
Cas12c (C2c3)	TA
Cas12d (CasY)	TA
Cas12e (CasX)	5′-TTCN-3′
Cas1	5′-CTT-3′
Cas8e	5′-ATG-3′
Type I-A	5′-CCN-3′
Type I-B	TTC, ACT, TAA, TAT, TAG, and CAC
Type I-C	NTTC
Type I-E	5′-AAG-3′
TypeI-F	GG

[0121] In a preferred embodiment, the CRISPR effector protein may recognize a 3' PAM. In one example embodiment, the CRISPR effector protein may recognize a 3' PAM which is 5'H, wherein H is A, C or U.

[0122] Further, engineering of the PAM Interacting (PI) domain on the Cas protein may allow programing of PAM specificity, improve target site recognition fidelity, and increase the versatility of the CRISPR-Cas protein, for example as described for Cas9 in Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592. As further detailed herein, the skilled person will understand that Cas13 proteins may be modified analogously. Gao *et al*, "Engineered Cpf1 Enzymes with Altered PAM Specificities," bioRxiv 091611; doi: http://dx.doi.org/10.1101/091611 (Dec. 4, 2016). Doench et al. created a pool of sgRNAs, tiling across all possible target sites of a panel of six endogenous mouse and three endogenous human

genes and quantitatively assessed their ability to produce null alleles of their target gene by antibody staining and flow cytometry. The authors showed that optimization of the PAM improved activity and also provided an on-line tool for designing sgRNAs.

[0123] PAM sequences can be identified in a polynucleotide using an appropriate design tool, which are commercially available as well as online. Such freely available tools include, but are not limited to, CRISPRFinder and CRISPRTarget. Mojica et al. 2009. Microbiol. 155(Pt. 3):733-740; Atschul et al. 1990. J. Mol. Biol. 215:403-410; Biswass et al. 2013 RNA Biol. 10:817-827; and Grissa et al. 2007. Nucleic Acid Res. 35:W52-57. Experimental approaches to PAM identification can include, but are not limited to, plasmid depletion assays (Jiang et al. 2013. Nat. Biotechnol. 31:233-239; Esvelt et al. 2013. Nat. Methods. 10:1116-1121; Kleinstiver et al. 2015. Nature. 523:481-485), screened by a high-throughput *in vivo* model called PAM-SCNAR (Pattanayak et al. 2013. Nat. Biotechnol. 31:839-843 and Leenay et al. 2016.Mol. Cell. 16:253), and negative screening (Zetsche et al. 2015. Cell. 163:759-771).

[0124] As previously mentioned, CRISPR-Cas systems that target RNA do not typically rely on PAM sequences. Instead, such systems typically recognize protospacer flanking sites (PFSs) instead of PAMs Thus, Type VI CRISPR-Cas systems typically recognize protospacer flanking sites (PFSs) instead of PAMs. PFSs represents an analogue to PAMs for RNA targets. Type VI CRISPR-Cas systems employ a Cas13. Some Cas13 proteins analyzed to date, such as Cas13a (C2c2) identified from Leptotrichia shahii (LShCAs13a) have a specific discrimination against G at the 3'end of the target RNA. The presence of a C at the corresponding crRNA repeat site can indicate that nucleotide pairing at this position is rejected. However, some Cas13 proteins (e.g., LwaCAs13a and PspCas13b) do not seem to have a PFS preference. See e.g., Gleditzsch et al. 2019. RNA Biology. 16(4):504-517.

[0125] Some Type VI proteins, such as subtype B, have 5′-recognition of D (G, T, A) and a 3′-motif requirement of NAN or NNA. One example is the Cas13b protein identified in Bergeyella zoohelcum (BzCas13b). See e.g., Gleditzsch et al. 2019. RNA Biology. 16(4):504-517.

[0126] Overall Type VI CRISPR-Cas systems appear to have less restrictive rules for substrate (e.g., target sequence) recognition than those that target DNA (e.g., Type V and type II).

Sequences related to nucleus targeting and transportation

[0127] In some embodiments, one or more components (e.g., the Cas protein) in the composition for engineering cells may comprise one or more sequences related to nucleus targeting and transportation. Such sequences may facilitate the one or more components in the composition for targeting a sequence within a cell. In order to improve targeting of the CRISPR-Cas protein used in the methods of the present disclosure to the nucleus, it may be advantageous to provide one or both of these components with one or more nuclear localization sequences (NLSs).

[0128] In one example embodiment, the NLSs used in the context of the present disclosure are heterologous to the proteins. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV (SEQ ID NO:1) or PKKKRKVEAS (SEQ ID NO:2); the NLS from nucleoplasmin (e.g., the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID NO:3)); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO:4) or RQRRNELKRSP (SEQ ID NO:5); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:6); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO:7) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO:8) and PPKKARED (SEQ ID NO:9) of the myoma T protein; the sequence PQPKKKPL (SEQ ID NO:10) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO:11) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO:12) and PKQKKRK (SEQ ID NO:13) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO:14) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO:15) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO:16) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO:17) of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs are of sufficient strength to drive accumulation of the DNA-targeting Cas protein in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the CRISPR-Cas protein, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the nucleic acid-targeting protein, such that location within a cell may be visualized, such as in combination

with a means for detecting the location of the nucleus (e.g., a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of nucleic acid-targeting complex formation (e.g., assay for deaminase activity) at the target sequence, or assay for altered gene expression activity affected by DNA-targeting complex formation and/or DNA-targeting), as compared to a control not exposed to the Cas protein, or exposed to a Cas protein lacking the one or more NLSs.

[0129] The Cas proteins may be provided with 1 or more, such as with, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more heterologous NLSs. In some embodiments, the proteins comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g., zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. In preferred embodiments of the Cas proteins, an NLS attached to the C-terminal of the protein.

[0130] In certain embodiments, the CRISPR-Cas protein and a functional domain protein (described further herein) are delivered to the cell or expressed within the cell as separate proteins. In these embodiments, each of the CRISPR-Cas and functional domain protein can be provided with one or more NLSs as described herein. In certain embodiments, the CRISPR-Cas and functional domain protein are delivered to the cell or expressed with the cell as a fusion protein. In these embodiments one or both of the CRISPR-Cas and functional domain protein is provided with one or more NLSs. Where the functional domain protein is fused to an adaptor protein (such as MS2) as described above, the one or more NLS can be provided on the adaptor protein, provided that this does not interfere with aptamer binding. In particular embodiments, the one or more NLS sequences may also function as linker sequences between the functional domain protein and the CRISPR-Cas protein.

[0131] In certain embodiments, guides of the disclosure comprise specific binding sites (e.g., aptamers) for adapter proteins, which may be linked to or fused to a functional domain protein or catalytic domain thereof. When such a guide forms a CRISPR complex (e.g., CRISPR-Cas protein binding to guide and target), the adapter proteins bind and the functional domain protein or catalytic domain thereof associated with the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective.

[0132] The skilled person will understand that modifications to the guide which allow for binding of the adapter + nucleotide deaminase, but not proper positioning of the adapter + nucleotide deaminase (e.g., due to steric hindrance within the three-dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and in some cases at both the tetra loop and stem loop 2.

[0133] In some embodiments, a component (e.g., the dead Cas protein, the functional domain protein or catalytic domain thereof, or a combination thereof) in the systems may comprise one or more nuclear export signals (NES), one or more nuclear localization signals (NLS), or any combinations thereof. In some cases, the NES may be an HIV Rev NES. In certain cases, the NES may be MAPK NES. When the component is a protein, the NES or NLS may be at the C terminus of component. Alternatively or additionally, the NES or NLS may be at the N terminus of component. In some examples, the Cas protein and optionally said functional domain protein or catalytic domain thereof comprise one or more heterologous nuclear export signal(s) (NES(s)) or nuclear localization signal(s) (NLS(s)), preferably an HIV Rev NES or MAPK NES, preferably C-terminal.

CRISPR-Cas Cleavage

[0134] In one example embodiment, the CRISPR-Cas system may induce a double- or single-stranded break at a designated site in the target sequence. The CRISPR-Cas system may introduce an indel, which, as used herein, refers to insertions or deletions of the DNA at particular locations on the chromosome. The site of CRISPR-Cas cleavage, for most CRISPR-Cas systems, is dictated by distance from a protospacer-adjacent motif (PAM). Accordingly, a guide sequence may be selected to direct the CRISPR-Cas system to induce cleavage at a desired target site at or near the one or more variants.

NHEJ-Based Editing

[0135] In one example embodiment, the CRISPR-Cas system is used to introduce one or more insertions or deletions to a target sequence on the gene or enhancer associated with the gene such that one or more indels or insertions reduce expression or activity of the one or more polypeptides. More than one guide sequence may be selected to insert multiple insertion, deletions, or combination thereof. Likewise, more than one Cas protein type may be used, for example, to maximize targets sites adjacent to different PAMs. In one example embodiment, a guide sequence is selected that directs the CRISPR-Cas system to make one or more insertions or deletions within the enhancer region. In one example embodiment, a guide is selected that directs the CRISPR-Cas system to make an insertion 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 base pairs upstream of an enhancer controlling expression of a target gene. In one example embodiment, a guide sequence is selected to that directs the CRISPR-Cas system to make an insertion 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 base pairs downstream of an enhancer controlling expression of a target gene. In one example embodiment, a guide sequence is selected that directs the CRISPR-Cas system to make a deletion 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 base pairs downstream of an enhancer controlling expression of a target gene. In one example embodiment, a guide sequence is selected that directs the CRISPR-Cas system to make a deletion 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 base pairs downstream of an enhancer controlling expression of a target gene.

HDR Template Based Editing

[0136] In one example embodiment, a donor template is provided to replace a genomic sequence in a target gene or sequence controlling expression of the target gene. A donor template may comprise an insertion sequence flanked by two homology regions. The insertion sequence comprises an edited sequence to be inserted in place of the target sequence (e.g., a portion of genomic DNA to be edited). The homology regions comprise sequences that are homologous to the genomic DNA strands at the site of the CRISPR-Cas induced double-strand break. Cellular HDR mechanisms then facilitate insertion of the insertion sequence at the site of the DSB.

[0001] Accordingly, in certain example embodiments, a donor template and guide sequence are selected to direct excision and replacement of a section of genome DNA

comprising an enhancer controlling expression of a target gene or a section of genome DNA within the gene that is required for activity of the target gene. In one example embodiment, the insertion sequence comprises a transcription factor binding site that recruits a repressor to the gene.

[0137] The donor template may include a sequence which results in a change in sequence of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more nucleotides of the target sequence.

[0138] A donor template may be of any suitable length, such as about or more than about 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In an embodiment, the template nucleic acid may be 20+/- 10, 30+/- 10, 40+/- 10, 50+/- 10, 60+/- 10, 70+/- 10, 80+/- 10, 90+/- 10, 100+/- 10, 1 10+/- 10, 120+/- 10, 130+/- 10, 140+/- 10, 150+/- 10, 160+/- 10, 170+/- 10, 180+/- 10, 190+/- 10, 200+/- 10, 210+/-10, or 220+/- 10 nucleotides in length. In an embodiment, the template nucleic acid may be 30+/-20, 40+/-20, 50+/-20, 60+/- 20, 70+/- 20, 80+/-20, 90+/-20, 100+/-20, 1 10+/-20, 120+/-20, 130+/-20, 140+/-20, 150+/-20, 160+/-20, 170+/-20, 180+/-20, 190+/-20, 200+/-20, 210+/-20, or 220+/-20 nucleotides in length. In an embodiment, the template nucleic acid is 10 to 1 ,000, 20 to 900, 30 to 800, 40 to 700, 50 to 600, 50 to 500, 50 to 400, 50 to300, 50 to 200, or 50 to 100 nucleotides in length.

[0139] The homology regions of the donor template may be complementary to a portion of a polynucleotide comprising the target sequence. When optimally aligned, a donor template might overlap with one or more nucleotides of a target sequences (e.g., about or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more nucleotides). In some embodiments, when a template sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the template polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

[0140] The donor template comprises a sequence to be integrated (e.g., a mutated gene). The sequence for integration may be a sequence endogenous or exogenous to the cell. Examples of a sequence to be integrated include polynucleotides encoding a protein or a non-coding RNA (e.g., a microRNA). Thus, the sequence for integration may be operably linked to an appropriate control sequence or sequences. Alternatively, the sequence to be integrated may provide a regulatory function.

[0141] Homology arms of the donor template may comprise from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200,

1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence have about 200 bp to about 2000 bp, about 600 bp to about 1000 bp, or more particularly about 700 bp to about 1000.

[0142] In one example embodiment, one or both homology arms may be shortened to avoid including certain sequence repeat elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In other embodiments, a 3' homology arm may be shortened to avoid a sequence repeat element. In some embodiments, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

[0143] The donor template may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The donor template of the disclosure can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996).

[0144] In one example embodiment, a donor template is a single-stranded oligonucleotide. When using a single-stranded oligonucleotide, 5' and 3' homology arms may range up to about 200 base pairs (bp) in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length.

[0145] Suzuki *et al.* describe *in vivo* genome editing via CRISPR/Cas9 mediated homology-independent targeted integration (2016, Nature 540:144–149).

Templates

[0146] In some embodiments, a composition for engineering cells comprises a template, e.g., a recombination template. A template may be a component of another vector as described herein, contained in a separate vector, or provided as a separate polynucleotide. In some embodiments, a recombination template is designed to serve as a template in homologous recombination, such as within or near a target sequence nicked or cleaved by a nucleic acid-targeting effector protein as a part of a nucleic acid-targeting complex.

[0147] In an embodiment, the template nucleic acid alters the sequence of the target position. In an embodiment, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

[0148] The template sequence may undergo a breakage mediated or catalyzed recombination with the target sequence. In an embodiment, the template nucleic acid may include sequence that corresponds to a site on the target sequence that is cleaved by a Cas protein mediated cleavage event. In an embodiment, the template nucleic acid may include a

sequence that corresponds to both, a first site on the target sequence that is cleaved in a first Cas protein mediated event, and a second site on the target sequence that is cleaved in a second Cas protein mediated event.

[0149] In certain embodiments, the template nucleic acid can include a sequence which results in an alteration in the coding sequence of a translated sequence, e.g., one which results in the substitution of one amino acid for another in a protein product, e.g., transforming a mutant allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or introducing a stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a nonsense mutation. In certain embodiments, the template nucleic acid can include a sequence which results in an alteration in a non-coding sequence, e.g., an alteration in an exon or in a 5' or 3' non-translated or non-transcribed region. Such alterations include an alteration in a control element, e.g., a promoter, enhancer, and an alteration in a cis-acting or trans-acting control element.

[0150] A template nucleic acid having homology with a target position in a target gene may be used to alter the structure of a target sequence. The template sequence may be used to alter an unwanted structure, e.g., an unwanted or mutant nucleotide. The template nucleic acid may include a sequence which, when integrated, results in decreasing the activity of a positive control element; increasing the activity of a positive control element; decreasing the activity of a negative control element; increasing the activity of a negative control element; decreasing the expression of a gene; increasing the expression of a gene; increasing resistance to a disorder or disease; increasing resistance to viral entry; correcting a mutation or altering an unwanted amino acid residue conferring, increasing, abolishing or decreasing a biological property of a gene product, e.g., increasing the enzymatic activity of an enzyme, or increasing the ability of a gene product to interact with another molecule.

[0151] The template nucleic acid may include a sequence which results in a change in sequence of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more nucleotides of the target sequence.

[0152] A template polynucleotide may be of any suitable length, such as about or more than about 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In an embodiment, the template nucleic acid may be 20+/- 10, 30+/- 10, 40+/- 10, 50+/- 10, 60+/- 10, 70+/- 10, 80+/- 10, 90+/- 10, 100+/- 10, 1 10+/- 10, 120+/- 10, 130+/- 10, 140+/- 10, 150+/- 10, 160+/- 10, 170+/- 10, 180+/- 10, 190+/- 10, 200+/- 10, 210+/-10, or 220+/- 10 nucleotides in length. In an embodiment, the template nucleic acid may be 30+/-20, 40+/-20, 50+/-20, 60+/-

20, 70+/- 20, 80+/-20, 90+/-20, 100+/-20, 110+/-20, 120+/-20, 130+/-20, 140+/-20, 150+/-20, 160+/-20, 170+/-20, 180+/-20, 190+/-20, 200+/-20, 210+/-20, or 220+/-20 nucleotides in length. In an embodiment, the template nucleic acid is 10 to 1 ,000, 20 to 900, 30 to 800, 40 to 700, 50 to 600, 50 to 500, 50 to 400, 50 to300, 50 to 200, or 50 to 100 nucleotides in length.

[0153] In some embodiments, the template polynucleotide is complementary to a portion of a polynucleotide comprising the target sequence. When optimally aligned, a template polynucleotide might overlap with one or more nucleotides of a target sequences (e.g., about or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more nucleotides). In some embodiments, when a template sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the template polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

[0154] The exogenous polynucleotide template comprises a sequence to be integrated (e.g., a mutated gene). The sequence for integration may be a sequence endogenous or exogenous to the cell. Examples of a sequence to be integrated include polynucleotides encoding a protein or a non-coding RNA (e.g., a microRNA). Thus, the sequence for integration may be operably linked to an appropriate control sequence or sequences. Alternatively, the sequence to be integrated may provide a regulatory function.

[0155] An upstream or downstream sequence may comprise from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence have about 200 bp to about 2000 bp, about 600 bp to about 1000 bp, or more particularly about 700 bp to about 1000.

[0156] An upstream or downstream sequence may comprise from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence have about 200 bp to about 2000 bp, about 600 bp to about 1000 bp, or more particularly about 700 bp to about 1000 bp.

[0157] In certain embodiments, one or both homology arms may be shortened to avoid including certain sequence repeat elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In other embodiments, a 3' homology arm may be shortened

to avoid a sequence repeat element. In some embodiments, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

[0158] In some methods, the exogenous polynucleotide template may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The exogenous polynucleotide template of the disclosure can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996).

[0159] In certain embodiments, a template nucleic acid for correcting a mutation may designed for use as a single-stranded oligonucleotide. When using a single-stranded oligonucleotide, 5' and 3' homology arms may range up to about 200 base pairs (bp) in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length.

[0160] Suzuki et al. describe *in vivo* genome editing via CRISPR/Cas9 mediated homology-independent targeted integration (2016, Nature 540:144–149).

Specialized Cas-based Systems

[0161] In some embodiments, the system is a Cas-based system that is capable of performing a specialized function or activity. For example, the Cas protein may be fused, operably coupled to, or otherwise associated with one or more functionals domains. In certain example embodiments, the Cas protein may be a catalytically dead Cas protein ("dCas") and/or have nickase activity. A nickase is a Cas protein that cuts only one strand of a double stranded target. In such embodiments, the dCas or nickase provide a sequence specific targeting functionality that delivers the functional domain to or proximate a target sequence. Example functional domains that may be fused to, operably coupled to, or otherwise associated with a Cas protein can be or include, but are not limited to a nuclear localization signal (NLS) domain, a nuclear export signal (NES) domain, a translational activation domain, a transcriptional activation domain (e.g. VP64, p65, MyoD1, HSF1, RTA, and SET7/9), a translation initiation domain, a transcriptional repression domain (e.g., a KRAB domain, NuE domain, NcoR domain, and a SID domain such as a SID4X domain), a nuclease domain (e.g., FokI), a histone modification domain (e.g., a histone acetyltransferase), a light inducible/controllable domain, a chemically inducible/controllable domain, a transposase domain, a homologous recombination machinery domain, a recombinase domain, an integrase domain, and combinations thereof. Methods for generating catalytically dead Cas9 or a nickase Cas9 (WO 2014/204725, Ran et al. Cell. 2013 Sept 12; 154(6):1380-1389), Cas12 (Liu et al. Nature

Communications, 8, 2095 (2017) , and Cas13 (International Patent Publication Nos. WO 2019/005884 and WO2019/060746) are known in the art and incorporated herein by reference.

[0162] In some embodiments, the functional domains can have one or more of the following activities: methylase activity, demethylase activity, translation activation activity, translation initiation activity, translation repression activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity, molecular switch activity, chemical inducibility, light inducibility, and nucleic acid binding activity. In some embodiments, the one or more functional domains may comprise epitope tags or reporters. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporters include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and auto-fluorescent proteins including blue fluorescent protein (BFP).

[0163] The one or more functional domain(s) may be positioned at, near, and/or in proximity to a terminus of the effector protein (e.g., a Cas protein). In embodiments having two or more functional domains, each of the two can be positioned at or near or in proximity to a terminus of the effector protein (e.g., a Cas protein). In some embodiments, such as those where the functional domain is operably coupled to the effector protein, the one or more functional domains can be tethered or linked via a suitable linker (including, but not limited to, GlySer linkers) to the effector protein (e.g., a Cas protein). When there is more than one functional domain, the functional domains can be same or different. In some embodiments, all the functional domains are the same. In some embodiments, all of the functional domains are different from each other. In some embodiments, at least two of the functional domains are different from each other. In some embodiments, at least two of the functional domains are the same as each other.

[0164] Other suitable functional domains can be found, for example, in International Patent Publication No. WO 2019/018423.

Split CRISPR-Cas systems

[0165] In one example embodiment, the CRISPR-Cas system is a split CRISPR-Cas system. *See* e.g., Zetche et al., 2015. Nat. Biotechnol. 33(2): 139-142 and International Patent Publication WO 2019/018423, the compositions and techniques of which can be used in and/or adapted for use with the present invention. Split CRISPR-Cas proteins are set forth herein and in documents incorporated herein by reference in further detail herein. In certain embodiments, each part of a split CRISPR protein is attached to a member of a specific binding pair, and when bound with each other, the members of the specific binding pair maintain the parts of the CRISPR protein in proximity. In certain embodiments, each part of a split CRISPR protein is associated with an inducible binding pair. An inducible binding pair is one which is capable of being switched "on" or "off" by a protein or small molecule that binds to both members of the inducible binding pair. In some embodiments, CRISPR proteins may preferably split between domains, leaving domains intact. In particular embodiments, said Cas split domains (e.g., RuvC and HNH domains in the case of Cas9) can be simultaneously or sequentially introduced into the cell such that said split Cas domain(s) process the target nucleic acid sequence in the algae cell. The reduced size of the split Cas compared to the wild type Cas allows other methods of delivery of the systems to the cells, such as the use of cell penetrating peptides as described herein.

DNA and RNA Base Editing

[0166] In one example embodiment, the gene editing system configured to modify the gene encoding the one or more polypeptides disclosed herein is a base editing system. In one example embodiment, a Cas protein is connected or fused to a nucleotide deaminase. As used herein, "base editing" refers generally to the process of polynucleotide modification via a CRISPR-Cas-based or Cas-based system that does not include excising nucleotides to make the modification. Base editing can convert base pairs at precise locations without generating excess undesired editing byproducts that can be made using traditional CRISPR-Cas systems.

[0167] In one example embodiment, the nucleotide deaminase may be a DNA base editor used in combination with a DNA binding Cas protein such as, but not limited to, Class 2 Type II and Type V systems. Two classes of DNA base editors are generally known: cytosine base editors (CBEs) and adenine base editors (ABEs). CBEs convert a C•G base pair into a T•A base pair (Komor et al. 2016. Nature. 533:420-424; Nishida et al. 2016. Science. 353; and Li et al. Nat. Biotech. 36:324-327) and ABEs convert an A•T base pair to a G•C base pair.

Collectively, CBEs and ABEs can mediate all four possible transition mutations (C to T, A to G, T to C, and G to A). Rees and Liu. 2018.Nat. Rev. Genet. 19(12): 770-788, particularly at Figures 1b, 2a-2c, 3a-3f, and Table 1. In some embodiments, the base editing system includes a CBE and/or an ABE. In some embodiments, a polynucleotide of the present invention described elsewhere herein can be modified using a base editing system. Rees and Liu. 2018. Nat. Rev. Gent. 19(12):770-788. Base editors also generally do not need a DNA donor template and/or rely on homology-directed repair. Komor et al. 2016. Nature. 533:420-424; Nishida et al. 2016. Science. 353; and Gaudeli et al. 2017. Nature. 551:464-471. Upon binding to a target locus in the DNA, base pairing between the guide RNA of the system and the target DNA strand leads to displacement of a small segment of ssDNA in an "R-loop". Nishimasu et al. Cell. 156:935-949. DNA bases within the ssDNA bubble are modified by the enzyme component, such as a deaminase. In some systems, the catalytically disabled Cas protein can be a variant or modified Cas can have nickase functionality and can generate a nick in the non-edited DNA strand to induce cells to repair the non-edited strand using the edited strand as a template. Komor et al. 2016. Nature. 533:420-424; Nishida et al. 2016. Science. 353; and Gaudeli et al. 2017. Nature. 551:464-471.

[0168] Other Example Type V base editing systems are described in International Patent Publication Nos. WO 2018/213708, WO 2018/213726, and International Patent Applications No. PCT/US2018/067207, PCT/US2018/067225, and PCT/US2018/067307, each of which is incorporated herein by reference.

[0169] In one example embodiment, the base editing system may be an RNA base editing system. As with DNA base editors, a nucleotide deaminase capable of converting nucleotide bases may be fused to a Cas protein. However, in these embodiments, the Cas protein will need to be capable of binding RNA. Example RNA binding Cas proteins include, but are not limited to, RNA-binding Cas9s such as *Francisella novicida* Cas9 ("FnCas9"), and Class 2 Type VI Cas systems. The nucleotide deaminase may be a cytidine deaminase or an adenosine deaminase, or an adenosine deaminase engineered to have cytidine deaminase activity. In certain example embodiments, the RNA base editor may be used to delete or introduce a post-translation modification site in the expressed mRNA. In contrast to DNA base editors, whose edits are permanent in the modified cell, RNA base editors can provide edits where finer, temporal control may be needed, for example in modulating a particular immune response. Example Type VI RNA-base editing systems are described in Cox et al. 2017. Science 358:

1019-1027, International Patent Publication Nos. WO 2019/005884, WO 2019/005886, and WO 2019/071048, and International Patent Application Nos. PCT/US20018/05179 and PCT/US2018/067207, which are incorporated herein by reference. An example FnCas9 system that may be adapted for RNA base editing purposes is described in International Patent Publication No. WO 2016/106236, which is incorporated herein by reference.

[0170] An example method for delivery of base-editing systems, including use of a split-intein approach to divide CBE and ABE into reconstitutable halves, is described in Levy et al. Nature Biomedical Engineering doi.org/10.1038/s41441-019-0505-5 (2019), which is incorporated herein by reference.

Prime Editors

[0171] In one example embodiment, the gene editing system configured to modify a gene encoding the one or more polypeptides disclosed herein is a prime editing system. See e.g., Anzalone et al. 2019. Nature. 576: 149-157. In one example embodiment, a genomic sequence in a target gene or sequence controlling expression of the target gene is replaced or deleted using a prime editing system. Like base editing systems, prime editing systems can be capable of targeted modification of a polynucleotide without generating double stranded breaks. Further prime editing systems are capable of all 12 possible combination swaps. Prime editing may operate via a "search-and-replace" methodology and can mediate targeted insertions, deletions, of all 12 possible base-to-base conversion and combinations thereof. Generally, a prime editing system, as exemplified by PE1, PE2, and PE3 (*Id.*), can include a reverse transcriptase fused or otherwise coupled or associated with an RNA-programmable nickase and a prime-editing extended guide RNA (pegRNA) to facility direct copying of genetic information from the extension on the pegRNA into the target polynucleotide. Embodiments that can be used with the present invention include these and variants thereof. Prime editing can have the advantage of lower off-target activity.

[0172] In some embodiments, the prime editing guide molecule can specify both the target polynucleotide information (e.g., sequence) and contain a new polynucleotide cargo that replaces target polynucleotides. To initiate transfer from the guide molecule to the target polynucleotide, the PE system can nick the target polynucleotide at a target side to expose a 3'hydroxyl group, which can prime reverse transcription of an edit-encoding extension region of the guide molecule (e.g., a prime editing guide molecule or peg guide molecule) directly

into the target site in the target polynucleotide. See e.g., Anzalone et al. 2019. Nature. 576: 149-157, particularly at Figures 1b, 1c, related discussion, and Supplementary discussion.

[0173] In some embodiments, a prime editing system can be composed of a Cas polypeptide having nickase activity, a reverse transcriptase, and a guide molecule. The Cas polypeptide can lack nuclease activity. The guide molecule can include a target binding sequence as well as a primer binding sequence and a template containing the edited polynucleotide sequence. The guide molecule, Cas polypeptide, and/or reverse transcriptase can be coupled together or otherwise associate with each other to form an effector complex and edit a target sequence. In some embodiments, the Cas polypeptide is a Class 2, Type V Cas polypeptide. In some embodiments, the Cas polypeptide is a Cas9 polypeptide (e.g., is a Cas9 nickase). In some embodiments, the Cas polypeptide is fused to the reverse transcriptase. In some embodiments, the Cas polypeptide is linked to the reverse transcriptase.

[0174] In some embodiments, the prime editing system can be a PE1 system or variant thereof, a PE2 system or variant thereof, or a PE3 (e.g., PE3, PE3b) system. See e.g., Anzalone et al. 2019. Nature. 576: 149-157, particularly at pgs. 2-3, Figs. 2a, 3a-3f, 4a-4b, Extended data Figs. 3a-3b, 4.

[0175] The peg guide molecule can be about 10 to about 200 or more nucleotides in length, such as 10 to/or 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, or 200 or more nucleotides in length. Optimization of the peg guide molecule can be accomplished as described in Anzalone et al. 2019. Nature. 576: 149-157, particularly at pg. 3, Fig. 2a-2b, and Extended Data Figs. 5a-c.

CRISPR Associated Transposase (CAST) Systems

[0176] In one example embodiment, the gene editing system configured to modify a gene encoding the one or more polypeptides disclosed herein is a CRISPR associated transposase

system (CAST). In one example embodiment, a CAST system is used to replace all or a portion of an enhancer controlling the target gene expression. CAST system can include a Cas protein that is catalytically inactive, or engineered to be catalytically active, and further comprises a transposase (or subunits thereof) that catalyze RNA-guided DNA transposition. Such systems are able to insert DNA sequences at a target site in a DNA molecule without relying on host cell repair machinery. CAST systems can be Class1 or Class 2 CAST systems. An example Class 1 system is described in Klompe *et al.* Nature, doi:10.1038/s41586-019-1323, which is in incorporated herein by reference. An example Class 2 system is described in Strecker et al. Science. 10/1126/science. aax9181 (2019), and PCT/US2019/066835 which are incorporated herein by reference.

Epigenetic Editing

[0177] In one example embodiment, the one or more agents is an epigenetic modification polypeptide comprising a DNA binding domain linked to or otherwise capable of associating with an epigenetic modification domain such that binding of the DNA binding domain at target sequence on genomic DNA (e.g., chromatin) results in one or more epigenetic modifications by the epigenetic modification domain that increases or decreases expression of the one or more polypeptides disclosed herein. As used herein, "linked to or otherwise capable of associating with" refers to a fusion protein or a recruitment domain or an adaptor protein, such as an aptamer (e.g., MS2) or an epitope tag. The recruitment domain or an adaptor protein can be linked to an epigenetic modification domain or the DNA binding domain (e.g., an adaptor for an aptamer). The epigenetic modification domain can be linked to an antibody specific for an epitope tag fused to the DNA binding domain. An aptamer can be linked to a guide sequence.

[0178] In example embodiments, the DNA binding domain is a programmable DNA binding protein linked to or otherwise capable of associating with an epigenetic modification domain. Programmable DNA binding proteins for modifying the epigenome include, but are not limited to CRISPR systems, transcription activator-like effectors (TALEs), Zn finger proteins and meganucleases (see, e.g., Thakore PI, Black JB, Hilton IB, Gersbach CA. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat Methods. 2016;13(2):127-137; and described further herein). In example embodiments, the DNA binding domain is a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient endonuclease enzyme. In example embodiments, a CRISPR system having an inactivated nuclease activity (e.g., dCas) is used as the DNA binding domain.

[0179] In example embodiments, the epigenetic modification domain is a functional domain and includes, but is not limited to, a histone methyltransferase (HMT) domain, histone demethylase domain, histone acetyltransferase (HAT) domain, histone deacetylation (HDAC) domain, DNA methyltransferase domain, DNA demethylation domain, histone phosphorylation domain (e.g., serine and threonine, or tyrosine), histone ubiquitylation domain, histone sumoylation domain, histone ADP ribosylation domain, histone proline isomerization domain, histone biotinylation domain, histone citrullination domain (see, e.g., Epigenetics, Second Edition, 2015, Edited by C. David Allis; Marie-Laure Caparros; Thomas Jenuwein; Danny Reinberg; Associate Editor Monika Lachlan; Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150(1):12-27; Syding LA, Nickl P, Kasparek P, Sedlacek R. CRISPR/Cas9 Epigenome Editing Potential for Rare Imprinting Diseases: A Review. Cells. 2020;9(4):993; and Zhang Y. Transcriptional regulation by histone ubiquitination and deubiquitination. Genes Dev. 2003;17(22):2733-2740). Example epigenetic modification domains can be obtained from, but are not limited to chromatin modifying enzymes, such as, DNA methyltransferases (e.g., DNMT1, DNMT3a and DNMT3b), TET1, TET2, thymine-DNA glycosylase (TDG), GCN5-related N-acetyltransferases family (GNAT), MYST family proteins (e.g., MOZ and MORF), and CBP/p300 family proteins (e.g., CBP, p300), Class I HDACs (e.g., HDAC 1-3 and HDAC8), Class II HDACs (e.g., HDAC 4-7 and HDAC 9-10), Class III HDACs (e.g., sirtuins), HDAC11, SET domain containing methyltransferases (e.g., SET7/9 (KMT7, NCBI Entrez Gene: 80854), KMT5A (SET8), MMSET, EZH2, and MLL family members), DOT1L, LSD1, Jumonji demethylases (e.g., KDM5A (JARID1A), KDM5C (JARID1C), and KDM6A (UTX)), kinases (e.g., Haspin, VRK1, PKCα, PKCβ, PIM1, IKKα, Rsk2, PKB/Akt, Aurora B, MSK1/2, JNK1, MLTKα, PRK1, Chk1, Dlk/ZIP, PKCδ, MST1, AMPK, JAK2, Abl, BMK1, CaMK, S6K1, SIK1), Ubp8, ubiquitin C-terminal hydrolases (UCH), the ubiquitin-specific processing proteases (UBP), and poly(ADP-ribose) polymerase 1 (PARP-1). See, also, US Patent US11001829B2 for additional domains.

[0180] In example embodiments, histone acetylation is targeted to a target sequence using a CRISPR system (see, e.g., Hilton IB, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol. 2015). In example embodiments, histone deacetylation is targeted to a target sequence (see, e.g., Cong et al., 2012; and Konermann S, et al. Optical control of mammalian endogenous transcription

and epigenetic states. Nature. 2013;500:472–476). In example embodiments, histone methylation is targeted to a target sequence (see, e.g., Snowden AW, Gregory PD, Case CC, Pabo CO. Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. Curr Biol. 2002;12:2159–2166; and Cano-Rodriguez D, Gjaltema RA, Jilderda LJ, et al. Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat Commun. 2016;7:12284). In example embodiments, histone demethylation is targeted to a target sequence (see, e.g., Kearns NA, Pham H, Tabak B, et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat Methods. 2015;12(5):401-403). In example embodiments, histone phosphorylation is targeted to a target sequence (see, e.g., Li J, Mahata B, Escobar M, et al. Programmable human histone phosphorylation and gene activation using a CRISPR/Cas9-based chromatin kinase. Nat Commun. 2021;12(1):896). In example embodiments, DNA methylation is targeted to a target sequence (see, e.g., Rivenbark AG, et al. Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics. 2012;7:350–360; Siddique AN, et al. Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. J Mol Biol. 2013;425:479–491; Bernstein DL, Le Lay JE, Ruano EG, Kaestner KH. TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J Clin Invest. 2015;125:1998–2006; Liu XS, Wu H, Ji X, et al. Editing DNA Methylation in the Mammalian Genome. Cell. 2016;167(1):233-247.e17; Stepper P, Kungulovski G, Jurkowska RZ, et al. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. Nucleic Acids Res. 2017;45(4):1703-1713; and Pflueger C., Tan D., Swain T., Nguyen T., Pflueger J., Nefzger C., Polo J.M., Ford E., Lister R. A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. Genome Res. 2018;28:1193–1206). In example embodiments, DNA demethylation is targeted to a target sequence using a CRISPR system (see, e.g., TET1, see Xu et al, Cell Discov. 2016 May 3;2: 16009; Choudhury et al, Oncotarget. 2016 Jul 19;7(29):46545-46556; and Kang JG, Park JS, Ko JH, Kim YS. Regulation of gene expression by altered promoter methylation using a CRISPR/Cas9-mediated epigenetic editing system. Sci Rep. 2019;9(1):11960). In example embodiments, DNA demethylation is targeted to a target sequence (see, e.g., TDG, see, Gregory DJ, Zhang Y, Kobzik L, Fedulov AV. Specific

transcriptional enhancement of inducible nitric oxide synthase by targeted promoter demethylation. Epigenetics. 2013;8:1205–1212).

[0181] Example epigenetic modification domains can be obtained from, but are not limited to transcription activators, such as, VP64 (see, e.g., Ji Q, et al. Engineered zinc-finger transcription factors activate OCT4 (POU5F1), SOX2, KLF4, c-MYC (MYC) and miR302/367. Nucleic Acids Res. 2014;42:6158–6167; Perez-Pinera P, et al. Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat Methods. 2013;10:239–242; Farzadfard F, Perli SD, Lu TK. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. ACS Synth Biol. 2013;2:604–613; Black JB, Adler AF, Wang HG, et al. Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. Cell Stem Cell. 2016;19(3):406-414; and Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013;10(10):977-979), p65 (see, e.g., Liu PQ, et al. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. J Biol Chem. 2001;276:11323–11334; and Konermann S, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015;517:583–588), HSF1, and RTA (see, e.g., Chavez A, et al. Highly efficient Cas9-mediated transcriptional programming. Nat Methods. 2015;12:326–328). Example epigenetic modification domains can be obtained from, but are not limited to transcription repressors, such as, KRAB (see, e.g., Beerli RR, Segal DJ, Dreier B, Barbas CF., 3rd Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci U S A. 1998;95:14628–14633; Cong L, Zhou R, Kuo YC, Cunniff M, Zhang F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun. 2012;3:968; Gilbert LA, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154:442–451; and Yeo NC, Chavez A, Lance-Byrne A, et al. An enhanced CRISPR repressor for targeted mammalian gene regulation. Nat Methods. 2018;15(8):611-616).

[0182] In example embodiments, the epigenetic modification domain linked to a DNA binding domain recruits an epigenetic modification protein to a target sequence. In example embodiments, a transcriptional activator recruits an epigenetic modification protein to a target

sequence. For example, VP64 can recruit DNA demethylation, increased H3K27ac and H3K4me. In example embodiments, a transcriptional repressor protein recruits an epigenetic modification protein to a target sequence. For example, KRAB can recruit increased H3K9me3 (see, e.g., Thakore PI, D'Ippolito AM, Song L, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat Methods. 2015;12(12):1143-1149). In an example embodiment, methyl-binding proteins linked to a DNA binding domain, such as MBD1, MBD2, MBD3, and MeCP2 recruits an epigenetic modification protein to a target sequence. In an example embodiment, Mi2/NuRD, Sin3A, or Co-REST recruit HDACs to a target sequence.

[0183] In example embodiments, the epigenetic modification domain can be a eukaryotic or prokaryotic (e.g., bacteria or Archaea) protein. In example embodiments, the eukaryotic protein can be a mammalian, insect, plant, or yeast protein and is not limited to human proteins (e.g., a yeast, insect, plant chromatin modifying protein, such as yeast HATs, HDACs, methyltransferases, etc.

[0184] In one aspect of the invention, is provided a fusion protein (epigenetic modification polypeptide) comprising from N-terminus to C-terminus, an epigenetic modification domain, an XTEN linker, and a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient endonuclease enzyme.

[0185] In aspects, the epigenetic modification polypeptide further comprises a transcriptional activator. In aspects, the transcriptional activator is VP64, p65, RTA, or a combination of two or more thereof. In another aspect, the epigenetic modification polypeptide further comprises one or more nuclear localization sequences. In embodiments, the epigenetic modification polypeptide comprises the nuclease-deficient RNA-guided DNA endonuclease enzyme. In embodiments, the fusion protein comprises the nuclease-deficient DNA endonuclease enzyme.

[0186] In some embodiments, the functional domains associated with the adaptor protein or the CRISPR enzyme is a transcriptional activation domain comprising VP64, p65, MyoD1, HSF1, RTA or SET7/9. Other references herein to activation (or activator) domains in respect of those associated with the adaptor protein(s) include any known transcriptional activation domain and specifically VP64, p65, MyoD1, HSF1, RTA or SET7/9 (see, e.g., US Patent, US11001829B2).

[0187] In certain embodiments, the present invention provides a fusion protein comprising from N-terminus to C-terminus, an RNA-binding sequence, an XTEN linker, and a transcriptional activator. In aspects, the transcriptional activator is VP64, p65, RTA, or a combination of two or more thereof. In aspects, the fusion protein further comprises a demethylation domain, a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient endonuclease enzyme, a nuclear localization sequence, or a combination of two or more thereof. In embodiments, the fusion protein comprises the nuclease-deficient RNA-guided DNA endonuclease enzyme. In embodiments, the fusion protein comprises the nuclease-deficient DNA endonuclease enzyme.

[0188] In certain embodiments, the present invention provides a method of activating a target nucleic acid sequence in a cell, the method comprising: (i) delivering a first polynucleotide encoding a epigenetic modification polypeptide described herein including embodiments thereof to a cell containing the silenced target nucleic acid; and (ii) delivering to the cell a second polynucleotide comprising: (a) a sgRNA or (b) a cr:tracrRNA; thereby reactivating the silenced target nucleic acid sequence in the cell. In aspects, the sgRNA comprises at least one MS2 stem loop. In aspects, the second polynucleotide comprises a transcriptional activator. In aspects, the second polynucleotide comprises two or more sgRNA.

Zinc Finger Nucleases

[0189] In some embodiments, the polynucleotide is modified using a Zinc Finger nuclease or system thereof. One type of programmable DNA-binding domain is provided by artificial zinc-finger (ZF) technology, which involves arrays of ZF modules to target new DNA-binding sites in the genome. Each finger module in a ZF array targets three DNA bases. A customized array of individual zinc finger domains is assembled into a ZF protein (ZFP).

[0190] ZFPs can comprise a functional domain. The first synthetic zinc finger nucleases (ZFNs) were developed by fusing a ZF protein to the catalytic domain of the Type IIS restriction enzyme FokI. (Kim, Y. G. et al., 1994, Chimeric restriction endonuclease, Proc. Natl. Acad. Sci. U.S.A. 91, 883−887; Kim, Y. G. et al., 1996, Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci. U.S.A. 93, 1156−1160). Increased cleavage specificity can be attained with decreased off target activity by use of paired ZFN heterodimers, each targeting different nucleotide sequences separated by a short spacer. (Doyon, Y. et al., 2011, Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat. Methods 8, 74−79). ZFPs can also be designed as

transcription activators and repressors and have been used to target many genes in a wide variety of organisms. Exemplary methods of genome editing using ZFNs can be found for example in U.S. Patent Nos. 6,534,261, 6,607,882, 6,746,838, 6,794,136, 6,824,978, 6,866,997, 6,933,113, 6,979,539, 7,013,219, 7,030,215, 7,220,719, 7,241,573, 7,241,574, 7,585,849, 7,595,376, 6,903,185, and 6,479,626, all of which are specifically incorporated by reference.

TALE Nucleases

[0191] In some embodiments, a TALE nuclease or TALE nuclease system can be used to modify a polynucleotide. In some embodiments, the methods provided herein use isolated, non-naturally occurring, recombinant or engineered DNA binding proteins that comprise TALE monomers or TALE monomers or half monomers as a part of their organizational structure that enable the targeting of nucleic acid sequences with improved efficiency and expanded specificity.

[0192] Naturally occurring TALEs or "wild type TALEs" are nucleic acid binding proteins secreted by numerous species of proteobacteria. TALE polypeptides contain a nucleic acid binding domain composed of tandem repeats of highly conserved monomer polypeptides that are predominantly 33, 34 or 35 amino acids in length and that differ from each other mainly in amino acid positions 12 and 13. In advantageous embodiments the nucleic acid is DNA. As used herein, the term "polypeptide monomers", "TALE monomers" or "monomers" will be used to refer to the highly conserved repetitive polypeptide sequences within the TALE nucleic acid binding domain and the term "repeat variable di-residues" or "RVD" will be used to refer to the highly variable amino acids at positions 12 and 13 of the polypeptide monomers. As provided throughout the disclosure, the amino acid residues of the RVD are depicted using the IUPAC single letter code for amino acids. A general representation of a TALE monomer which is comprised within the DNA binding domain is X_{1-11}-($X_{12}X_{13}$)-X_{14-33} or $_{34}$ or $_{35}$, where the subscript indicates the amino acid position and X represents any amino acid. $X_{12}X_{13}$ indicate the RVDs. In some polypeptide monomers, the variable amino acid at position 13 is missing or absent and in such monomers, the RVD consists of a single amino acid. In such cases the RVD may be alternatively represented as X*, where X represents X_{12} and (*) indicates that X_{13} is absent. The DNA binding domain comprises several repeats of TALE monomers and this may be represented as (X_{1-11}-($X_{12}X_{13}$)-X_{14-33} or $_{34}$ or $_{35}$)$_z$, where in an advantageous embodiment, z is at least 5 to 40. In a further advantageous embodiment, z is at least 10 to 26.

[0193] The TALE monomers can have a nucleotide binding affinity that is determined by the identity of the amino acids in its RVD. For example, polypeptide monomers with an RVD of NI can preferentially bind to adenine (A), monomers with an RVD of NG can preferentially bind to thymine (T), monomers with an RVD of HD can preferentially bind to cytosine (C) and monomers with an RVD of NN can preferentially bind to both adenine (A) and guanine (G). In some embodiments, monomers with an RVD of IG can preferentially bind to T. Thus, the number and order of the polypeptide monomer repeats in the nucleic acid binding domain of a TALE determines its nucleic acid target specificity. In some embodiments, monomers with an RVD of NS can recognize all four base pairs and can bind to A, T, G or C. The structure and function of TALEs is further described in, for example, Moscou et al., Science 326:1501 (2009); Boch et al., Science 326:1509-1512 (2009); and Zhang et al., Nature Biotechnology 29:149-153 (2011).

[0194] The polypeptides used in methods of the invention can be isolated, non-naturally occurring, recombinant or engineered nucleic acid-binding proteins that have nucleic acid or DNA binding regions containing polypeptide monomer repeats that are designed to target specific nucleic acid sequences.

[0195] As described herein, polypeptide monomers having an RVD of HN or NH preferentially bind to guanine and thereby allow the generation of TALE polypeptides with high binding specificity for guanine containing target nucleic acid sequences. In some embodiments, polypeptide monomers having RVDs RN, NN, NK, SN, NH, KN, HN, NQ, HH, RG, KH, RH and SS can preferentially bind to guanine. In some embodiments, polypeptide monomers having RVDs RN, NK, NQ, HH, KH, RH, SS and SN can preferentially bind to guanine and can thus allow the generation of TALE polypeptides with high binding specificity for guanine containing target nucleic acid sequences. In some embodiments, polypeptide monomers having RVDs HH, KH, NH, NK, NQ, RH, RN and SS can preferentially bind to guanine and thereby allow the generation of TALE polypeptides with high binding specificity for guanine containing target nucleic acid sequences. In some embodiments, the RVDs that have high binding specificity for guanine are RN, NH RH and KH. Furthermore, polypeptide monomers having an RVD of NV can preferentially bind to adenine and guanine. In some embodiments, monomers having RVDs of H*, HA, KA, N*, NA, NC, NS, RA, and S* bind to adenine, guanine, cytosine and thymine with comparable affinity.

[0196] The predetermined N-terminal to C-terminal order of the one or more polypeptide monomers of the nucleic acid or DNA binding domain determines the corresponding predetermined target nucleic acid sequence to which the polypeptides of the invention will bind. As used herein the monomers and at least one or more half monomers are "specifically ordered to target" the genomic locus or gene of interest. In plant genomes, the natural TALE-binding sites always begin with a thymine (T), which may be specified by a cryptic signal within the non-repetitive N-terminus of the TALE polypeptide; in some cases, this region may be referred to as repeat 0. In animal genomes, TALE binding sites do not necessarily have to begin with a thymine (T) and polypeptides of the invention may target DNA sequences that begin with T, A, G or C. The tandem repeat of TALE monomers always ends with a half-length repeat or a stretch of sequence that may share identity with only the first 20 amino acids of a repetitive full-length TALE monomer and this half repeat may be referred to as a half-monomer. Therefore, it follows that the length of the nucleic acid or DNA being targeted is equal to the number of full monomers plus two.

[0197] As described in Zhang et al., Nature Biotechnology 29:149-153 (2011), TALE polypeptide binding efficiency may be increased by including amino acid sequences from the "capping regions" that are directly N-terminal or C-terminal of the DNA binding region of naturally occurring TALEs into the engineered TALEs at positions N-terminal or C-terminal of the engineered TALE DNA binding region. Thus, in certain embodiments, the TALE polypeptides described herein further comprise an N-terminal capping region and/or a C-terminal capping region.

[0198] An exemplary amino acid sequence of a N-terminal capping region is:

[0199] M D P I R S R T P S P A R E L L S G P Q P D G V Q P T A D R G V S P P A G G P L D G L P A R R T M S R T R L P S P P A P S P A F S A D S F S D L L R Q F D P S L F N T S L F D S L P P F G A H H T E A A T G E W D E V Q S G L R A A D A P P P T M R V A V T A A R P P R A K P A P R R R A A Q P S D A S P A A Q V D L R T L G Y S Q Q Q Q E K I K P K V R S T V A Q H H E A L V G H G F T H A H I V A L S Q H P A A L G T V A V K Y Q D M I A A L P E A T H E A I V G V G K Q W S G A R A L E A L L T V A G E L R G P P L Q L D T G Q L L K I A K R G G V T A V E A V H A W R N A L T G A P L N (SEQ ID NO:18)

[0200] An exemplary amino acid sequence of a C-terminal capping region is:

[0201] R P A L E S I V A Q L S R P D P A L A A L T N D H L V A L A C L G G R P A L D A V K K G L P H A P A L I K R T N R R I P E R T S H R V A D H A Q V V R V L G F F Q C H S H P A Q A F D D A M T Q F G M S R H G L L Q L F R R V G V T E L E A R S G T L P P A S Q R W D R I L Q A S G M K R A K P S P T S T Q T P D Q A S L H A F A D S L E R D L D A P S P M H E G D Q T R A S (SEQ ID NO:19)

[0202] As used herein the predetermined "N-terminus" to "C terminus" orientation of the N-terminal capping region, the DNA binding domain comprising the repeat TALE monomers and the C-terminal capping region provide structural basis for the organization of different domains in the d-TALEs or polypeptides of the invention.

[0203] The entire N-terminal and/or C-terminal capping regions are not necessary to enhance the binding activity of the DNA binding region. Therefore, in certain embodiments, fragments of the N-terminal and/or C-terminal capping regions are included in the TALE polypeptides described herein.

[0204] In certain embodiments, the TALE polypeptides described herein contain a N-terminal capping region fragment that included at least 10, 20, 30, 40, 50, 54, 60, 70, 80, 87, 90, 94, 100, 102, 110, 117, 120, 130, 140, 147, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260 or 270 amino acids of an N-terminal capping region. In certain embodiments, the N-terminal capping region fragment amino acids are of the C-terminus (the DNA-binding region proximal end) of an N-terminal capping region. As described in Zhang et al., *Nature Biotechnology* 29:149-153 (2011), N-terminal capping region fragments that include the C-terminal 240 amino acids enhance binding activity equal to the full length capping region, while fragments that include the C-terminal 147 amino acids retain greater than 80% of the efficacy of the full length capping region, and fragments that include the C-terminal 117 amino acids retain greater than 50% of the activity of the full-length capping region.

[0205] In some embodiments, the TALE polypeptides described herein contain a C-terminal capping region fragment that included at least 6, 10, 20, 30, 37, 40, 50, 60, 68, 70, 80, 90, 100, 110, 120, 127, 130, 140, 150, 155, 160, 170, or 180 amino acids of a C-terminal capping region. In certain embodiments, the C-terminal capping region fragment amino acids are of the N-terminus (the DNA-binding region proximal end) of a C-terminal capping region. As described in Zhang et al., Nature Biotechnology 29:149-153 (2011), C-terminal capping region fragments that include the C-terminal 68 amino acids enhance binding activity equal to

the full-length capping region, while fragments that include the C-terminal 20 amino acids retain greater than 50% of the efficacy of the full-length capping region.

[0206] In certain embodiments, the capping regions of the TALE polypeptides described herein do not need to have identical sequences to the capping region sequences provided herein. Thus, in some embodiments, the capping region of the TALE polypeptides described herein have sequences that are at least 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical or share identity to the capping region amino acid sequences provided herein. Sequence identity is related to sequence homology. Homology comparisons may be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs may calculate percent (%) homology between two or more sequences and may also calculate the sequence identity shared by two or more amino acid or nucleic acid sequences. In some preferred embodiments, the capping region of the TALE polypeptides described herein have sequences that are at least 95% identical or share identity to the capping region amino acid sequences provided herein.

[0207] Sequence homologies can be generated by any of a number of computer programs known in the art, which include but are not limited to BLAST or FASTA. Suitable computer programs for carrying out alignments like the GCG Wisconsin Bestfit package may also be used. Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

[0208] In some embodiments described herein, the TALE polypeptides of the invention include a nucleic acid binding domain linked to the one or more effector domains. The terms "effector domain" or "regulatory and functional domain" refer to a polypeptide sequence that has an activity other than binding to the nucleic acid sequence recognized by the nucleic acid binding domain. By combining a nucleic acid binding domain with one or more effector domains, the polypeptides of the invention may be used to target the one or more functions or activities mediated by the effector domain to a particular target DNA sequence to which the nucleic acid binding domain specifically binds.

[0209] In some embodiments of the TALE polypeptides described herein, the activity mediated by the effector domain is a biological activity. For example, in some embodiments the effector domain is a transcriptional inhibitor (i.e., a repressor domain), such as an mSin

interaction domain (SID). SID4X domain or a Krüppel-associated box (KRAB) or fragments of the KRAB domain. In some embodiments, the effector domain is an enhancer of transcription (i.e., an activation domain), such as the VP16, VP64 or p65 activation domain. In some embodiments, the nucleic acid binding is linked, for example, with an effector domain that includes but is not limited to a transposase, integrase, recombinase, resolvase, invertase, protease, DNA methyltransferase, DNA demethylase, histone acetylase, histone deacetylase, nuclease, transcriptional repressor, transcriptional activator, transcription factor recruiting, protein nuclear-localization signal or cellular uptake signal.

[0210] In some embodiments, the effector domain is a protein domain which exhibits activities which include but are not limited to transposase activity, integrase activity, recombinase activity, resolvase activity, invertase activity, protease activity, DNA methyltransferase activity, DNA demethylase activity, histone acetylase activity, histone deacetylase activity, nuclease activity, nuclear-localization signaling activity, transcriptional repressor activity, transcriptional activator activity, transcription factor recruiting activity, or cellular uptake signaling activity. Other preferred embodiments of the invention may include any combination of the activities described herein.

Meganucleases

[0211] In some embodiments, a meganuclease or system thereof can be used to modify a polynucleotide. Meganucleases, which are endodeoxyribonucleases characterized by a large recognition site (double-stranded DNA sequences of 12 to 40 base pairs). Exemplary methods for using meganucleases can be found in US Patent Nos. 8,163,514, 8,133,697, 8,021,867, 8,119,361, 8,119,381, 8,124,369, and 8,129,134, which are specifically incorporated herein by reference.

ARCUS Based Editing

[0212] In one example embodiment, a target gene is modified with an ARCUS base editing system. Exemplary methods for using ARCUS can be found in US Patent No. 10,851,358, US Publication No. 2020-0239544, and WIPO Publication No. 2020/206231 which are incorporated herein by reference.

RNAi and antisense oligonucleotides (ASO)

[0213] In example embodiments, IP6Ks, PPIP5Ks, and/or FGF23 are targeted with RNAi or antisense oligonucleotides (ASO). As used herein, “gene silencing” or “gene silenced” in reference to an activity of an RNAi molecule, for example a siRNA or miRNA refers to a

decrease in the mRNA level in a cell for a target gene by at least about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 95%, about 99%, about 100% of the mRNA level found in the cell without the presence of the miRNA or RNA interference molecule. In one preferred embodiment, the mRNA levels are decreased by at least about 70%, about 80%, about 90%, about 95%, about 99%, about 100%. Additionally, inhibitory nucleic acid molecules such as RNAi and ASOs can be used *in vivo* (see, e.g., Yan Y, Liu XY, Lu A, Wang XY, Jiang LX, Wang JC. Non-viral vectors for RNA delivery. J Control Release. 2022;342:241-279).

[0214] As used herein, the term "RNAi" refers to any type of interfering RNA, including but not limited to, siRNAi, shRNAi, endogenous microRNA and artificial microRNA. For instance, it includes sequences previously identified as siRNA, regardless of the mechanism of down-stream processing of the RNA (i.e., although siRNAs are believed to have a specific method of in vivo processing resulting in the cleavage of mRNA, such sequences can be incorporated into the vectors in the context of the flanking sequences described herein). The term "RNAi" can include both gene silencing RNAi molecules, and also RNAi effector molecules which activate the expression of a gene.

[0215] As used herein, a "siRNA" refers to a nucleic acid that forms a double stranded RNA, which double stranded RNA has the ability to reduce or inhibit expression of a gene or target gene when the siRNA is present or expressed in the same cell as the target gene. The double stranded RNA siRNA can be formed by the complementary strands. In one embodiment, a siRNA refers to a nucleic acid that can form a double stranded siRNA. The sequence of the siRNA can correspond to the full-length target gene, or a subsequence thereof. Typically, the siRNA is at least about 15-50 nucleotides in length (e.g., each complementary sequence of the double stranded siRNA is about 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs in length, preferably about 19-30 base nucleotides, preferably about 20-25 nucleotides in length, e.g., 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length).

[0216] As used herein "shRNA" or "small hairpin RNA" (also called stem loop) is a type of siRNA. In one embodiment, these shRNAs are composed of a short, e.g., about 19 to about 25 nucleotide, antisense strand, followed by a nucleotide loop of about 5 to about 9 nucleotides, and the analogous sense strand. Alternatively, the sense strand can precede the nucleotide loop structure and the antisense strand can follow.

[0217] The terms "microRNA" or "miRNA" are used interchangeably herein are endogenous RNAs, some of which are known to regulate the expression of protein-coding genes at the posttranscriptional level. Endogenous microRNAs are small RNAs naturally present in the genome that are capable of modulating the productive utilization of mRNA. The term artificial microRNA includes any type of RNA sequence, other than endogenous microRNA, which is capable of modulating the productive utilization of mRNA. MicroRNA sequences have been described in publications such as Lim, et al., Genes & Development, 17, p. 991 - 1008 (2003), Lim et al Science 299, 1540 (2003), Lee and Ambros Science, 294, 862 (2001), Lau et al., Science 294, 858-861 (2001), Lagos-Quintana et al, Current Biology, 12, 735-739 (2002), Lagos Quintana et al, Science 294, 853- 857 (2001), and Lagos-Quintana et al, RNA, 9, 175- 179 (2003), which are incorporated herein by reference. Multiple microRNAs can also be incorporated into a precursor molecule. Furthermore, miRNA-like stem-loops can be expressed in cells as a vehicle to deliver artificial miRNAs and short interfering RNAs (siRNAs) for the purpose of modulating the expression of endogenous genes through the miRNA and or RNAi pathways.

[0218] As used herein, "double stranded RNA" or "dsRNA" refers to RNA molecules that are comprised of two strands. Double-stranded molecules include those comprised of a single RNA molecule that doubles back on itself to form a two-stranded structure. For example, the stem loop structure of the progenitor molecules from which the single-stranded miRNA is derived, called the pre-miRNA (Bartel et al. 2004. Cell 1 16:281 -297), comprises a dsRNA molecule.

[0219] Antisense therapy is a form of treatment that uses antisense oligonucleotides (ASOs) to target messenger RNA (mRNA). ASOs are capable of altering mRNA expression through a variety of mechanisms, including ribonuclease H mediated decay of the pre-mRNA, direct steric blockage, and exon content modulation through splicing site binding on pre-mRNA (see, e.g., Crooke ST, Liang XH, Baker BF, Crooke RM. Antisense technology: A review. J Biol Chem. 2021;296:100416. doi:10.1016/j.jbc.2021.100416). Antisense oligonucleotides (ASO) generally inhibit their target by binding target mRNA and sterically blocking expression by obstructing the ribosome. ASOs can also inhibit their target by binding target mRNA thus forming a DNA-RNA hybrid that can be a substance for RNase H. Commonly used antisense mechanisms to degrade target RNAs include RNase H1-dependent

and RISC-dependent mechanisms. Preferred ASOs include Locked Nucleic Acid (LNA), Peptide Nucleic Acid (PNA), and morpholinos.

Administration of Therapeutic Agents

[0220] In one example, therapeutic agents are administered in a combination with other agents, e.g., therapeutic agents, that are useful for treating pathological conditions or disorders, such as various forms of cancer. The term "in combination" in this context means that the agents are given substantially contemporaneously, either simultaneously or sequentially. If given sequentially, at the onset of administration of the second agent, the first of the two agents is in some cases still detectable at effective concentrations at the site of treatment.

[0221] The administration of an agent or a combination of agents for the treatment of a neoplasia may be by any suitable means that results in a concentration of the therapeutic that, combined with other components, is effective in ameliorating, reducing, or stabilizing a neoplasia. The agent may be contained in any appropriate amount in any suitable carrier substance, and is generally present in an amount of 1-95% by weight of the total weight of the composition. The composition may be provided in a dosage form that is suitable for parenteral (e.g., subcutaneously, intravenously, intramuscularly, or intraperitoneally) administration route. The pharmaceutical compositions may be formulated according to conventional pharmaceutical practice (see, e.g., Remington: The Science and Practice of Pharmacy (20^{th} ed.), ed. A. R. Gennaro, Lippincott Williams & Wilkins, 2000 and Encyclopedia of Pharmaceutical Technology, eds. J. Swarbrick and J. C. Boylan, 1988-1999, Marcel Dekker, New York).

[0222] In some examples, the prophylactic and/or therapeutic regimen comprises administration of an agent of the invention in combination with one or more additional anticancer therapeutics. In one example, the dosages of the one or more additional anticancer therapeutics used in the combination therapy is lower than those which have been or are currently being used to prevent, treat, and/or manage cancer. The recommended dosages of the one or more additional anticancer therapeutics currently used for the prevention, treatment, and/or management of cancer can be obtained from any reference in the art including, but not limited to, Hardman et al., eds., Goodman & Gilman's The Pharmacological Basis of Therapeutics, 10^{th} ed., McGraw-Hill, New York, 2001; Physician's Desk Reference (60^{th} ed., 2006), which is incorporated herein by reference in its entirety.

[0223] The agent of the invention and the one or more additional anticancer therapeutics can be administered separately, simultaneously, or sequentially. In various aspects, the agent of the invention and the additional anticancer therapeutic are administered less than 5 minutes apart, less than 30 minutes apart, less than 1 hour apart, at about 1 hour apart, at about 1 to about 2 hours apart, at about 2 hours to about 3 hours apart, at about 3 hours to about 4 hours apart, at about 4 hours to about 5 hours apart, at about 5 hours to about 6 hours apart, at about 6 hours to about 7 hours apart, at about 7 hours to about 8 hours apart, at about 8 hours to about 9 hours apart, at about 9 hours to about 10 hours apart, at about 10 hours to about 11 hours apart, at about 11 hours to about 12 hours apart, at about 12 hours to 18 hours apart, 18 hours to 24 hours apart, 24 hours to 36 hours apart, 36 hours to 48 hours apart, 48 hours to 52 hours apart, 52 hours to 60 hours apart, 60 hours to 72 hours apart, 72 hours to 84 hours apart, 84 hours to 96 hours apart, or 96 hours to 120 hours part. In another example, two or more anticancer therapeutics are administered within the same patient visit.

[0224] In certain aspects, the agent of the invention and the additional anticancer therapeutic are cyclically administered. Cycling therapy involves the administration of one anticancer therapeutic for a period of time, followed by the administration of a second anticancer therapeutic for a period of time and repeating this sequential administration, i.e., the cycle, in order to reduce the development of resistance to one or both of the anticancer therapeutics, to avoid or reduce the side effects of one or both of the anticancer therapeutics, and/or to improve the efficacy of the therapies. In one example, cycling therapy involves the administration of a first anticancer therapeutic for a period of time, followed by the administration of a second anticancer therapeutic for a period of time, optionally, followed by the administration of a third anticancer therapeutic for a period of time and so forth, and repeating this sequential administration, i.e., the cycle in order to reduce the development of resistance to one of the anticancer therapeutics, to avoid or reduce the side effects of one of the anticancer therapeutics, and/or to improve the efficacy of the anticancer therapeutics.

[0225] In another example, the anticancer therapeutics are administered concurrently to a subject in separate compositions. The combination anticancer therapeutics of the invention may be administered to a subject by the same or different routes of administration.

[0226] When an agent of the invention and the additional anticancer therapeutic are administered to a subject concurrently, the term "concurrently" is not limited to the administration of the anticancer therapeutics at exactly the same time, but rather, it is meant

that they are administered to a subject in a sequence and within a time interval such that they can act together (e.g., synergistically to provide an increased benefit than if they were administered otherwise). For example, the anticancer therapeutics may be administered at the same time or sequentially in any order at different points in time; however, if not administered at the same time, they should be administered sufficiently close in time so as to provide the desired therapeutic effect, preferably in a synergistic fashion. The combination anticancer therapeutics of the invention can be administered separately, in any appropriate form and by any suitable route. When the components of the combination anticancer therapeutics are not administered in the same pharmaceutical composition, it is understood that they can be administered in any order to a subject in need thereof. For example, an agent of the invention can be administered prior to (e.g., 5 minutes, 15 minutes, 30 minutes, 45 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours, 24 hours, 48 hours, 72 hours, 96 hours, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 8 weeks, or 12 weeks before), concomitantly with, or subsequent to (e.g., 5 minutes, 15 minutes, 30 minutes, 45 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours, 24 hours, 48 hours, 72 hours, 96 hours, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 8 weeks, or 12 weeks after) the administration of the additional anticancer therapeutic, to a subject in need thereof. In various aspects, the anticancer therapeutics are administered 1 minute apart, 10 minutes apart, 30 minutes apart, less than 1 hour apart, 1 hour apart, 1 hour to 2 hours apart, 2 hours to 3 hours apart, 3 hours to 4 hours apart, 4 hours to 5 hours apart, 5 hours to 6 hours apart, 6 hours to 7 hours apart, 7 hours to 8 hours apart, 8 hours to 9 hours apart, 9 hours to 10 hours apart, 10 hours to 11 hours apart, 11 hours to 12 hours apart, no more than 24 hours apart or no more than 48 hours apart. In one example, the anticancer therapeutics are administered within the same office visit. In another example, the combination anticancer therapeutics of the invention are administered at 1 minute to 24 hours apart.

[0227] For therapeutic uses, the compositions or agents described herein may be administered systemically, for example, formulated in a pharmaceutically-acceptable buffer such as physiological saline. Preferable routes of administration include, for example, subcutaneous, intravenous, interperitoneal, intramuscular, or intradermal injections that provide continuous, sustained levels of the drug in the patient. Treatment of human patients or other animals will be carried out using a therapeutically effective amount of a therapeutic identified herein in a physiologically-acceptable carrier. Suitable carriers and their formulation

are described, for example, in Remington's Pharmaceutical Sciences by E. W. Martin. The amount of the therapeutic agent to be administered varies depending upon the manner of administration, the age and body weight of the patient, and with the clinical symptoms of the neoplasia. Generally, amounts will be in the range of those used for other agents used in the treatment of other diseases associated with neoplasia, although in certain instances lower amounts will be needed because of the increased specificity of the compound. For example, a therapeutic compound is administered at a dosage that is cytotoxic to a neoplastic cell.

[0228] Human dosage amounts can initially be determined by extrapolating from the amount of compound used in mice, as a skilled artisan recognizes it is routine in the art to modify the dosage for humans compared to animal models. In example embodiments, it is envisioned that the dosage may vary from between about 1 μg compound/Kg body weight to about 5000 mg compound/Kg body weight; or from about 5 mg/Kg body weight to about 4000 mg/Kg body weight or from about 10 mg/Kg body weight to about 3000 mg/Kg body weight; or from about 50 mg/Kg body weight to about 2000 mg/Kg body weight; or from about 100 mg/Kg body weight to about 1000 mg/Kg body weight; or from about 150 mg/Kg body weight to about 500 mg/Kg body weight. In other cases, this dose may be about 1, 5, 10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, 2000, 2500, 3000, 3500, 4000, 4500, or 5000 mg/Kg body weight. In other aspects, it is envisaged that doses may be in the range of about 5 mg compound/Kg body to about 20 mg compound/Kg body. In other embodiments, the doses may be about 8, 10, 12, 14, 16 or 18 mg/Kg body weight. Of course, this dosage amount may be adjusted upward or downward, as is routinely done in such treatment protocols, depending on the results of the initial clinical trials and the needs of a particular patient.

[0229] In some cases, the compound or composition of the invention is administered at a dose that is lower than the human equivalent dosage (HED) of the no observed adverse effect level (NOAEL) over a period of three months, four months, six months, nine months, 1 year, 2 years, 3 years, 4 years or more. The NOAEL, as determined in animal studies, is useful in determining the maximum recommended starting dose for human clinical trials. For instance, the NOAELs can be extrapolated to determine human equivalent dosages. Typically, such extrapolations between species are conducted based on the doses that are normalized to body surface area (i.e., mg/m^2). In specific embodiments, the NOAELs are determined in mice,

hamsters, rats, ferrets, guinea pigs, rabbits, dogs, primates, primates (monkeys, marmosets, squirrel monkeys, baboons), micropigs or minipigs. For a discussion on the use of NOAELs and their extrapolation to determine human equivalent doses, see Guidance for Industry Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers, U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER), Pharmacology and Toxicology, July 2005, incorporated herein by reference.

[0230] The amount of an agent of the invention used in the prophylactic and/or therapeutic regimens which will be effective in the prevention, treatment, and/or management of cancer can be based on the currently prescribed dosage of the agent as well as assessed by methods disclosed herein and known in the art. The frequency and dosage will vary also according to factors specific for each patient depending on the specific compounds administered, the severity of the cancerous condition, the route of administration, as well as age, body, weight, response, and the past medical history of the patient. For example, the dosage of an agent of the invention which will be effective in the treatment, prevention, and/or management of cancer can be determined by administering the compound to an animal model such as, e.g., the animal models disclosed herein or known to those skilled in the art. In addition, in vitro assays may optionally be employed to help identify optimal dosage ranges.

[0231] In some aspects, the prophylactic and/or therapeutic regimens comprise titrating the dosages administered to the patient so as to achieve a specified measure of therapeutic efficacy. Such measures include a reduction in the cancer cell population in the patient.

[0232] In certain cases, the dosage of the compound of the invention in the prophylactic and/or therapeutic regimen is adjusted so as to achieve a reduction in the number or amount of cancer cells found in a test specimen extracted from a patient after undergoing the prophylactic and/or therapeutic regimen, as compared with a reference sample. Here, the reference sample is a specimen extracted from the patient undergoing therapy, wherein the specimen is extracted from the patient at an earlier time point. In one aspect, the reference sample is a specimen extracted from the same patient, prior to receiving the prophylactic and/or therapeutic regimen. For example, the number or amount of cancer cells in the test specimen is at least 2%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or 99% lower than in the reference sample.

[0233] In some cases, the dosage of the compound of the invention in the prophylactic and/or therapeutic regimen is adjusted so as to achieve a number or amount of cancer cells that falls within a predetermined reference range. In these embodiments, the number or amount of cancer cells in a test specimen is compared with a predetermined reference range.

[0234] In other embodiments, the dosage of the compound of the invention in prophylactic and/or therapeutic regimen is adjusted so as to achieve a reduction in the number or amount of cancer cells found in a test specimen extracted from a patient after undergoing the prophylactic and/or therapeutic regimen, as compared with a reference sample, wherein the reference sample is a specimen is extracted from a healthy, noncancer-afflicted patient. For example, the number or amount of cancer cells in the test specimen is at least within 60%, 50%, 40%, 30%, 20%, 15%, 10%, 5%, or 2% of the number or amount of cancer cells in the reference sample.

[0235] In treating certain human patients having solid tumors, extracting multiple tissue specimens from a suspected tumor site may prove impracticable. In these cases, the dosage of the compounds of the invention in the prophylactic and/or therapeutic regimen for a human patient is extrapolated from doses in animal models that are effective to reduce the cancer population in those animal models. In the animal models, the prophylactic and/or therapeutic regimens are adjusted so as to achieve a reduction in the number or amount of cancer cells found in a test specimen extracted from an animal after undergoing the prophylactic and/or therapeutic regimen, as compared with a reference sample. The reference sample can be a specimen extracted from the same animal, prior to receiving the prophylactic and/or therapeutic regimen. In specific embodiments, the number or amount of cancer cells in the test specimen is at least 2%, 5%, 10%, 15%, 20%, 30%, 40%, 50% or 60% lower than in the reference sample. The doses effective in reducing the number or amount of cancer cells in the animals can be normalized to body surface area (e.g., mg/m^2) to provide an equivalent human dose.

[0236] The prophylactic and/or therapeutic regimens disclosed herein comprise administration of compounds of the invention or pharmaceutical compositions thereof to the patient in a single dose or in multiple doses (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 10, 15, 20, or more doses).

[0237] In one aspect, the prophylactic and/or therapeutic regimens comprise administration of the compounds of the invention or pharmaceutical compositions thereof in multiple doses. When administered in multiple doses, the compounds or pharmaceutical compositions are administered with a frequency and in an amount sufficient to prevent, treat, and/or manage the condition. For example, the frequency of administration ranges from once a day up to about

once every eight weeks. In another example, the frequency of administration ranges from about once a week up to about once every six weeks. In another example, the frequency of administration ranges from about once every three weeks up to about once every four weeks.

[0238] Generally, the dosage of a compound of the invention administered to a subject to prevent, treat, and/or manage cancer is in the range of 0.01 to 500 mg/kg, e.g., in the range of 0.1 mg/kg to 100 mg/kg of the subject's body weight. For example, the dosage administered to a subject is in the range of 0.1 mg/kg to 50 mg/kg, or 1 mg/kg to 50 mg/kg, of the subject's body weight, more preferably in the range of 0.1 mg/kg to 25 mg/kg, or 1 mg/kg to 25 mg/kg, of the patient's body weight. In another example, the dosage of a compound of the invention administered to a subject to prevent, treat, and/or manage cancer in a patient is 500 mg/kg or less, preferably 250 mg/kg or less, 100 mg/kg or less, 95 mg/kg or less, 90 mg/kg or less, 85 mg/kg or less, 80 mg/kg or less, 75 mg/kg or less, 70 mg/kg or less, 65 mg/kg or less, 60 mg/kg or less, 55 mg/kg or less, 50 mg/kg or less, 45 mg/kg or less, 40 mg/kg or less, 35 mg/kg or less, 30 mg/kg or less, 25 mg/kg or less, 20 mg/kg or less, 15 mg/kg or less, 10 mg/kg or less, 5 mg/kg or less, 2.5 mg/kg or less, 2 mg/kg or less, 1.5 mg/kg or less, or 1 mg/kg or less of a patient's body weight.

[0239] In another example, the dosage of a compound of the invention administered to a subject to prevent, treat, and/or manage cancer in a patient is a unit dose of 0.1 to 50 mg, 0.1 mg to 20 mg, 0.1 mg to 15 mg, 0.1 mg to 12 mg, 0.1 mg to 10 mg, 0.1 mg to 8 mg, 0.1 mg to 7 mg, 0.1 mg to 5 mg, 0.1 to 2.5 mg, 0.25 mg to 20 mg, 0.25 to 15 mg, 0.25 to 12 mg, 0.25 to 10 mg, 0.25 to 8 mg, 0.25 mg to 7 mg, 0.25 mg to 5 mg, 0.5 mg to 2.5 mg, 1 mg to 20 mg, 1 mg to 15 mg, 1 mg to 12 mg, 1 mg to 10 mg, 1 mg to 8 mg, 1 mg to 7 mg, 1 mg to 5 mg, or 1 mg to 2.5 mg.

[0240] In another example, the dosage of a compound of the invention administered to a subject to prevent, treat, and/or manage cancer in a patient is in the range of 0.01 to 10 g/m^2, and more typically, in the range of 0.1 g/m^2 to 7.5 g/m^2, of the subject's body weight. For example, the dosage administered to a subject is in the range of 0.5 g/m^2 to 5 g/m^2, or 1 g/m^2 to 5 g/m^2 of the subject's body's surface area.

[0241] In another example, the prophylactic and/or therapeutic regimen comprises administering to a patient one or more doses of an effective amount of a compound of the invention, wherein the dose of an effective amount achieves a plasma level of at least 0.1 μg/mL, at least 0.5 μg/mL, at least 1 μg/mL, at least 2 μg/mL, at least 5 μg/mL, at least 6

μg/mL, at least 10 μg/mL, at least 15 μg/mL, at least 20 μg/mL, at least 25 μg/mL, at least 50 μg/mL, at least 100 μg/mL, at least 125 μg/mL, at least 150 μg/mL, at least 175 μg/mL, at least 200 μg/mL, at least 225 μg/mL, at least 250 μg/mL, at least 275 μg/mL, at least 300 μg/mL, at least 325 μg/mL, at least 350 μg/mL, at least 375 μg/mL, or at least 400 μg/mL of the compound of the invention.

[0242] In another example, the prophylactic and/or therapeutic regimen comprises administering to a patient a plurality of doses of an effective amount of a compound of the invention, wherein the plurality of doses maintains a plasma level of at least 0.1 μg/mL, at least 0.5 μg/mL, at least 1 μg/mL, at least 2 μg/mL, at least 5 μg/mL, at least 6 μg/mL, at least 10 μg/mL, at least 15 μg/mL, at least 20 μg/mL, at least 25 μg/mL, at least 50 μg/mL, at least 100 μg/mL, at least 125 μg/mL, at least 150 μg/mL, at least 175 μg/mL, at least 200 μg/mL, at least 225 μg/mL, at least 250 μg/mL, at least 275 μg/mL, at least 300 μg/mL, at least 325 μg/mL, at least 350 μg/mL, at least 375 μg/mL, or at least 400 μg/mL of the compound of the invention for at least 1 day, 1 month, 2 months, 3 months, 4 months, 5 months, 6 months, 7 months, 8 months, 9 months, 10 months, 11 months, 12 months, 15 months, 18 months, 24 months or 36 months.

[0243] In other embodiments, the prophylactic and/or therapeutic regimen comprises administering to a patient a plurality of doses of an effective amount of a compound of the invention, wherein the plurality of doses maintains a plasma level of at least 0.1 μg/mL, at least 0.5 μg/mL, at least 1 μg/mL, at least 2 μg/mL, at least 5 μg/mL, at least 6 μg/mL, at least 10 μg/mL, at least 15 μg/mL, at least 20 μg/mL, at least 25 μg/mL, at least 50 μg/mL, at least 100 μg/mL, at least 125 μg/mL, at least 150 μg/mL, at least 175 μg/mL, at least 200 μg/mL, at least 225 μg/mL, at least 250 μg/mL, at least 275 μg/mL, at least 300 μg/mL, at least 325 μg/mL, at least 350 μg/mL, at least 375 μg/mL, or at least 400 μg/mL of the compound of the invention for at least 1 day, 1 month, 2 months, 3 months, 4 months, 5 months, 6 months, 7 months, 8 months, 9 months, 10 months, 11 months, 12 months, 15 months, 18 months, 24 months or 36 months.

Release of pharmaceutical compositions

[0244] Pharmaceutical compositions according to the invention may be formulated to release the active compound substantially immediately upon administration or at any predetermined time or time period after administration. The latter types of compositions are generally known as controlled release formulations, which include (i) formulations that create

a substantially constant concentration of the drug within the body over an extended period of time; (ii) formulations that after a predetermined lag time create a substantially constant concentration of the drug within the body over an extended period of time; (iii) formulations that sustain action during a predetermined time period by maintaining a relatively, constant, effective level in the body with concomitant minimization of undesirable side effects associated with fluctuations in the plasma level of the active substance (sawtooth kinetic pattern); (iv) formulations that localize action by, e.g., spatial placement of a controlled release composition adjacent to or in contact with the thymus; (v) formulations that allow for convenient dosing, such that doses are administered, for example, once every one or two weeks; and (vi) formulations that target a neoplasia by using carriers or chemical derivatives to deliver the therapeutic agent to a particular cell type (e.g., neoplastic cell). For some applications, controlled release formulations obviate the need for frequent dosing during the day to sustain the plasma level at a therapeutic level.

[0245] Any of a number of strategies can be pursued in order to obtain controlled release in which the rate of release outweighs the rate of metabolism of the compound in question. In one example, controlled release is obtained by appropriate selection of various formulation parameters and ingredients, including, e.g., various types of controlled release compositions and coatings. Thus, the therapeutic is formulated with appropriate excipients into a pharmaceutical composition that, upon administration, releases the therapeutic in a controlled manner. Examples include single or multiple unit tablet or capsule compositions, oil solutions, suspensions, emulsions, microcapsules, microspheres, molecular complexes, nanoparticles, patches, and liposomes.

Parenteral Compositions

[0246] The pharmaceutical composition may be administered parenterally by injection, infusion or implantation (subcutaneous, intravenous, intramuscular, intraperitoneal, or the like) in dosage forms, formulations, or via suitable delivery devices or implants containing conventional, non-toxic pharmaceutically acceptable carriers and adjuvants. The formulation and preparation of such compositions are well known to those skilled in the art of pharmaceutical formulation. Formulations can be found in Remington: The Science and Practice of Pharmacy, supra.

[0247] Compositions for parenteral use may be provided in unit dosage forms (e.g., in single-dose ampoules), or in vials containing several doses and in which a suitable preservative

may be added (see below). The composition may be in the form of a solution, a suspension, an emulsion, an infusion device, or a delivery device for implantation, or it may be presented as a dry powder to be reconstituted with water or another suitable vehicle before use. Apart from the active agent that reduces or ameliorates a neoplasia, the composition may include suitable parenterally acceptable carriers and/or excipients. The active therapeutic agent(s) may be incorporated into microspheres, microcapsules, nanoparticles, liposomes, or the like for controlled release. Furthermore, the composition may include suspending, solubilizing, stabilizing, pH-adjusting agents, tonicity adjusting agents, and/or dispersing, agents.

[0248] As indicated above, the pharmaceutical compositions according to the invention may be in the form suitable for sterile injection. To prepare such a composition, the suitable active antineoplastic therapeutic(s) are dissolved or suspended in a parenterally acceptable liquid vehicle. Among acceptable vehicles and solvents that may be employed are water, water adjusted to a suitable pH by addition of an appropriate amount of hydrochloric acid, sodium hydroxide or a suitable buffer, 1,3-butanediol, Ringer's solution, and isotonic sodium chloride solution and dextrose solution. The aqueous formulation may also contain one or more preservatives (e.g., methyl, ethyl or n-propyl p-hydroxybenzoate). In cases where one of the compounds is only sparingly or slightly soluble in water, a dissolution enhancing or solubilizing agent can be added, or the solvent may include 10-60% w/w of propylene glycol.

Controlled Release Parenteral Compositions

[0249] Controlled release parenteral compositions may be in form of aqueous suspensions, microspheres, microcapsules, magnetic microspheres, oil solutions, oil suspensions, or emulsions. Alternatively, the active drug may be incorporated in biocompatible carriers, liposomes, nanoparticles, implants, or infusion devices.

[0250] Materials for use in the preparation of microspheres and/or microcapsules are, e.g., biodegradable/bioerodible polymers such as polygalactin, poly-(isobutyl cyanoacrylate), poly(2-hydroxyethyl-L-glutam- nine) and, poly(lactic acid). Biocompatible carriers that may be used when formulating a controlled release parenteral formulation are carbohydrates (e.g., dextrans), proteins (e.g., albumin), lipoproteins, or antibodies. Materials for use in implants can be non-biodegradable (e.g., polydimethyl siloxane) or biodegradable (e.g., poly(caprolactone), poly(lactic acid), poly(glycolic acid) or poly(ortho esters) or combinations thereof).

SCREENING METHODS

[0251] In example embodiments, cancers sensitive to phosphate dysregulation are screened by contacting tumors cells with inhibitors of PP-InsP synthesis as described herein and detecting the inositol pyrophosphate level in the cell or cell population, wherein the cancer is sensitive if the inositol pyrophosphate level is decreased as compared to a control cell or population not treated with the inhibitor; or detecting the phosphate concentration in the cell or cell population, wherein the cancer is sensitive if the phosphate concentration is increased as compared to a control cell or population not treated with the inhibitor. In example embodiments, the concentration of inhibitors of PP-InsP synthesis is selected such that the cells are still viable. In example embodiments, a titration of inhibitors is used to determine whether the tumor is sensitive to phosphate dysregulation. In example embodiments, the tumor cells are compared to a reference value for tumor cells that are sensitive to phosphate dysregulation. In example embodiments, tumor cells obtained from a subject are screened. Screening may be performed *in vitro* or *in vivo*. For example, agents that modulate phosphate export in a tumor microenvironment may be screened *in vivo*. For example, the invention provides for identifying an agent using a cell or complex cell population (e.g., multicellular systems, such as, organoid, tissue explant, or organ on a chip) and translating the agent to an *in vivo* system (e.g., therapeutic agents). In certain embodiments, cancer cell lines can be assayed for phosphate efflux as described herein.

[0252] In example embodiments, agents suitable for treating phosphate dysregulation sensitive tumors can be identified by applying candidate agents to tumor cells that are sensitive to phosphate dysregulation; and detecting modulation of phosphate efflux in the cell or cell population by the candidate agent. In example embodiments, IP6K inhibitors are tested. For example, IP6K inhibitors that are modified to improve specificity to IP6Ks or improve inhibition activity may be screened.

[0253] In example embodiments, cancer cell lines vulnerable to IP6K inhibitors or XPR1/phosphate efflux inhibition can be identified by treating cancer cell lines with inhibitors and determining viability. In one example, the cancer cell lines are barcoded and screened in a pooled fashion (see, e.g., Yu C, Mannan AM, Yvone GM, et al. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. Nat Biotechnol. 2016;34(4):419-423; and Corsello SM, Nagari RT, Spangler RD, et al. Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. Nat Cancer.

2020;1(2):235-248). PRISM allows pooled screening of mixtures of cancer cell lines by labeling each cell line with 24-nucleotide barcodes. The enrichment or depletion of barcodes can then be determined by sequencing to identify sensitive cancer cell lines.

[0254] Further embodiments are illustrated in the following Examples which are given for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLES

Example 1 – Inhibition of inositol pyrophosphate (PP-InsP) synthesis for the treatment of cancer

[0255] The SPX domain of XPR1 is de-repressed by binding to inositol pyrophosphates (InsP7/InspP8), which are synthesized by PPIP5K and IP6K enzymes (**Fig. 1** and **Fig. 2**). Applicants hypothesized that inhibition of inositol pyrophosphate (InsPP) synthesis would kill tumor cells sensitive to phosphate regulation through XPR1. Analysis of cells in the DepMap showed that IP6K1/2 are broadly expressed in the cell lines, while IP6K3 is expressed primarily in skeletal muscle cells (**Fig. 3**). Analysis of cells in the DepMap using chronos scores showed that ovarian and uterine cancer cell lines displayed a high degree of dependency on both XPR1 and KIDINS220, however dependencies in any of the cell lines was not observed for IP6K1, IP6K2, or IP6K3 (**Fig. 4**). DepMap analysis comparing phosphate transporters and InsPP synthesizing enzymes showed a strong co-dependency with KIDINS220, however, InsPP-producing enzymes are not well-correlated with XPR1 (**Fig. 5**). Therefore, DepMap analysis would not identify IP6K1, IP6K2, or IP6K3. Applicants hypothesized that the three IP6Ks paralogs are functionally redundant.

[0256] Applicants analyzed N2-(m-(trifluoromethyl)benzyl) N6-(p-nitrobenzyl)purine (TNP) (also known as, N2-(m-trifluorobenzyl), N6-(p-nitrobenzyl) purine), which is a commercially available IP6K inhibitor. Applicants did not observe a decrease in phosphate efflux in XPR1-dependent cell lines (RMGI) using TNP as compared to RBD, which is a protein inhibitor of XPR1 (see, e.g., International Patent Application PCT/US2021/045227) (**Fig. 6A**). Proliferation effects were observed at high doses of TNP, but the effects were not specific to XPR1-dependent cell lines (**Fig. 6B**).

[0257] Applicants analyzed SC-919, which inhibits all three IP6Ks with an *in vitro* IC_{50} ~5 nM, has a favorable PK/PD (pharmacokinetic/pharmacodynamic) profile in rats and monkeys, and showed improvement in blood chemistry and kidney fibrosis in a rat model of chronic

kidney disease (CKD) (see, e.g., Moritoh Y, Abe SI, Akiyama H, et al. The enzymatic activity of inositol hexakisphosphate kinase controls circulating phosphate in mammals. Nat Commun. 2021;12(1):4847). Applicants observed a cellular efflux IC_{50} of 50 nM in XPR1-dependent ovarian cell lines (**Fig. 7A**). Applicants observed selectivity for cell viability in ovarian cancer cells treated with SC-919, whereby SLC34A2 high, XPR1-dependent ovarian cancer cells had decreased viability as compared to SLC34A2 low, XPR1-non-dependent ovarian cancer cells (**Fig. 7B**). Applicants observed selectivity for cell viability in the RMGI ovarian clear cell carcinoma cell line treated with SC-919, whereby the wildtype, SLC34A2 high cancer cell line had decreased viability as compared to an isogenic, SLC34A2 inactivated RMGI cancer cell line (**Fig. 7C**). Applicants also tested the effects of SC-919 on a panel of ovarian cancer cell lines and found cell proliferation defects in *SLC34A2*-HIGH cell lines (e.g., OVISE, IGROV1, OVCAR3, RMGI). In contrast, *SLC34A2*-LOW cell lines were unaffected by treatment with the drug (59M, ES2, and RMGI after *SLC34A2* inactivation) (**Fig. 8**). In the ovarian cancer cell line OVISE after treatment with SC-919, Applicants observed an increase in dead and dying cells as measured by flow cytometric analyses of Annexin V staining and DAPI uptake. Doxycycline was used to induce an shRNA targeting *XPR1* (**Fig. 9**).

[0258] Applicants evaluated SC919 in PRISM. Barcoded cancer cell lines were pooled together, treated with increasing concentrations of SC-919, and barcode abundance was quantified using next generation sequencing. 400 cancer cell lines were observed in the highest dose group of SC919 (**Fig. 10**). Across the 400 cancer cell lines treated with 10 μM SC-919, applicants observed the highest correlation with cell lines that are sensitive to knockout of *XPR1* or *KIDINS220* and cell lines that express *SLC34A2* (**Fig. 11**).

***

[0259] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the

invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

CLAIMS

What is claimed is:

1. A method of treating cancer in a subject in need thereof comprising administering to the subject one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis.

2. The method of claim 1, wherein the cancer is characterized by high expression of SLC34A2 in tumor tissue as compared to expression in normal tissue.

3. A method of treating cancer in a subject in need thereof comprising:

detecting tumors sensitive to phosphate dysregulation by detecting in the subject increased expression of SLC34A2 relative to a control, wherein

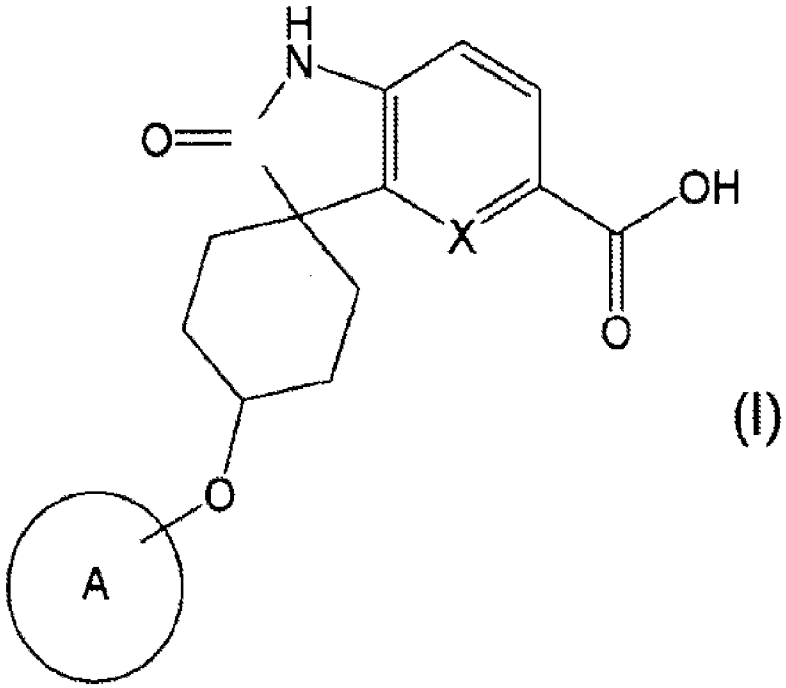
if the subject has a tumor sensitive to phosphate dysregulation, including administration of one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis;

if the subject does not have a tumor sensitive to phosphate dysregulation, administering a standard of care treatment that does not include administration of one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis.

4. The method of any of claims 1 to 3, wherein the one or more therapeutic agents is capable of inhibiting one or more inositol hexakisphosphate kinases (IP6Ks) selected from the group consisting of IP6K1, IP6K2, and IP6K3.

5. The method of claim 4, wherein the one or more agents are capable of inhibiting enzymatic activity of IP6K1, IP6K2, and/or IP6K3 with an *in vitro* IC_{50} of less than 10 nM, preferably about 5 nM.

6. The method of claim 4 or 5, wherein the inhibitor is represented by the formula:



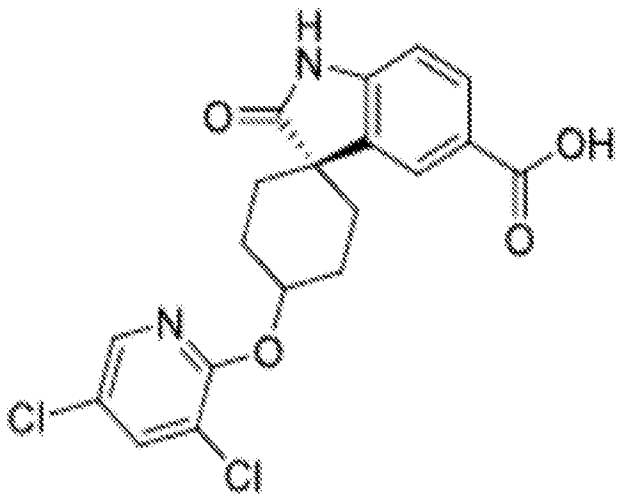
(I)

or a salt thereof, wherein

ring A is an optionally substituted aromatic ring;

X is CH or N.

7. The method of claim 6, wherein the inhibitor is represented by the formula:



or a salt thereof.

8. The method of any of claims 1 to 7, wherein the one or more therapeutic agents is capable of inhibiting one or more diphosphoinositol pentakisphosphate kinases (PPIP5Ks) selected from the group consisting of PPIP5K1 and PPIP5K2.

9. The method of any of claims 1 to 8, wherein the cancer is selected from the group consisting of ovarian cancer, uterine cancer, breast cancer, bile duct cancer, liver and lung cancer.

10. The method of any of claims 1 to 9, further comprising administering to the subject one or more therapeutic agents capable of inhibiting the suppression of SLC34A2.

11. The method of any of claims 1 to 10, wherein the one or more therapeutic agents capable of inhibiting inositol pyrophosphate (PP-InsP) synthesis are co-administered within a standard of care treatment regimen.

12. The method of claim 11, wherein the standard of care treatment regimen comprises surgery and chemotherapy.

13. The method of claim 11, wherein the standard of care treatment regimen comprises administration of an immunotherapy or a PARP inhibitor.

14. The method of claim 13, wherein the immunotherapy is a checkpoint blockade therapy.

15. A method for identifying a cancer sensitive to phosphate dysregulation, comprising:

applying an IP6K inhibitor to a cancer cell or cell population; and

detecting the inositol pyrophosphate level in the cell or cell population, wherein the cancer is sensitive if the inositol pyrophosphate level is decreased as compared to a control cell or population not treated with the inhibitor; or

detecting the phosphate concentration in the cell or cell population, wherein the cancer is sensitive if the phosphate concentration is increased as compared to a control cell or population not treated with the inhibitor.

16. The method of claim 15, wherein the cancer cell or population is obtained or derived from a subject in need thereof.

17. A method for identifying an agent capable of inhibiting XPR1-mediated phosphate export, comprising:

applying a candidate agent derived from an IP6K inhibitor to a cancer cell or cell population; and

detecting modulation of phosphate efflux in the cell or cell population by the candidate agent, thereby identifying the agent.

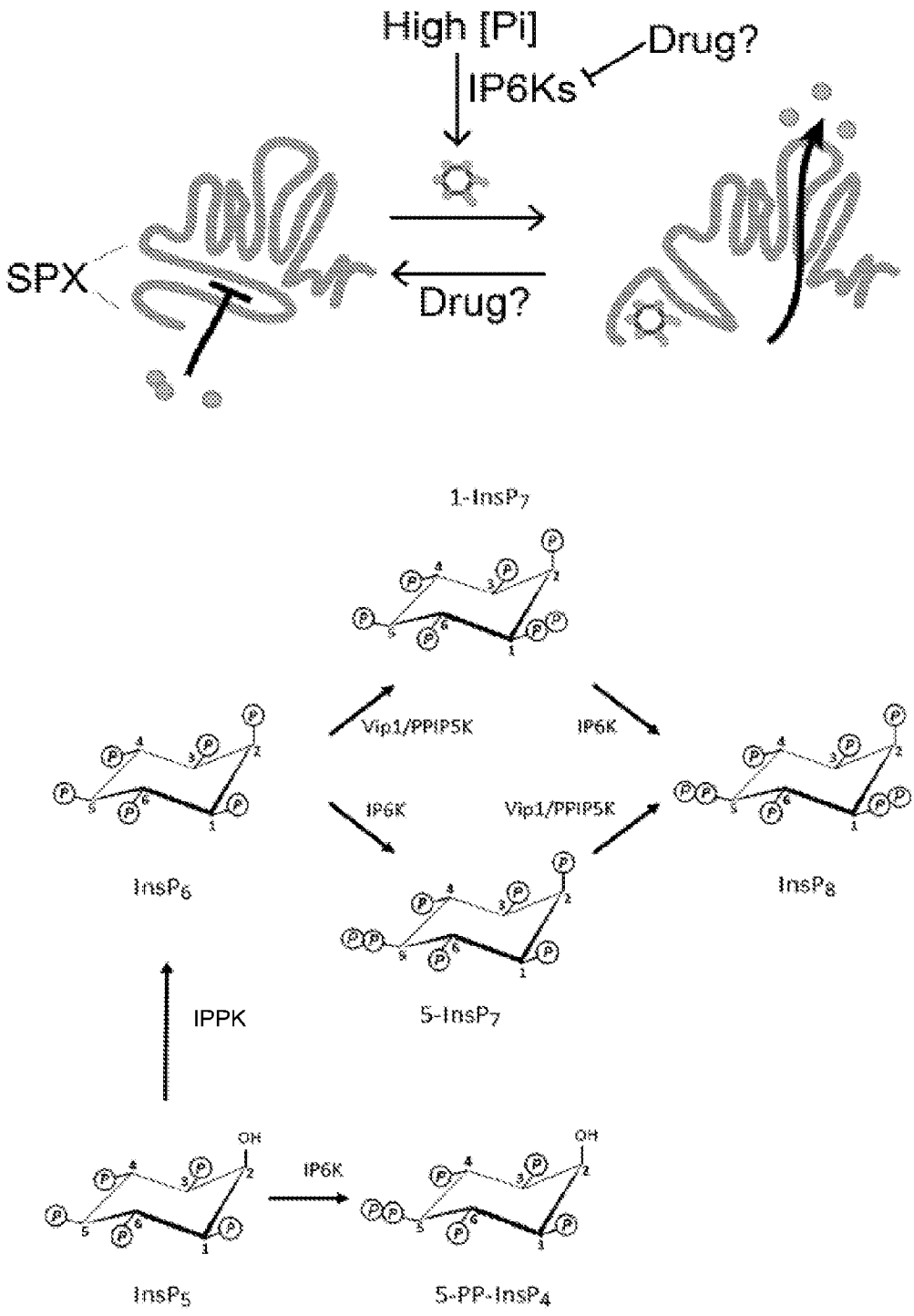

High [Pi]
IP6Ks
Drug?
SPX
Drug?
1-InsP7
Vip1/PPIP5K
IP6K
IP6K
Vip1/PPIP5K
InsP6
InsP8
5-InsP7
IPPK
IP6K
InsP5
5-PP-InsP4

FIG. 1

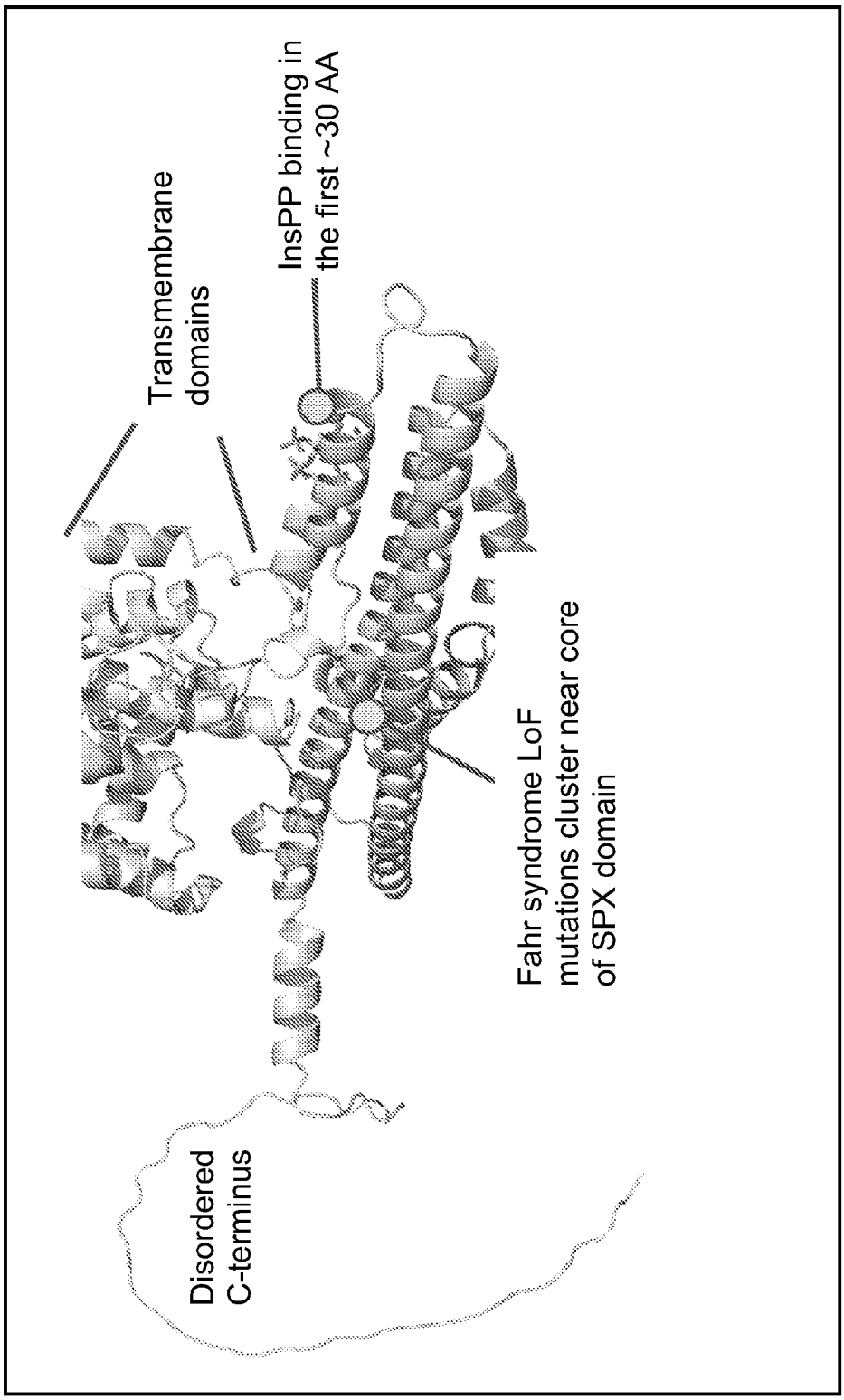

Transmembrane domains
InsPP binding in the first ~30 AA
Fahr syndrome LoF mutations cluster near core of SPX domain
Disordered C-terminus

FIG. 2

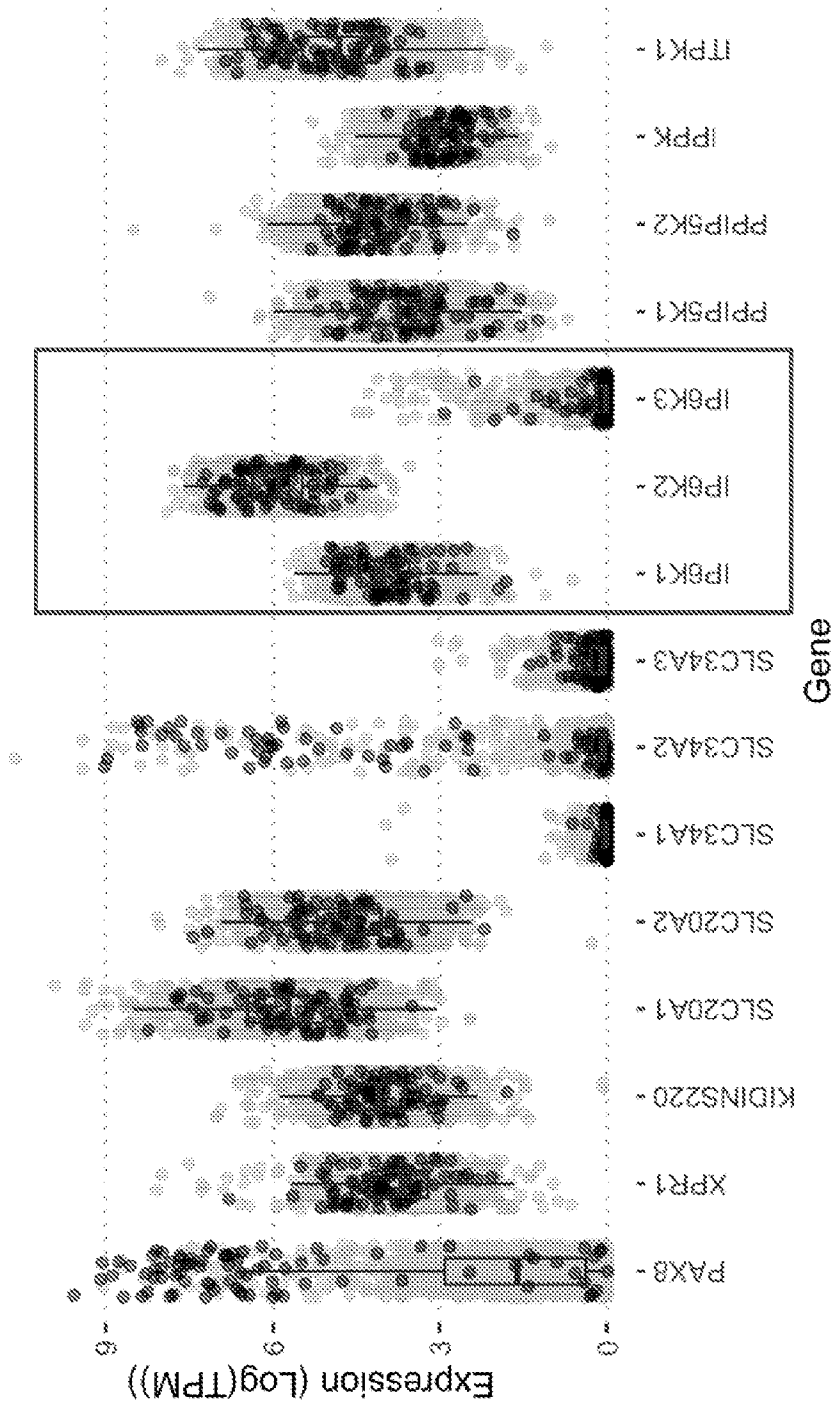

ITPK1
IPPK
PPIP5K2
PPIP5K1
IP6K3
IP6K2
IP6K1
SLC34A3
SLC34A2
SLC34A1
SLC20A2
SLC20A1
KIDINS220
XPR1
PAX8
Gene
Expression (Log(TPM))

FIG. 3

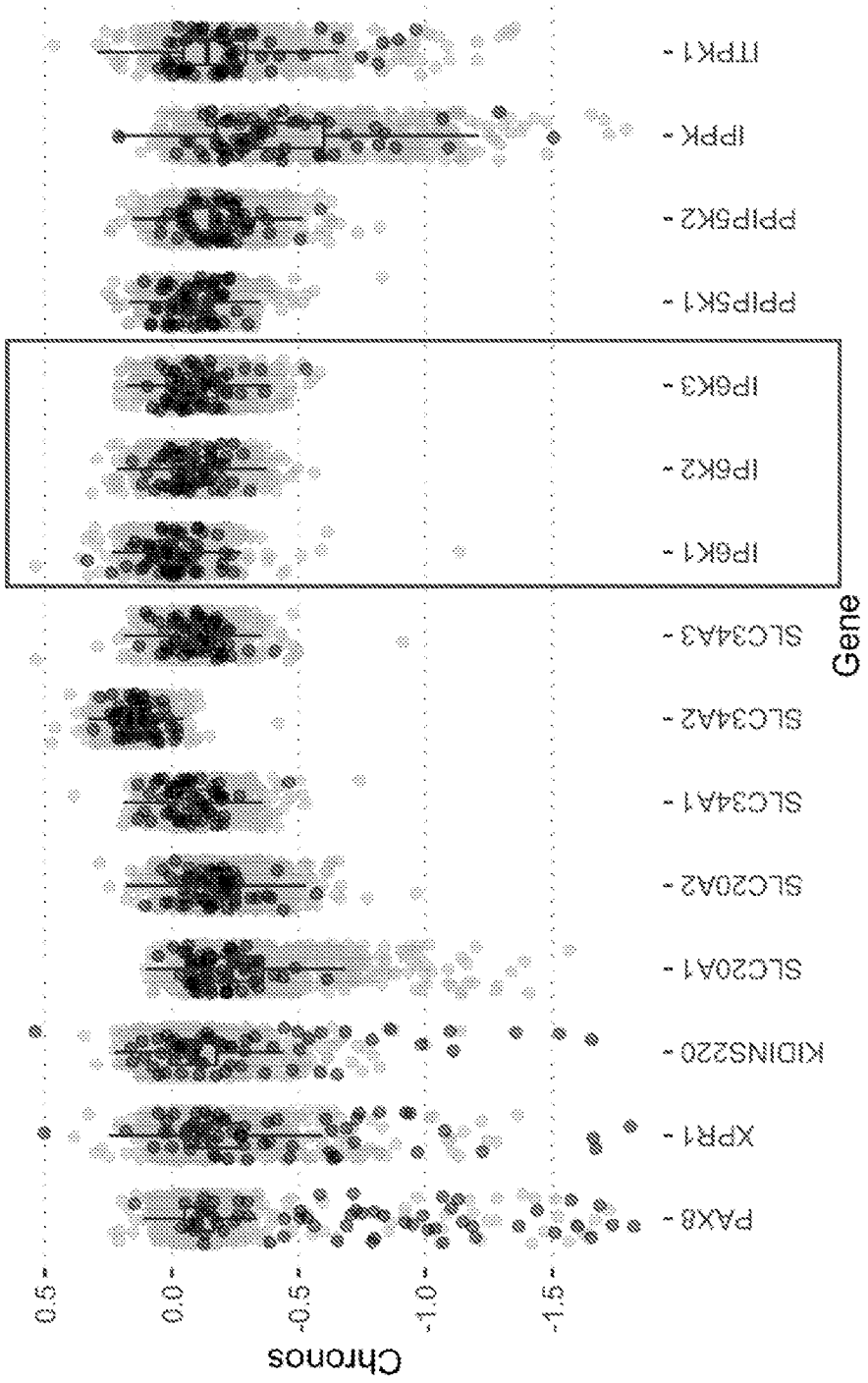

ITPK1
IPPK
PPIP5K2
PPIP5K1
IP6K3
IP6K2
IP6K1
SLC34A3
SLC34A2
SLC34A1
SLC20A2
SLC20A1
KIDINS220
XPR1
PAXB
Gene
0.5
0.0
-0.5
-1.0
-1.5
Chronos

FIG. 4

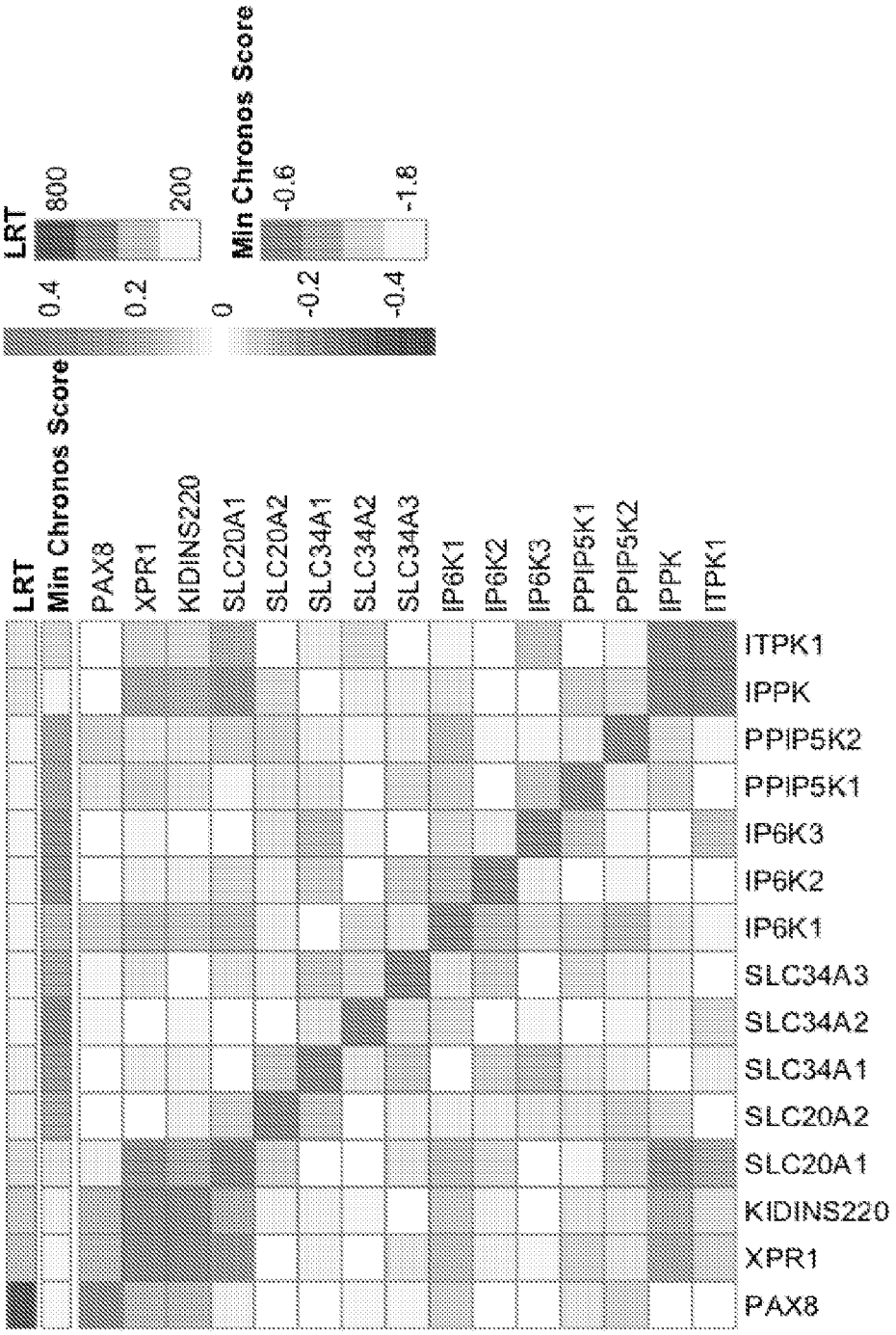

LRT
800
200
0
Min Chronos Score
-0.6
-1.8
-0.2
-0.4
0.4
0.2
LRT
Min Chronos Score
PAX8
XPR1
KIDINS220
SLC20A1
SLC20A2
SLC34A1
SLC34A2
SLC34A3
IP6K1
IP6K2
IP6K3
PPIP5K1
PPIP5K2
IPPK
ITPK1
ITPK1
IPPK
PPIP5K2
PPIP5K1
IP6K3
IP6K2
IP6K1
SLC34A3
SLC34A2
SLC34A1
SLC20A2
SLC20A1
KIDINS220
XPR1
PAX8

FIG. 5

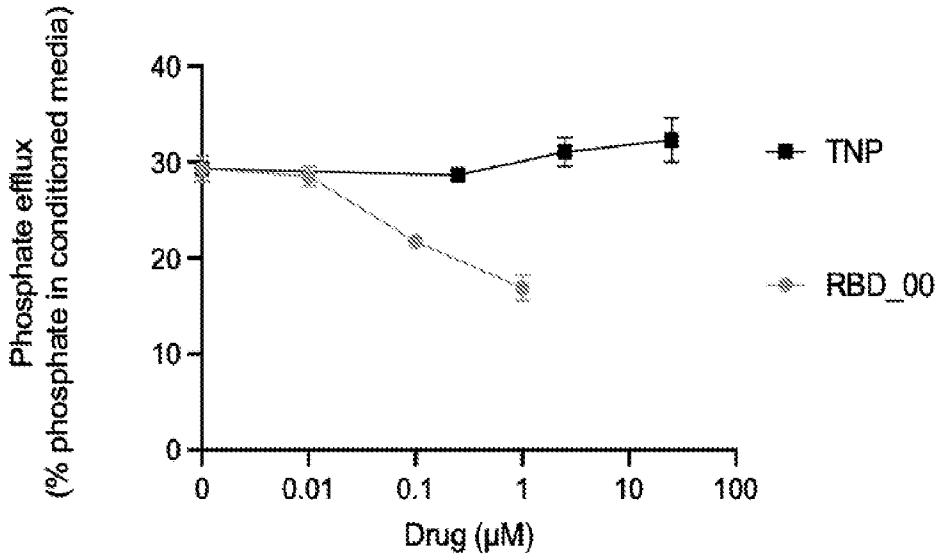

Phosphate efflux
(% phosphate in conditioned media)
40
30
20
10
0
0
0.01
0.1
1
10
100
Drug (μM)
TNP
RBD_00

FIG. 6A

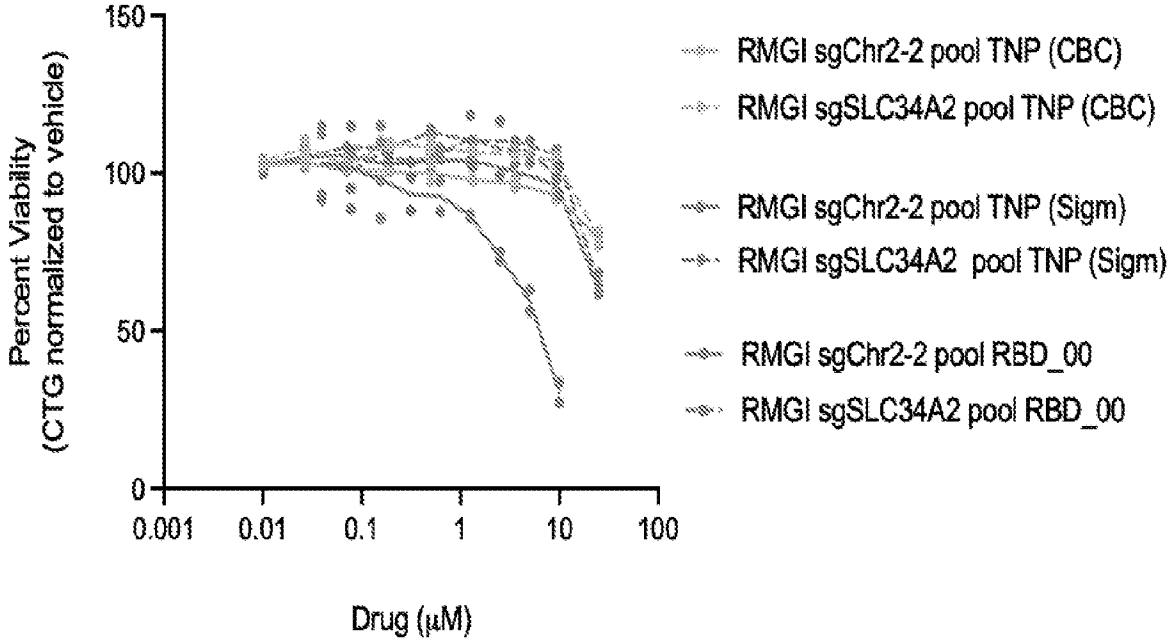

Percent Viability
(CTG normalized to vehicle)
150
100
50
0
0.001
0.01
0.1
1
10
100
Drug (μM)
RMGI sgChr2-2 pool TNP (CBC)
RMGI sgSLC34A2 pool TNP (CBC)
RMGI sgChr2-2 pool TNP (Sigm)
RMGI sgSLC34A2 pool TNP (Sigm)
RMGI sgChr2-2 pool RBD_00
RMGI sgSLC34A2 pool RBD_00

FIG. 6B

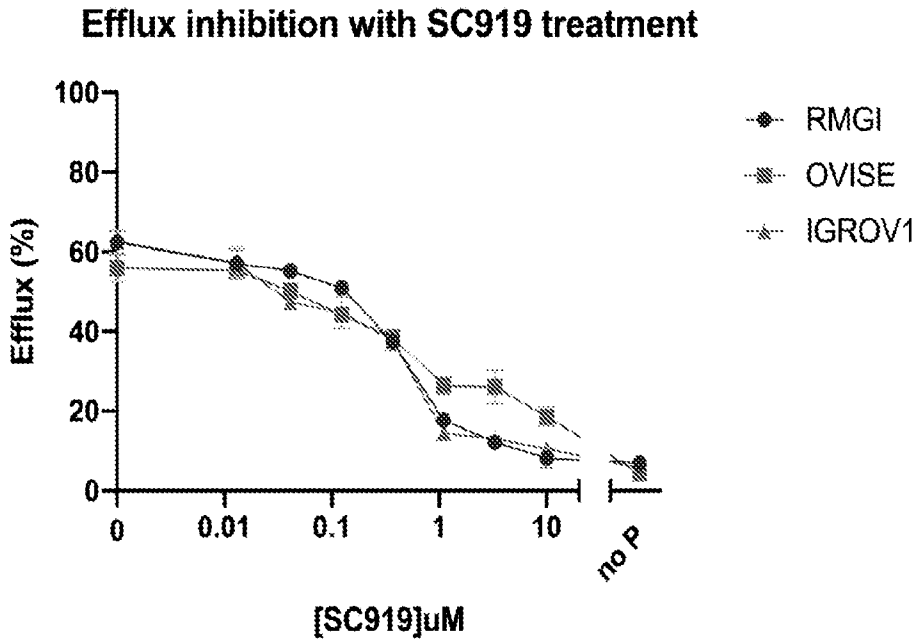

Efflux inhibition with SC919 treatment
Efflux (%)
100
80
60
40
20
0
0
0.01
0.1
1
10
no P
[SC919]uM
RMGI
OVISE
IGROV1

FIG. 7A

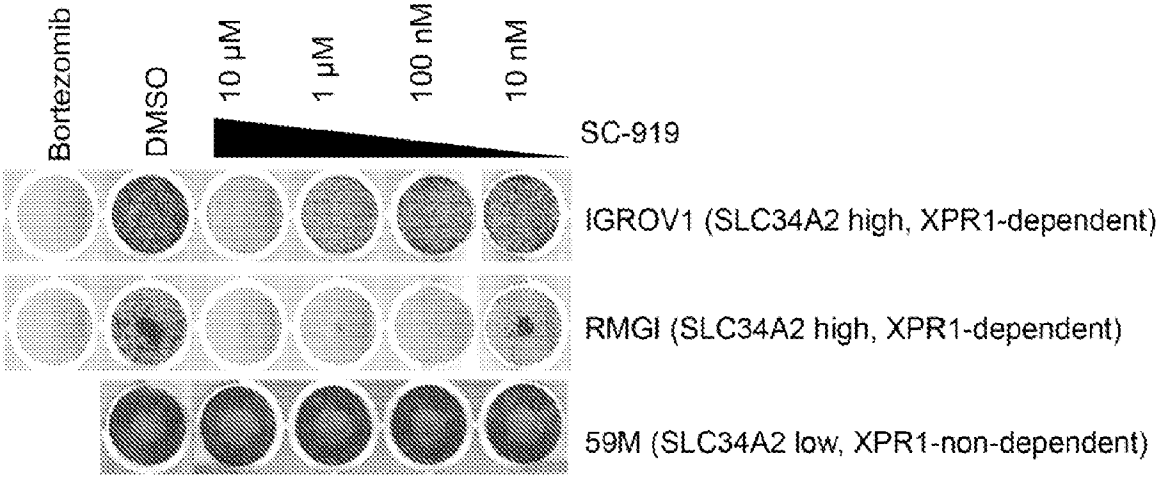

Bortezomib
DMSO
10 µM
1 µM
100 nM
10 nM
SC-919
IGROV1 (SLC34A2 high, XPR1-dependent)
RMGI (SLC34A2 high, XPR1-dependent)
59M (SLC34A2 low, XPR1-non-dependent)

FIG. 7B

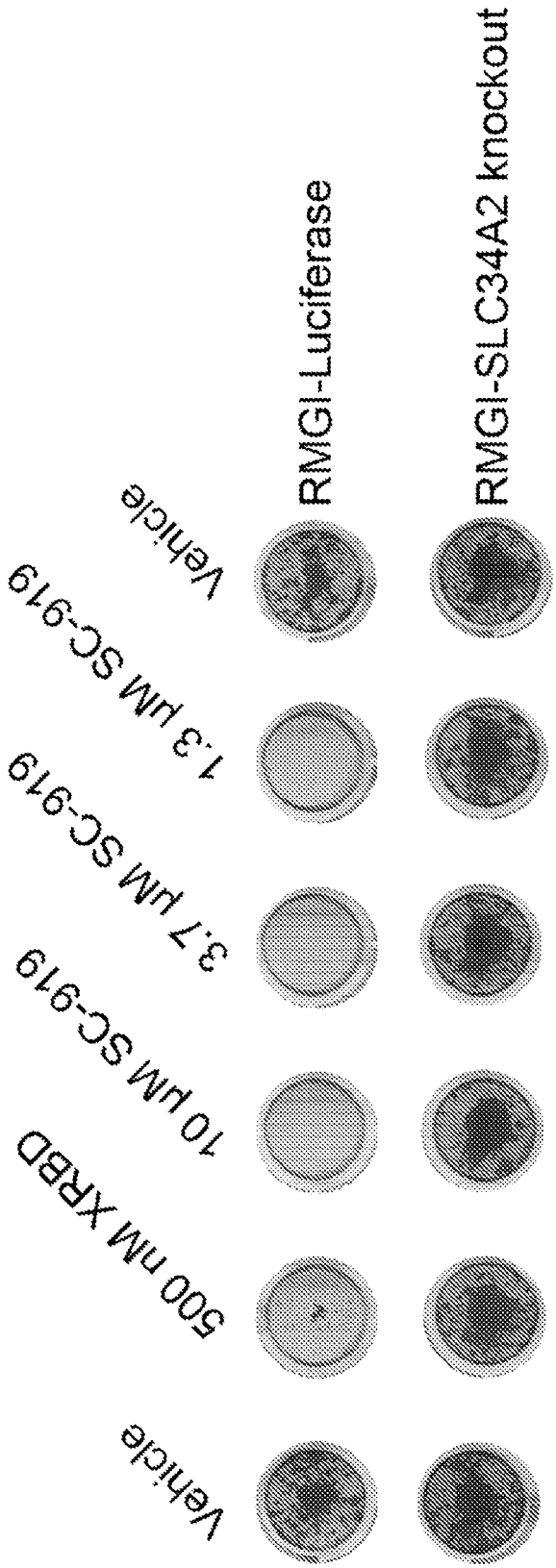

RMGI-SLC34A2 knockout
RMGI-Luciferase
Vehicle
1.3 μM SC-919
3.7 μM SC-919
10 μM SC-919
500 nM XRBD
Vehicle

FIG. 7C

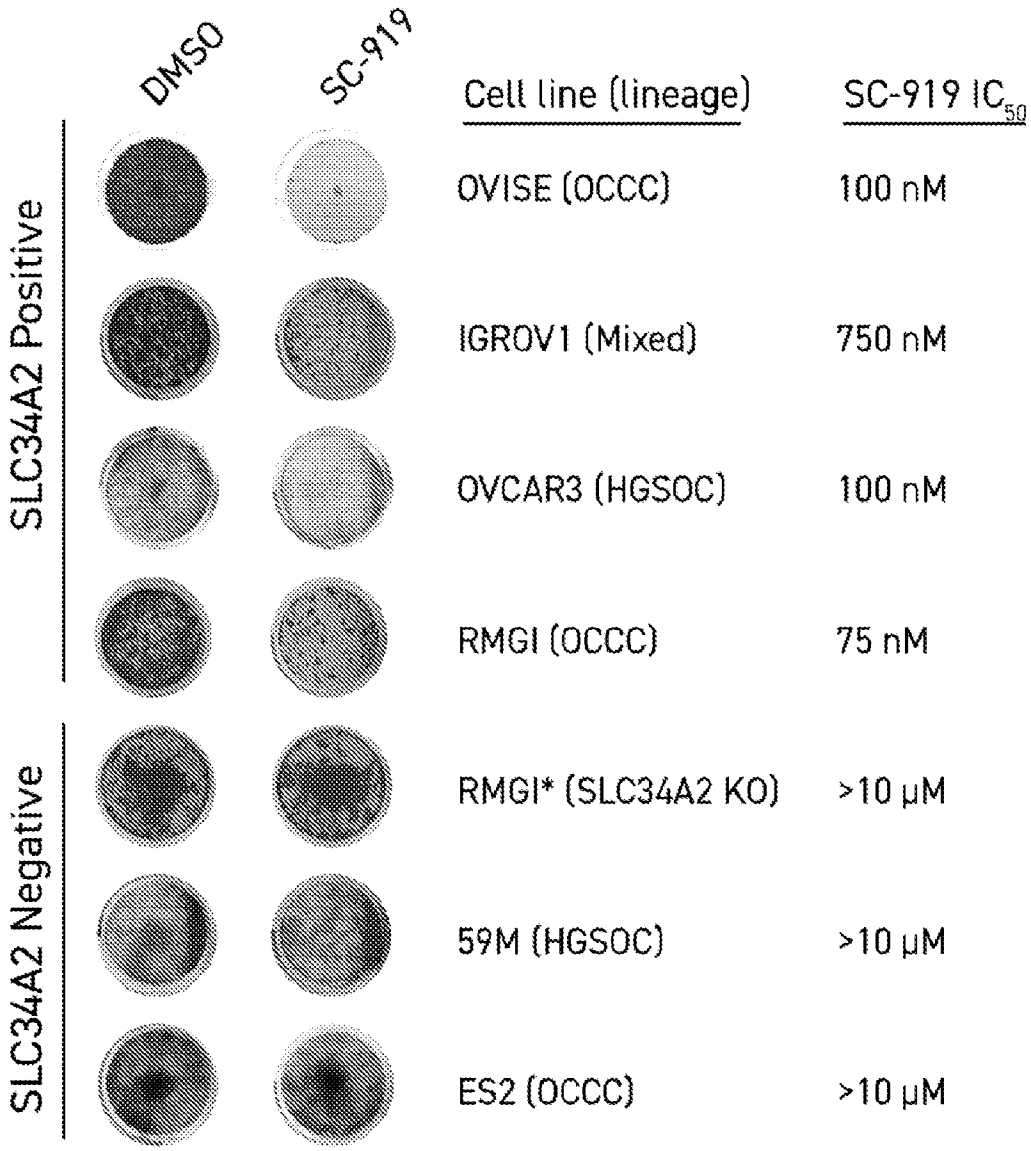

DMSO
SC-919
Cell line (lineage)
SC-919 IC50
SLC34A2 Positive
SLC34A2 Negative
OVISE (OCCC)
100 nM
IGROV1 (Mixed)
750 nM
OVCAR3 (HGSOC)
100 nM
RMGI (OCCC)
75 nM
RMGI* (SLC34A2 KO)
>10 µM
59M (HGSOC)
>10 µM
ES2 (OCCC)
>10 µM

FIG. 8

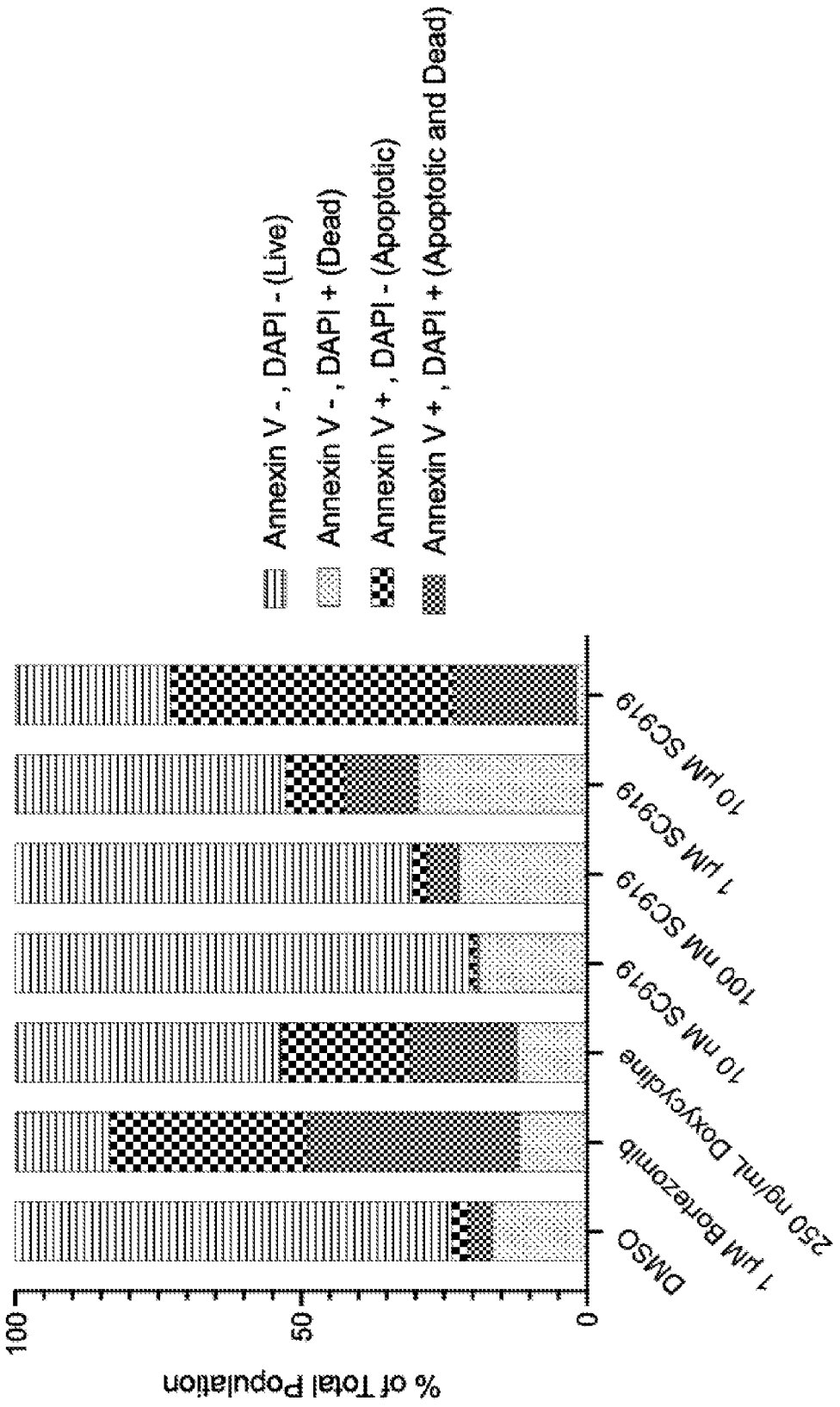

Annexin V -, DAPI - (Live)
Annexin V -, DAPI + (Dead)
Annexin V +, DAPI - (Apoptotic)
Annexin V +, DAPI + (Apoptotic and Dead)
% of Total Population
100
50
0
DMSO
1 μM Bortezomib
250 ng/mL Doxycycline
10 nM SC919
100 nM SC919
1 μM SC919
10 μM SC919

FIG. 9

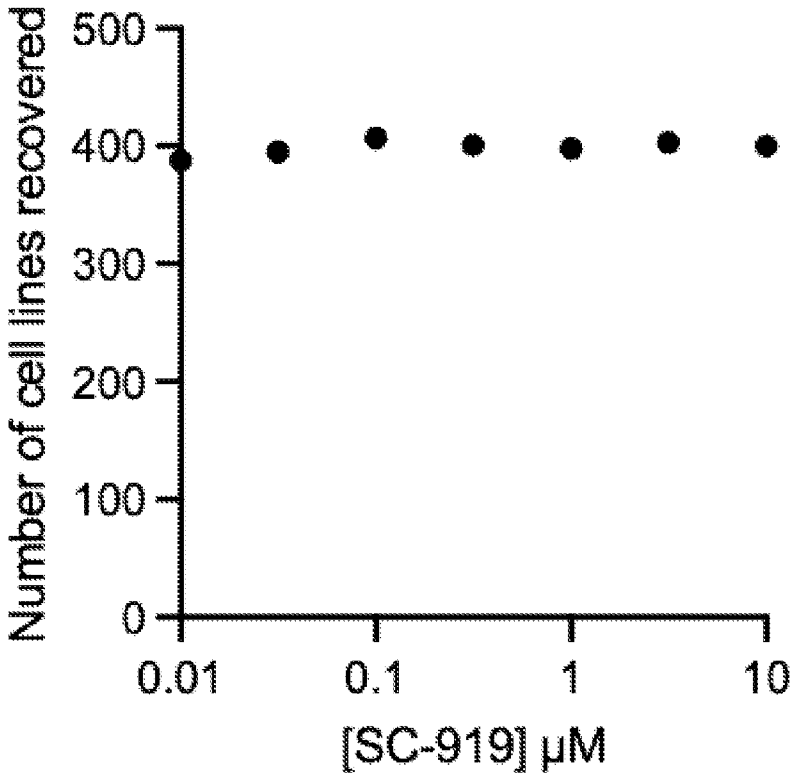

Number of cell lines recovered
500
400
300
200
100
0
0.01
0.1
1
10
[SC-919] μM

FIG. 10

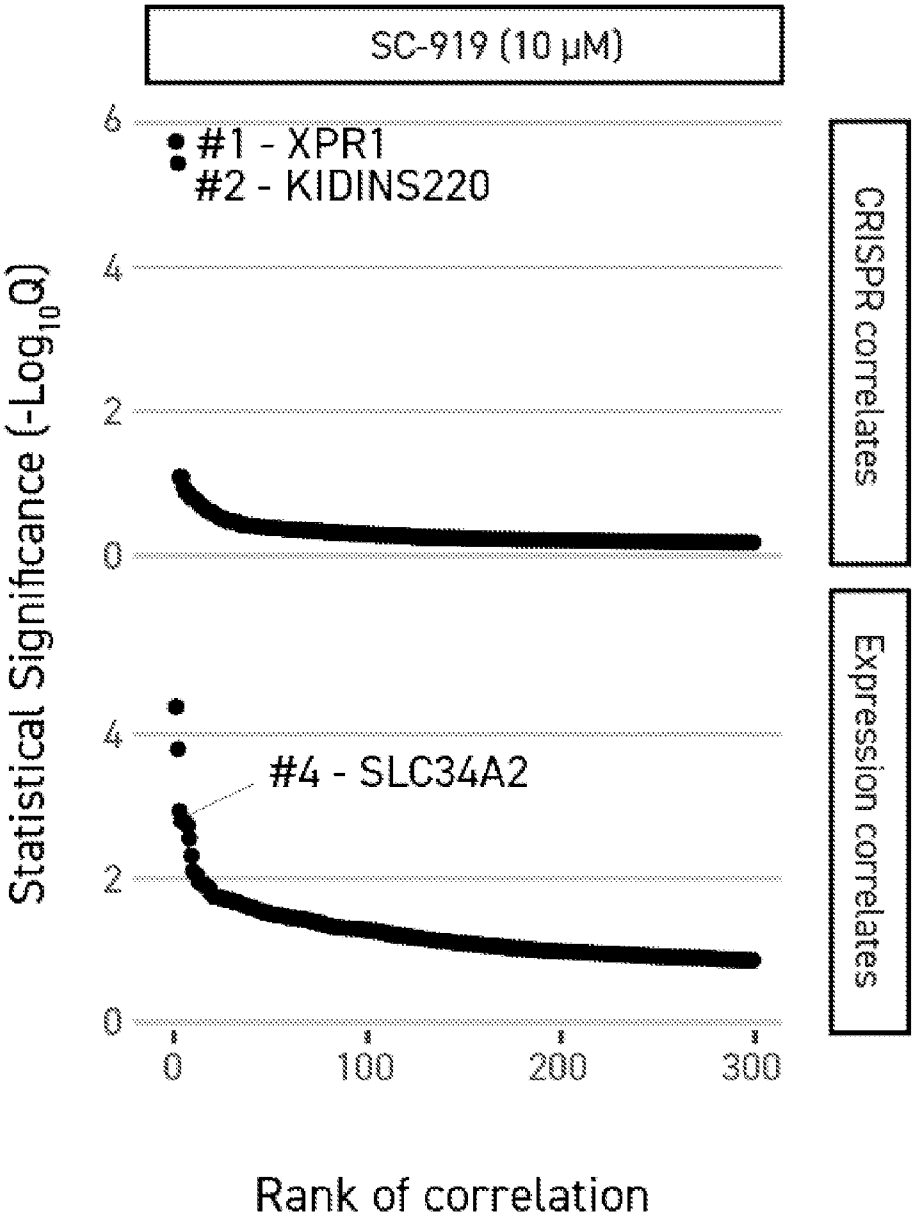

SC-919 (10 μM)
#1 - XPR1
#2 - KIDINS220
#4 - SLC34A2
CRISPR correlates
Expression correlates
Statistical Significance (-Log10Q)
Rank of correlation
0
2
4
6
0
100
200
300

FIG. 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 23/69476

A. CLASSIFICATION OF SUBJECT MATTER

IPC - INV. A61K 31/395, A61K 31/025, A61K 31/015 (2023.01)
ADD. A61K 31/33 (2023.01)

CPC - INV. A61K 31/395, A61K 31/025, A61K 31/015
ADD. A61K 31/33

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)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---- Y	Jadav et al. Deletion of inositol hexakisphosphate kinase 1 (IP6K1) reduces cell migration and invasion, conferring protection from aerodigestive tract carcinoma in mice. Cellular Signalling 28 (2016) 1124–1136; title, abstract, pg1133, col2, ln8-9, 12-13;	1, 1/4, 5/4/1 ---------- 2, 4/2, 5/4/2
Y	WO 2022/032228 A1 (The Broad Institute, Inc. et al.) 10 February 2022 (10.02.2022) abstract, para [0008]-[0010], [0013]-[0014]	2-3, 4/(2-3), 5/4/(2-3), 15-17
Y	Wilson et al. The inositol hexakisphosphate kinases IP6K1 and -2 regulate human cellular phosphate homeostasis, including XPR1- mediated phosphate export. title,J.Biol.Chem.(2019)294(30)11597–11608, pg11597, col1, ln13-14, 19-20; pg11606, col1, ln11;	3, 4/3, 5/4/3, 15-17
A	Minini et al. The Key Role of IP6K: A Novel Target for Anticancer Treatments. Molecules 2020, 25, 4401; pg1-20	1-5, 15-17
A	Rao et al. Inositol pyrophosphates promote tumor growth andmetastasis by antagonizing liver kinase B1. PNAS February 10, 2015, vol. 112, no. 6, 1773–1778	1-5, 15-17
A	US 2010/0129806 A1 (Snyder et al.) 27 May 2010 (27.05.2010) entire document	1-5, 15-17
X, P	WO 2023/154466 A1 (THE IINIVERSITY OF NORTH CAROLINA AT CHAPEL HILL) 17 August 2023 (17.08.2023) entire document	1-5, 15-17

☐ 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
"D" document cited by the applicant in the international application
"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 another 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
12 October 2023 (10.12.2023)	NOV 0 8 2023
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Kari Rodriquez Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 23/69476

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as fi led.

   b. ☐ furnished subsequent to the international fi ling date for the purposes of international search (Rule 13ter.1(a)),

      ☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 23/69476

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6-14
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.