

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

30 December 2020 (30.12.2020)



(10) International Publication Number

WO 2020/260899 A1

(51) International Patent Classification:

G01N 33/50 (2006.01)

C12N 9/22 (2006.01)

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:

PCT/GB2020/051559

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

(22) International Filing Date:

26 June 2020 (26.06.2020)

Published:

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

— with sequence listing part of description (Rule 5.2(a))

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1909228.7

27 June 2019 (27.06.2019)

GB

(71) Applicants: AZERIA THERAPEUTICS LIMITED

[GB/GB]; CRUK Cambridge Institute, Robinson Way, Cambridge CB2 0RE (GB). **CANCER RESEARCH TECHNOLOGY LIMITED** [GB/GB]; 2 Redman Place, London E20 1JQ (GB).

(72) Inventors: CARROLL, Jason Scott; C/o Azeria Therapeutics Limited, CRUK Cambridge Institute, Robinson Way, Cambridge CB2 0RE (GB). HOLMES, Kelly-Ann; 15 Rowley Road, St Neots Cambridgeshire PE19 1UF (GB). JOSEPH, Nimesh; C/o Azeria Therapeutics Limited, CRUK Cambridge Institute, Robinson Way, Cambridge CB2 0RE (GB).

(74) Agent: HGF LIMITED; Document Handling - HGF (Manchester), 1 City Walk, LEEDS Yorkshire LS11 9DX (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available):

AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: SCREEN FOR INHIBITORS

(57) Abstract: The present invention relates to methods for screening for a putative regulator of a DNA binding protein that inhibits the DNA binding of said protein, using a gain of function assay endpoint in a cellular context. The method comprises bringing into contact (i) a host cell that expresses an inducible fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide; and, (ii) a test compound, and determining whether the test compound reduces the ability of the fusion protein to cleave cellular DNA.



WO 2020/260899 A1

SCREEN FOR INHIBITORS

FIELD OF THE INVENTION

The present invention relates to methods for screening for a putative regulator of a DNA binding protein that inhibits the DNA binding of said protein, using a gain of function assay endpoint in a cellular context.

BACKGROUND TO THE INVENTION

The gene expression programs that establish and maintain specific cell states are controlled by DNA binding proteins including transcription factors and chromatin regulators. DNA binding proteins are commonly deregulated in the pathogenesis of many human diseases including diabetes, inflammatory disorders, cardiovascular disease and many cancers (Lee and Young, *Cell*. 152(6):1237–1251, 2013). Consequently, the targeting of these mechanisms can be highly effective in treating disease. For example, agents that target the nuclear hormone receptors androgen receptor (AR) and oestrogen receptor (ER), examples of transcription factors, have clinical efficacy in the treatment of prostate and breast cancer, respectively.

In disease contexts DNA binding proteins are often amplified, deleted, rearranged via chromosomal translocation, or subjected to mutations that result in a gain- or loss-of-function (Darnell, J.E. *Nat Rev Cancer*. 2:740–749, 2002). For example, the autoimmune regulator (AIRE) is a transcription factor involved in promoting transcriptional elongation at genes with paused RNA polymerase II in the thymus, and mutations in AIRE protein are now recognised to cause type I autoimmune polyendocrinopathy syndrome (Kyewski and Klein, *Annu Rev Immunol*. 24:571–606, 2006). For example, mutations in pancreatic master transcription factors have been implicated in diabetes. The gene expression programs of pancreatic cells are controlled by a small set of key transcription factors, including HNF1 α , HNF1 β , HNF4 α , PDX1 and NeuroD1, and mutations in any of these factors can result in various forms of maturity-onset diabetes of the young (Maestro et al., *Endocr Dev*. 12:33-45, 2007; Malecki, M.T., *Diabetes Res Clin Pract*. 68 Suppl1:S10-21, 2005). Furthermore, TP53 and MYC, which encode the transcription factors (TFs) p53 (tumor protein 53) and c-Myc, are among the most commonly altered genes across all cancers (Lee and Muller, *Cold Spring Harb Perspect Biol*, Oct;2(10): a003236, 2010; Betrones et al., *Biochim Biophys Acta*. 1849(5), 506–516, 2015).

Many of the DNA binding proteins implicated in human diseases are considered intractable and difficult to target by drugs owing to a prevalence of unstructured domains and a lack of

druggable binding sites. Furthermore, conventional methods used to target transcription factor activity with molecules mimicking endogenous ligands commonly fail to achieve high specificity and are limited by a lack of identification of new molecular targets. In the notable exceptions where such conditions are met, drugging transcription factors has been highly successful, as evidenced by the clinical efficacy associated with the inhibition of ER and AR in breast and prostate cancers, respectively.

There is therefore significant unmet need for new approaches to identifying means to inhibit DNA binding protein activity. One such approach may be the inhibition of the ability of the DNA binding protein to interact with DNA or chromatin. For such an approach, cell-based or phenotypic screening methods are preferred, as opposed to more traditional *in vitro* biochemical target-based discovery methods. This is because DNA binding proteins often function in a highly cell or stimulus context-dependent manner, for example functioning only in the presence of certain co-factors (Lee and Young, *Cell*. 152(6):1237–1251, 2013). A phenotypic or cell-based screen addresses this issue by being more physiologically relevant and less artificial because intact cells and native cellular environment are preserved (Zheng et al., *Drug Discov Today*. 18(21-22), 1067–1073, 2013). Furthermore, DNA binding proteins are particularly appropriate for phenotypic or cell-based screening where often it is desirable to identify unknown regulatory mechanisms.

Generally, development of a phenotypic screening assay exploits a characteristic associated with the disease. Agents such as small molecules, or nucleic acid-based molecules such as small interfering RNA (siRNA), single guide RNA (sgRNA) and small hairpin RNA (shRNA) are screened in the assay to identify those agents that can ameliorate the disease phenotype, exemplified by selectively killing cancer cells, eliminating pathogens in culture, or reducing lysosomal cholesterol accumulation in Niemann Pick disease type C patient cells (Zheng et al., *Drug Discov Today*. 18(21-22):1067–1073, 2013). A common cell-based screening approach used in the art for identifying inhibitors of DNA binding proteins is a cell-based assay whereby the expression of a fluorescent protein or enzyme that produces bioluminescence (e.g. luciferase) is dependent upon the DNA binding protein of interest. Small molecules, siRNAs, sgRNAs, or shRNAs are then used to identify agents capable of reducing fluorescence or bioluminescence (Amante and Badr, *Methods Mol Biol*. 1098:185-95, 2014). However, a major challenge of such approaches is that they result in loss of signal (e.g. loss of bioluminescence, cancer cell killing), and in the complex environment of the cell, where many non-specific perturbations may result in loss of signal or cell death, this method is prone to false positives (Kaelin, W.G., *Nat Rev Cancer*. 17(7):425-440, 2017). In addition, these reporter systems assume that the one regulatory element used to make the

assay is representative of the thousands or tens of thousands of endogenous DNA binding sites and may therefore not be biologically meaningful.

Accordingly, there is a need in the field to identify gain of signal assays that can be applied to the discovery of novel agents capable of inhibiting the binding of DNA binding proteins to DNA and chromatin.

SUMMARY OF THE INVENTION

The present invention is based on work carried out to identify new methods for identifying inhibitors of the binding of a protein of interest to DNA or chromatin.

Suitably, such method (i) uses a gain of signal endpoint; (ii) has an endpoint that is specific to the loss of DNA or chromatin binding of the protein of interest; and (iii) retains the native cellular environment. Ideally, such gain of signal endpoint is not an endpoint associated with loss of cell fitness (for example, gain of apoptotic marker expression such as Annexin V), but should instead be a gain of cell fitness.

As described further herein, the invention arises from the discovery that the expression in cells of a protein chimera comprised of a fusion of a functionally active DNA endonuclease domain with a DNA binding domain from a DNA binding protein induces DNA damage via DNA cleavage and, typically, cell death in a manner that is (i) dependent on the DNA binding of the DNA binding domain; and (ii) can be inhibited using agents that disrupt the DNA binding of said DNA binding domain. This provides a cell-based assay whereby agents capable of inhibiting DNA or chromatin binding of the protein of interest can be identified using a method specific to the protein domain of interest, in the native cell environment, and where the assay endpoint is, for example, a gain of cell viability or increase in cell number or reduction in DNA damage.

According to a first aspect of the invention there is provided a method of screening for an inhibitor of a DNA binding protein comprising:

- a) bringing into contact (i) a host cell that expresses an inducible fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide; and, (ii) a test compound, under conditions where the fusion protein, in the absence of the test inhibitor compound, is capable of binding to and cleaving host cellular DNA; and
- b) determining whether the test compound reduces the ability of the fusion protein to cleave cellular DNA

wherein if the test compound inhibits the ability of the fusion protein to cleave cellular DNA the test compound is a putative inhibitor of the DNA binding protein.

There are various ways of determining whether the test compound interferes with the ability of the fusion protein to cleave cellular DNA. Suitably, it is the effect on cell growth, such as cell number or viability. In particular embodiments, the ability of the test compound to inhibit the ability of the fusion protein to cleave cellular DNA can be determined indirectly by measuring a host cell phenotype, such as cell number, cell growth, cell proliferation, cell viability, cellular DNA damage or cell cycle phase distribution. Suitably, cell growth or viability is used to determine if the test compound is an inhibitor of the DNA binding protein.

According to a variation of this aspect of the invention there is provided a method of screening for an inhibitor of a DNA binding protein comprising:

- a) bringing into contact (i) a host cell that expresses an inducible fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide; and, (ii) a test inhibitor compound, under conditions where the fusion protein, in the absence of the test inhibitor compound, is capable of binding to and cleaving host cellular DNA with the effect of reducing or eliminating the viability or proliferation of the cell; and
- b) determining cell number, cellular DNA damage, cell viability, cell growth, cell cycle phase distribution or cell proliferation.

wherein the amount of cells, cellular DNA damage, cell viability, cell growth, cell cycle phase distribution or cell proliferation determined indicates whether or not the test compound is an inhibitor of the DNA binding protein

Whilst the assay can be performed using a single cell, it is advantageously carried out using a clonal colony/population of a cell. Thus, when used herein the term "a cell" also includes cells plural.

As used herein, "method of screening for an inhibitor of a DNA binding protein" includes any assay for identifying an agent, e.g. compound, that inhibits the binding of the DNA binding protein to its target, when incubated with an appropriate test cell. This inhibition could be caused directly or indirectly (e.g. by inhibiting a protein that facilitates the binding of the DNA binding protein (e.g. transcription factor) to its target).

The assay is suitable for detecting inhibitors of any DNA binding protein. Exemplary DNA binding proteins include: AHR, AR, ARHGAP35, ARID5B, ARNTL, ASCL1, ASH1L, ATF1, BCL11A, BCL6, BCLAF1, BNC2, BRN2, C/EBPalpha or C/EBPbeta, CBFB, CIC, CLOCK,

CSL, CTCF, DMRT1, ELF1, ELF3, ELF4, ESR1, EZH2, FOXA1, FOXA2, FOXA3, FOXC2, FOXP1, FOXP3, GATA1, GATA2, GATA3, GATA4, HLF, HNF1A, IRF2, IRF6, IRF7, KLF4, KLF6, MAFA, MAX, MECOM, MEF2, MFE2L2, MGA, MYB, MYC, MYCN, MYOD, MYTL1, NFATC4, NGN3, NKX3-1, NR2F2, NR4A2, NR5A1, OCT4, PAX5, PDX1, PGR, PML-RARA, PRMD16, PRRX1, RUNX1, SMAD2, SMAD4, SOX2, SOX7, SOX9, TAL1, TBX3, TBX5, TCF12, TCF4, TCF7L2, TFDP1, and TFDP2.

In particular embodiments, the DNA binding protein is a transcription factor. A suitable type of transcription factor is a pioneer factor, such as one selected from the group consisting of: FOXA1, FOXA2, FOXA3, Zelda, Class V Pou, Oct3/4, Pou5f3, Group B1 Sox, Sox2, Klf4, Asc1, Pax7, PU.1, GATA1, GATA2, GATA4, CLOCK:BMAL1 and p53.

The fusion protein utilised in the invention comprises a first part (polypeptide) which is a DNA binding domain from a DNA binding protein and a second part (polypeptide) which is a functionally active DNA endonuclease polypeptide which lacks its endogenous DNA binding domain. The first part could be the full-length protein (e.g. transcription factor) or a truncated version that comprises the DNA binding domain region of the DNA binding protein.

In particular embodiments, the DNA endonuclease is a homing endonuclease from the LAGLIDADG, HNH, His-Cys box, and GIY-YIG families.

According to a second aspect of the invention there is provided an inhibitor of a DNA binding protein identified in accordance with the first aspect of the invention.

According to a third aspect of the invention there is provided a host cell for use in the method of the first aspect of the invention; i.e. one that expresses an inducible fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide.

According to a fourth aspect of the invention there is provided a pharmaceutical composition comprising an inhibitor of a DNA binding protein in accordance with the second aspect of the invention.

According to a fifth aspect of the invention there is provided method of preparing a pharmaceutical composition comprising an inhibitor of a DNA binding protein comprising: (i) selecting a compound that is an inhibitor of a DNA binding protein according to the first aspect of the invention; and, (ii) admixing the inhibitor with one or more pharmaceutically acceptable excipients.

According to a sixth aspect of the invention there is provided the inhibitor of a DNA binding protein identified in accordance with the first aspect of the invention or pharmaceutical composition according to the third aspect of the invention for use in therapy.

According to a seventh aspect of the invention there is provided a method of treatment of an individual in need of treatment with an inhibitor of a DNA binding protein comprising administering to said individual a therapeutically effective amount of a pharmaceutical composition comprising an inhibitor of a DNA binding protein, such as in accordance with aspect three of the invention.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and embodiments:

“DNA binding protein” as used herein refers to a protein capable of binding to DNA or chromatin, or which localises to cellular DNA or chromatin, in each case either directly or indirectly. These proteins typically have a DNA-binding domain that has specificity for binding to DNA. Sequence-specific DNA-binding proteins bind to particular sequences of single- or double-stranded DNA.

Exemplary DNA binding proteins include human transcription factors, which include Myc, members of the forkhead box family (including FOXP3, FOXC2, FOXA1, FOXA2, FOXA3, FOXO3a, FOXM1, and FOX3D), nuclear and orphan nuclear receptor (including androgen receptor, oestrogen receptor, hepatocyte nuclear factor 4alpha, peroxisome proliferator-activated receptors, nuclear receptor subfamily 1 group D, retinoic acid receptor-related orphan receptors) families (Lambert, S.A. et al. Cell.172(4): 650 – 665, 2018), and chromatin remodellers including BRG1, SNF2H, CHD3, CHD4, BRM, INO80, SNF2L and CHD1 (Giles, K.A. et al. Epigenetics Chromatin. 2019 12(1):12) and epigenetic modifiers (such as IDH1/2, KRAS, APC, TP53, STAT1/3, YAP1, CTCF), modulators (SMARCA4, PBRM1, ARID1A, ARID2, ARID1B, DNMT3A, TET2, MLL1/2/3, NSD1/2, SETD2, EZH2, BRD4), and mediators (including OCT4, NANOG, LIN28, SOX2, KLF4) (Feiberg, A.P. Nat. Rev Genet. 17(5):284-99, 2016).

“DNA binding domain” (or DBD) as used herein, refers to the part of the DNA binding protein that is required for binding of the DNA binding protein to cellular DNA or chromatin, or which is required for the localisation of the DNA binding protein to cellular DNA or chromatin, in each case either directly or indirectly. When we refer herein to the part of the fusion protein of the invention as comprising a DNA binding domain from a DNA binding protein said part of the fusion protein could be the full-length DNA binding protein or a truncated portion thereof. However, said part must comprise the DNA binding domain region because this ensures that the endonuclease part that is fused thereto is directed to DNA and able to cleave said DNA.

“DNA endonuclease” as used herein refers to a protein capable of inducing double or single strand DNA breaks, typically by cleaving the phosphodiester bond within a polynucleotide chain. Some, such as deoxyribonuclease I cut DNA relatively non-specifically (without regard to sequence), while others, typically called restriction endonucleases or restriction enzymes, cleave only at very specific nucleotide sequences. A “functionally active DNA endonuclease” is a polypeptide fragment of the full-length DNA endonuclease protein that retains the nuclease activity (i.e. comprises the nuclease domain). Suitably, said DNA endonuclease polypeptide lacks its endogenous DNA binding domain.

Exemplary DNA endonucleases include the LAGLIDADG, HNH, His-Cys box, and GIY-YIG families. Suitable examples of the GIY-YIG family include, T4 thymidylate synthase group I intron-encoded GIY-YIG endonuclease (I-TevI) and endonuclease encoded by the group I intron that interrupts the thymidylate synthase (TS) gene (thyA) of *Bacillus mojavensis* s87-18 (I-Bmol) (Guha and Edgell, *Int. J. Mol. Sci.* 18(12): 2565, 2017; Edgell, D.R., *Curr Biol.* 19 (3): R115–R117, 2009; Edgell and Shub, *Proc Natl Acad Sci. USA.* 98(14):7898-903, 2001).

Examples of suitable DNA endonucleases include: MegaTeV (Wolfs, J.M. et al. *Nucleic Acids Res.* 42(13):8816-29, 2014), I-TevI (Scharenberg, A.M. et al. *Curr. Gene Ther.* 13(4):291-303, 2013) and I-Bmol (Kleinstiver, B.P. *Nucleic Acids Res.* 41(10):5413-27, 2013).

MegaTeV (TeV169-onu), is a fusion of two homing endonuclease active sites in to a single polypeptide that makes two double strand breaks, 30nts apart (Wolfs JM et al *Nucleic Acids Research*, 2014, Vol. 42, No. 13; 8816–8829).

“Fusion protein” as used herein refers to any protein consisting of at least two polypeptide sequences that are normally encoded by separate genes that have been joined so that they are transcribed and translated as a single unit, producing a single polypeptide.

“Pioneer Factors” as used herein refers to any protein that performs a genetic function early in the activation of transcription and physically binds to the genome for a period prior to activation, and prior to other factors binding, and is able to bind target sites in condensed chromatin.

Exemplary pioneer factors include FOXA1, FOXA2, FOXA3, FOXD3, GATA factors, PU.1 (Zaret, K.S. and Carroll, J.S. *Genes Dev.* 25(21):2227-41, 2011), Zelda (Sun, Y. et al. *Genome Res.* 25: 1703-1714, 2015), Pou5f3 (Iwafuchi-Doi, M. *Genes & Dev.* 28: 2679-2692, 2014), Group B1 Sox and Sox2 (Kamachi, Y. and Kondoh, H. *Development* 140:4129-4144,

2013), Oct3/4, Klf4, Ascl1, Pax7, p53 (Iwafuchi-Doi, M. Wiley Interdiscip Rev Syst Biol Med. 11(1):e1427, 2019), and CLOCK:BMAL1 (Menet, J.S. Genes & Dev. 28:8-13, 2014).

“Host cell” means a eukaryotic or prokaryotic cell, including one that expresses the chimeric protein of interest.

Exemplary host cells useful for identifying inhibitors of DNA binding proteins involved in the pathogenesis of cancer include: 786-0, 22RV1, A498, A549, ACHN, BT474, CAKI-1, CCRF-CEM, COLO 205, DU-145, EKVX, HCC-2998, HCT-116, HCT-15, HL-60, HOP-62, HOP-92, HS 578T, HT29, IGR-OV1, K-562, KM12, LOX IMVI, M14, MALME-3M, MCF7, MDA-MB-231, MDA-MB-435, MDA-MB-468, MDA-N, MOLT-4, NCI/ADR-RES, NCI-H226, NCI-H23, NCI-H322M, NCI-H460, NCI-H522, OVCAR-3, OVCAR-4, OVCAR-5, OVCAR-8, PC-3, RPMI-8226, RXF 393, SF-268, SF-295, SF-539, SK-MEL-2, SK-MEL-28, SK-MEL-5, SK-OV-3, SN12C, SNB-19, SNB-75, SR, SW-620, TK-10, U251, U2OS, UACC-257, UACC-62, UO-31, BT-549, T-47D, LXFL 529, DMS 114, SHP-77, DLD-1, KM20L2, SNB-78, XF 498, RPMI-7951, M19-MEL, MEF, RXF-631, SN12K1, P388, and P388/ADR. Most of these cells are in the NCI-60 cell panel. They can be sourced from commercial suppliers, including: ATCC, ECACC, DSMZ.

Exemplary host cell useful for identifying inhibitors of DNA binding proteins involved in other disease settings include immortalized primary cells such as airway endothelial and epithelial cells, aortic endothelial cells, NTAP Schwann cells, Barrett's oesophageal epithelial cells, chondrocyte fibroblast cells, respiratory cells, dermal microvascular endothelial cells, endometrial fibroblast cells, foreskin keratinocytes, mammary epithelial cells, adipose-derived mesenchymal stem cells, pancreas duct cells, renal epithelial cells, prostate, retinal pigmented epithelial cells, and skin fibroblast cells. They can be sourced from commercial suppliers, including: ATCC, ECACC, DSMZ.

“Transcription factor” (or TF) as used herein is a protein involved in the process of converting, or transcribing, DNA into RNA whereby the protein has a DNA binding domain that gives it the ability to bind to specific sequences of DNA.

According to a first aspect of the invention there is provided a method of screening for an inhibitor of a DNA binding protein comprising:

- a) bringing into contact (i) a host cell that expresses an inducible fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide; and, (ii) a test compound, under conditions where the fusion protein, in the absence of the test inhibitor compound, is capable of binding to and cleaving host cellular DNA; and

- b) determining whether the test compound reduces the ability of the fusion protein to cleave cellular DNA,

wherein if the test compound inhibits the ability of the fusion protein to cleave cellular DNA, the test compound is a putative inhibitor of the DNA binding protein.

According to a variation of this aspect of the invention there is provided a method of screening for an inhibitor of a DNA binding protein comprising:

- a) bringing into contact (i) a host cell that expresses an inducible fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide; and, (ii) a test inhibitor compound, under conditions where the fusion protein, in the absence of the test inhibitor compound, is capable of binding to and cleaving host cellular DNA with the effect of reducing or eliminating the viability or proliferation of the cell; and
- b) determining cell number, cellular DNA damage, cell viability, cell growth, cell cycle phase distribution or cell proliferation.

wherein the amount of cellular DNA damage, cell viability, cell growth, cell cycle phase distribution, cell proliferation or number of cells determined indicates whether or not the test compound is an inhibitor of the DNA binding protein

In one embodiment, the inhibitor is one that interferes with the DNA binding of a DNA binding protein, either directly or indirectly, in other words, the ability of the DNA binding protein to bind DNA such as genomic DNA, or chromatin, or which interferes with the localisation of the DNA binding protein to genomic DNA or chromatin.

In one embodiment, the inhibitor is one that directly interferes with the DNA binding of a DNA binding protein.

In one embodiment, the inhibitor is one that indirectly interferes with the DNA binding of a DNA binding protein for example, by inhibiting the dimerization domain of dimeric DNA binding protein, inhibiting the required coactivators or mediators of DNA binding proteins, or inhibiting the ability of the DNA binding protein to bind its specific cognate DNA binding sequence via binding to the DNA itself (see Berg, T. Curr. Opin. Chem. Biol. 12(4):464–47, 2008; Majmudar, C.Y. and Mapp, A.K. Curr. Opin. Chem. Biol. 9(5):467–474, 2005; Arndt, H-D. Angew. Chem. Int. Ed. 45(28):4552–4560, 2006; Koehler, A.N. Curr. Opin. Chem. Biol. 14(3):331–340, 2010; Gniazdowski, M. et. al. Curr. Med. Chem. 10(11):909–924, 2003; and Gniazdowski, M. et. al. Expert Opin. Ther. Targets. 9(3):471–489, 2005).

In one embodiment, the inhibitor is one that inhibits pre-processing of a DNA binding protein. Examples may include the inhibition of post-translational modification of the DNA binding

protein, inhibition of the nuclear localisation of the DNA binding protein, inhibition of correct protein folding, or other protein-protein interactions of the of the DNA binding protein, in each case as required for the DNA or chromatin binding of the DNA binding protein.

Test inhibitor molecules that are more likely to indirectly cause inhibition of DNA binding of a DNA binding protein (e.g. transcription factor) are the nucleic-acid containing molecules (such as siRNA, ASO, gRNA) which could, for example inactivate a protein that is necessary to allow the DNA binding protein to bind to DNA. The assay of the present invention can therefore be used to identify novel targets that are required for DNA binding of a DNA binding protein, e.g. transcription factor, of interest. If blocking binding of the DNA binding protein is useful for therapeutic purposes, then the assay can be used to identify novel druggable targets, targeting of which could indirectly block DNA binding of the DNA binding protein. The assay therefore facilitates the identification of novel druggable targets and thus inhibitors of these targets that can indirectly inhibit DNA binding of the DNA binding protein of interest.

The indicator of whether a test compound inhibits the ability of the fusion protein to cleave cellular DNA can be cell viability, cell growth, cell number, cell cycle phase distribution, cell proliferation or cellular DNA damage. Suitably, the indicator is cell number, wherein if the test compound is an inhibitor of the DNA binding protein it results in an increase in total number of host cells compared to the number of cells grown in the absence of the test compound.

In particular embodiments, ability of the test compound to inhibit or reduce the ability of the fusion protein to cleave cellular DNA is gauged/determined by comparison to suitable reference or control values or control cells.

By way of example, the ability of the test compound to inhibit/reduce the ability of the fusion protein to cleave cellular DNA determined in step (b) is or can be compared to:

- (i) a reference cell viability, cell growth, cell number, cell cycle phase distribution, cell proliferation or cellular DNA damage;
- (ii) cell viability, cell growth, cell number, cell cycle phase distribution, cell proliferation or cellular DNA damage when in the absence of the test inhibitor compound;
- (iii) cell viability, cell growth, cell number, cell cycle phase distribution, cell proliferation or cellular DNA damage when in the presence of a control compound which is known to not inhibit the DNA binding protein;

- (iv) cell viability, cell growth, cell number, cell cycle phase distribution, cell proliferation or cellular DNA damage where the nuclease is not activated; or
- (v) cell viability, cell growth, cell number, cell cycle phase distribution, cell proliferation or cellular DNA damage where the DNA binding domain of said fusion protein has been mutated such that it cannot bind DNA,

An increase in cell number or cell viability, or reduction in cell cycle arrest or reduction in DNA damage in the presence of the test inhibitor compound compared to any of (i) – (v) In the absence of the test compound is indicative that the test inhibitor compound is an inhibitor or putative inhibitor of the DNA binding protein.

With option (iv) above, the nuclease can be defective or non-existent. In that way binding of the DNA binding protein to its target does not trigger cell killing or cell cycle arrest.

With option (v) above, the DNA binding protein cannot bind to its target and consequently the nuclease has a reduced ability to cleave DNA, as compared to a fusion protein wherein the DNA binding domain is functional, and therefore option (v) would have a reduced level of cell killing compared to a fusion protein with a functional DNA protein domain.

The reference values could be derived from cells that express endonuclease polypeptide alone that cannot bind to DNA, or fusion protein bearing cells wherein the fusion protein is not expressed by virtue of the absence of the inducer, or wherein the DNA binding domain of the fusion protein has been mutated or deleted thereby rendering it incapable of binding to its putative sites on the cellular DNA.

Aside from actual cell numbers or viability, e.g. cell survival determined by manual cell number counts, automated cell number counts with imaging software, or Trypan blue exclusion assays, surrogate markers of cell viability can be used to gauge cell viability. For example, in one embodiment, cell viability, proliferation, or cell cycle arrest is assessed by measuring adenosine triphosphate (ATP) consumption, wherein an increase in ATP consumption when in contact with the test inhibitor indicates that the test compound is an inhibitor (or putative inhibitor) of the DNA binding protein, since this is indicative of an increase in viability, cell number, or lack of cell cycle arrest; when cells lose membrane integrity, they lose the ability to synthesize ATP and endogenous ATPases rapidly deplete any remaining ATP from the cytoplasm. This approach is used in the CellTitre-Glo assay; the cells are the source for ATP in the luciferase reaction hence the luminescence produced is directly proportional to the number of viable cells.

In another embodiment, the effect of the host cells (e.g. total cell number or cell viability) is assessed by measuring cell confluence, wherein an increase in cell confluence in the

presence of the test inhibitor indicates that the test compound is an inhibitor (or putative inhibitor) of the DNA binding protein. Many other methods for measuring cell viability are known in the art (Sittampalam GS et al. The Assay Guidance Manual (2018), Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences).

The effect that the test compound has on the cell can also be assessed by determining the amount of cellular DNA damage that arises. The amount of cellular DNA damage can be gauged by measuring for one or more markers of cellular DNA damage including; gammaH2AX or RAD 51 foci produced by the cells following DNA damage. H2AX is a member of the histone H2A family and it has been established that elevated phosphorylation levels of H2AX on genomic DNA occur following DNA damage. The detection of γ H2AX protein phosphorylated at Serine-139 therefore allows the detection and quantification of DNA damage. Methods for detecting γ H2AX include using immunofluorescence, or flow cytometry in conjunction with secondary antibodies conjugated with fluorescein isothiocyanate (FITC) (Figuroa-González G and Pérez-Plasencia C, Oncol Lett. 2017 Jun; 13(6): 3982–3988). RAD51 recruitment to DNA is a marker of DNA break repair via homology-based mechanisms. Such mechanisms involve nuclease-dependent DNA end resection, which generates long tracts of single-stranded DNA required for checkpoint activation and loading of homologous recombination proteins including Rad 51. Measurement of Rad 51 loci is therefore a marker of such repair occurring at sites of DNA damage and can be used to quantify DNA damage. Many other methods are known in the art for measuring DNA damage in cells (e.g. Figuroa-González G and Pérez-Plasencia C, Oncol Lett. 2017 Jun; 13(6): 3982–3988).

In particular embodiments, the amount of gammaH2AX or RAD 51 foci produced by the cells is determined wherein a decrease in gammaH2AX or RAD51 foci production in the cells indicates that the compound is an inhibitor of the DNA binding protein.

Inhibitor Compound

The term “inhibitor” as used herein, refers to an entity/agent whose presence in a system in which an activity of interest is observed correlates with a decrease in level and/or nature of that activity as compared with that observed under otherwise comparable conditions when the inhibitor is absent. In some embodiments, an inhibitor interacts directly with a target whose activity is of interest. In some embodiments, an inhibitor affects level of a target of interest; alternatively, or additionally, in some embodiments, an inhibitor affects activity of a target of interest without affecting level of the target. In some embodiments, an inhibitor affects both level and activity of a target entity of interest, so that an observed difference in

activity is not entirely explained by or commensurate with an observed difference in level. The inhibitor can be any agent, e.g. small molecule compound, nucleic acid, antibody, and the like. The target can be a protein or a precursor thereof, or nucleic acid encoding said protein/precursor, e.g. genomic DNA or mRNA.

Inhibitors can be for example compounds that partially or totally block, decrease, prevent, delay, inactivate, desensitize or down regulate the function or effect of the DNA binding protein. The assay design provides that in the absence of an inhibitor of the DNA binding protein of interest, cell viability is decreased or cell cycle arrest occurs due to an increase in DNA damage imparted by an endonuclease. A decrease in cell viability will be reflected in the total number of cells in the culture after a set period of time and under equivalent conditions. The build-up of DNA damage, in particular double-strand breaks, is detrimental to cell viability, and sometimes lethal. Candidate compounds may thus be tested for their ability to inhibit binding of the DNA binding protein which restricts or prevents DNA damage caused by the endonuclease. Enhanced cell viability or lack of cell cycle arrest is thus a gain of function determinant. A gain of function assay reduces the number and type of false negative "hits" that arise with loss of function assays. The test compound can be any type of molecule, such as an organic or inorganic small molecule, a natural or derivatised carbohydrate, a protein, a polypeptide, a peptide, a glycoprotein, a nucleic acid, a DNA, a RNA, an oligonucleotide, a guide RNA, or a protein-nucleic acid (PNA). In another embodiment, the potential inhibitor compound is obtained from a library of compounds, such as small molecules with drug like properties.

A "small molecule" as used herein, is an organic molecule that is less than about 2 kilodaltons (KDa) in mass. In some embodiments, the small molecule is less than about 1.5 KDa, or less than about 1 KDa. Most small molecule compounds are less than about 800 daltons (Da). Accordingly, in some embodiments, the small molecule is less than about 800Da, less than about 600 Da, less than about 500 Da, less than about 400 Da, less than about 300 Da, less than about 200 Da, or less than about 100 Da. Often, a small molecule has a mass of at least 50 Da. In some embodiments, a small molecule is non-polymeric. In some embodiments, a small molecule contains multiple carbon-carbon bonds and can comprise one or more heteroatoms and/ or one or more functional groups important for structural interaction with proteins (e.g., hydrogen bonding), e.g., an amine, carbonyl, hydroxyl, or carboxyl group, and in some embodiments at least two functional groups. Small molecules often comprise one or more cyclic carbon or heterocyclic structures and/or aromatic or polyaromatic structures, optionally substituted with one or more of the above functional groups.

Unless otherwise stated, as used herein, the terms “about” or “approximately” when used in conjunction with a stated numerical value or range denotes somewhat more or somewhat less than the stated value or range, to within a range of $\pm 15\%$ of that stated, $\pm 10\%$ of that stated, $\pm 5\%$ of that stated in different embodiments.

According to another aspect of the invention there is provided an inhibitor of a DNA binding protein identified in accordance with the first aspect of the invention.

In a particular embodiment, the test compound and/or putative inhibitor is a small molecule compound, a nucleic acid containing molecule such as siRNA or antisense oligonucleotide (ASO), or a peptide.

Suitable nucleic acid containing molecules include short interfering RNA, messenger RNA, short hairpin RNA, micro RNA, single guide RNA or an antisense oligonucleotide. Such a molecule may become a nucleic acid based therapeutic. Suitable peptides may include stapled peptides, peptide mimics, peptide aptamers or peptide conjugates.

In particular embodiment, the nucleic acid containing molecule is an antisense oligonucleotide, a guide RNA (gRNA) or an RNAi molecule, such as siRNA.

Host cell

The assay of the invention is carried out in a live cell. The cell that is used in accordance with the present invention is one that expresses the fusion protein of the DNA binding protein and functional DNA endonuclease. Such cell(s) can be a prokaryotic cell, such as a bacterial cell, or a eukaryotic cell, such as fungi (including yeast), plant and mammalian cells. In a particular embodiment, the cells for use in accordance with the present invention are eukaryotic cells. Cells that are particularly useful in the context of the present invention are mammalian cells, including immortalized mammalian cells or cell lines.

Immortal cell lines can be those isolated from naturally occurring cancers (e.g. HeLa, U2-OS) or artificially made by the introduction of either viral genes such as adenovirus type 5E1 gene for the generation of HEK293, or by the artificial expression of key proteins such as hTERT that imparts immortality.

In particular embodiments, human cells of one of a variety of cell types (e.g. breast, prostate, bone, blood, liver, pancreas, lung, neural, colon etc.) are used. Suitably, the cell(s) are diseased human cells or cell lines, such as cancer cells.

A "eukaryotic cell" which can be used in the assay can be a transfected or transformed cell, provided that said eukaryotic cell is capable of expressing a fusion protein comprising the DNA-binding domain of a DNA binding protein and a functional endonuclease polypeptide.

Eukaryotic cells may be cultured cells, explants, cells *in vivo*, and the like. Eukaryotic cells may include yeast, insect, amphibian, or mammalian cells and the like. Preferably, eukaryotic cells used in the assay are mammalian cell lines, such as HaCaT, HeLa, MDCK, U-2 OS, CHO-K1, or primary cells. The cells may also be immortalized cell lines. The cells may be cultured under standard conditions, which will be determined easily by one skilled in the art.

There are numerous repositories of cells and cell lines that can be utilised in accordance with the first aspect of the invention (e.g. American Type Culture Collection – ATCC, or Coriell Institute).

Exemplary host cells useful for identifying inhibitors of DNA binding proteins involved in the pathogenesis of cancer include: 786-0, 22RV1, A498, A549, ACHN, BT474, CAKI-1, CCRF-CEM, COLO 205, DU-145, EKVX, HCC-2998, HCT-116, HCT-15, HL-60, HOP-62, HOP-92, HS 578T, HT29, IGR-OV1, K-562, KM12, LOX IMVI, M14, MALME-3M, MCF7, MDA-MB-231, MDA-MB-435, MDA-MB-468, MDA-N, MOLT-4, NCI/ADR-RES, NCI-H226, NCI-H23, NCI-H322M, NCI-H460, NCI-H522, OVCAR-3, OVCAR-4, OVCAR-5, OVCAR-8, PC-3, RPMI-8226, RXF 393, SF-268, SF-295, SF-539, SK-MEL-2, SK-MEL-28, SK-MEL-5, SK-OV-3, SN12C, SNB-19, SNB-75, SR, SW-620, TK-10, U251, U2OS, UACC-257, UACC-62, UO-31, BT-549, T-47D, LXFL 529, DMS 114, SHP-77, DLD-1, KM20L2, SNB-78, XF 498, RPMI-7951, M19-MEL, MEF, RXF-631, SN12K1, P388, and P388/ADR.

Exemplary host cell useful for identifying inhibitors of DNA binding proteins involved in other disease settings include immortalized primary cells such as airway endothelial and epithelial cells, aortic endothelial cells, NTAP Schwann cells, Barrett's Oesophageal epithelial cells, chondrocyte fibroblast cells, respiratory cells, dermal microvascular endothelial cells, endometrial fibroblast cells, foreskin keratinocytes, mammary epithelial cells, adipose-derived mesenchymal stem cells, pancreas duct cells, renal epithelial cells, prostate, retinal pigmented epithelial cells, and skin fibroblast cells.

The cells are grown in appropriate culture medium and the test compound is brought into contact with the cells and the amount of cellular DNA damage or cell viability is determined.

According to a third aspect of the invention there is provided a host cell suitable for use in the method of the first aspect of the invention; i.e. one that expresses a fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide.

Controls and validation

In order to minimise the number of false positive and false negative results it will be desirable to incorporate one or more control tests. Such control tests can be run in parallel with the test of the test compound. In one embodiment, the control is a compound which is known to inhibit the particular DNA binding protein. In one embodiment, the control is a compound which is known to not be an inhibitor of the particular DNA binding protein.

Other controls that can be incorporated into the assay include those that verify that the fusion protein has been expressed, and or is present; and/or that the endonuclease is functioning in the sense that it is cleaving the DNA it is expected to cleave.

It will be appreciated, that on occasion, a false positive identification of a putative inhibitor of the DNA binding protein might arise. For example, this might arise if, for example, cell viability alone was used as the determinant for a putative inhibitor of the DNA binding protein and the test compound inhibited the nuclease itself. In such circumstance, cell viability might be enhanced (because the endonuclease was not functioning properly) giving the impression that the test compound was an inhibitor of the DNA binding protein.

To ensure that the test compound is an actual inhibitor of the DNA binding protein further tests can be performed to validate that the test compound is or is not an inhibitor of the DNA binding protein and/or is not an inhibitor of the endonuclease. For example, cells expressing the endogenous protein of interest without the fusion protein may be treated with the test compound and levels of an appropriate biomarker for DNA binding of the protein of interest measured in treated cells and compared to a control. Such biomarkers may include appropriate gene expression, protein levels, histone modification, protein localisation, or cell viability.

Thus, in accordance with another embodiment, a test compound identified as a putative inhibitor of the DNA binding protein is further tested to confirm that it is not an inhibitor of the functionally active DNA endonuclease polypeptide. For example, such an inhibitor would be expected to inhibit the loss of cell viability or cell cycle arrest regardless of the DNA binding domain fused to the DNA endonuclease polypeptide, or of a full-length DNA endonuclease protein.

DNA binding protein

The assay of the invention can be employed to identify inhibitors or putative inhibitors of any DNA binding protein. Exemplary DNA binding proteins are known in the art including human transcription factors (Lambert, S.A. et al. *Cell*. 172 (4): 650 – 665, 2018) including Myc, members of the forkhead (including FOXF3, FOXF2, FOXA1, FOXA2, FOXA3, FOXO3a, FOXM1, and FOX3D) and nuclear and orphan nuclear receptor (including androgen receptor, oestrogen receptor, hepatocyte nuclear factor 4alpha, peroxisome proliferator-activated receptors, nuclear receptor subfamily 1 group D, retinoic acid receptor-related orphan receptors) families (Lambert, S.A. et al. *Cell*. 172 (4): 650 – 665, 2018), and chromatin remodellers (including BRG1, SNF2H, CHD3, CHD4, BRM, INO80, SNF2L and CHD1) (Giles, K.A. et al. *Epigenetics Chromatin*. 12(1):12, 2019) and epigenetic modifiers (such as IDH1/2, KRAS, APC, TP53, STAT1/3, YAP1, CTCF), modulators (SMARCA4, PBRM1, ARID1A, ARID2, ARID1B, DNMT3A, TET2, MLL1/2/3, NSD1/2, SETD2, EZH2, BRD4), and mediators (including OCT4, NANOG, LIN28, SOX2, KLF4) (Feiberg, A.P. *Nat. Rev Genet*. 17(5):284-99, 2016)

Transcription factors are proteins that regulate transcription. Each transcription factor typically binds to a specific sequence of DNA near the promoters of certain genes. Following binding, the transcription of the gene is typically, either switched on or off.

In one embodiment, the DNA binding protein is a transcription factor implicated in a disease. In a particular embodiment, the DNA binding protein is a transcription factor implicated in cancer, such as one selected from the group consisting of: Pterin-4 Alpha-Carbinolamine Dehydratase 2 (TCF1), Groucho, estrogen receptor alpha (ER α), estrogen receptor beta (ER β), Aryl Hydrocarbon Receptor (AHR), Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT), Retinoic Acid Receptor (RAR), Retinoid X Receptor (RXR), Jun Proto-Oncogene, AP-1 Transcription Factor Subunit (JUN), Fos Proto-Oncogene, AP-1 Transcription Factor Subunit (FOS), Activating Transcription Factor 2 (ATF2), ETS Domain-Containing Protein Elk-1 (ELK1), DNA Damage Inducible Transcript 3 (DDIT3 / GADD153), ABL Proto-Oncogene 1, Non-Receptor Tyrosine Kinase (c-ABL), Nuclear Factor Kappa B Subunit 1 (NF κ B), Inhibitor Of Nuclear Factor Kappa B Kinase Subunit Beta (I κ B), RB1 (RB Transcriptional Corepressor 1), E2F Transcription Factor 1 (E2F), TATA-Box Binding Protein (TBP), Tumor Protein P53 (TP53) and P21, BAX, FAS, APO1, BAD, BCL2, GADD45 (Nebert, D.W. *Toxicology Volumes* 181–182, 27 December 2002, Pages 131-141), androgen receptor (AR; Mainwaring, W.I. *Curr Top Mol Endocrinol*. 4:152-71, 1976), Rho GTPase Activating Protein 35 (ARHGAP35; Lawrence, M.S. et. al. *Nature*. 505(7484): 495–501, 2014), AT-Rich Interaction Domain 5B (ARID5B; Leong, W.Z. *Genes Dev*. 31(23-24): 2343–2360, 2017), Achaete-Scute Family BHLH Transcription Factor 1 (ASCL1), ASH1 Like Histone Lysine Methyltransferase (ASH1L; Zhu, L. et. al. *Cancer Discov*. 6(7):770-83, 2016), Activating Transcription Factor 1 (ATF1; Hao, Q. et. al. *Med Sci Monit Basic Res*. 23: 304–

312, 2017), BCL11A, BAF Complex Component (BCL11A; Khaled, W.T. Nat Commun. 6:5987, 2015), BCL6, Transcription Repressor (BCL6 Cardenas, M. G. Clin Cancer Res. 23(4):885–893, 2017), BCL2 Associated Transcription Factor 1 (BCLAF1; Dell'Aversana, C et al. Leukemia. 31(11): 2315–2325, 2017), Basonuclin 2 (BNC2; Cesaratto, L. et. al. Cell Death Dis. 7(9): e2374, 2016), POU Class 3 Homeobox 2 (BRN2; Chen, H-Y. J Exp Clin Cancer Res. 37:161, 2018), CCAAT Enhancer Binding Protein Alpha (C/EBPalpha, Life Sci Alliance. 2(1): e201800173, 2019 Feb) or CCAAT Enhancer Binding Protein Beta (C/EBPbeta; Gardiner, J.D. Oncotarget. 8(16): 26013–26026, 2017), Core-Binding Factor Subunit Beta (CBFB, Am J Hematol. 92(6): 520–528, 2017), Capicua Transcriptional Repressor (CIC, Chittaranjan, S. Oncotarget. 5(17): 7960–7979, 2014), Clock Circadian Regulator (CLOCK; Fu, L. and Kettner, N.M. Prog Mol Biol Transl Sci. 119: 221–282, 2013), CSL (Braune, E-B. et al. Stem Cell Reports. 6(5): 643–651, 2016), CCCTC-Binding Factor (CTCF; Aitken, S. et al. Genome Biol. 19: 106, 2018), E74 Like ETS Transcription Factors (ELF1, ELF3, ELF4 and ELF5; Luk I.Y. et al. Molecules. 23(9): 2191, 2018), Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit (EZH2; Christofides, A. Oncotarget. 7(51): 85624–85640, 2016), Forkhead Box A1 (FOXA1; Ross-Innes, C. S. et al. Nature. 481(7381):389–393, 2012), Forkhead Box A2 (FOXA2; Wang, B. et al. Exp Ther Med. 16(1): 133–140, 2018), Forkhead Box C2 (FOXC2; Gozo, M.C. et al. Oncotarget. 7(42): 68792–68802, 2016), Forkhead Box P1 (FOXP1; Mizunuma, M. Heliyon. 27;2(5):e00116, 2016), Forkhead Box P3 (FOXP3; Wing, J.B et al. Immunity. 19;50(2):302-316, 2019), GATA Binding Protein 1, 2, 3, and 4 (GATA1, GATA2, GATA3, and GATA4; Tremblay, M. et al. Development. 2018 Oct 22;145(20), Hepatocyte Nuclear Factor 4 Alpha (HNF4; Walesky, C. Gene Expr. 16(3):101-8, 2015), Interferon Regulatory Factor 2, 6 and 7 (IRF2, IRF6, and IRF7; Bhatellia, K. Cell Signal. Nov;26(11):2350-7, 2014), Kruppel Like Factor 4 and 6 (KLF4, KLF6; Rane, M.J. EBioMedicine. 40:743-750, 2019), MDS1 And EVI1 Complex Locus (MECOM; Choi, E. J. et al. Pathol Oncol Res. 23(1):145-149, 2017), Myocyte Enhancer Factor 2A (MEF2, Pon J. and Marra, M. Oncotarget. 7(3):2297-312, 2016), MYB Proto-Oncogene, Transcription Factor (MYB; Uttarkar, S. Exp Hematol. 47:31-35, 2017), MYC Proto-Oncogene, BHLH Transcription Factor (MYC; Dang, C. Cell. 149(1):22-35, 2012), MYCN Proto-Oncogene, BHLH Transcription Factor (MYCN; Aygun, N. Curr Pediatr Rev. 14(2):73-90, 2018), Nuclear Factor Of Activated T Cells 4 (NFATC4; Hessmann, E et al. Stem Cells Int. 2016:5272498, 2016), Neurogenin 3 (NGN3; Zhou C et al. Oncotarget. 13;8(33):54388-54401, 2017), NK3 Homeobox 1 (NKX3-1; Donadio, J. et al. Prostate. 79(5):462-467, 2019), Nuclear Factor, Erythroid 2 Like 2 (NR2F2; Wang, H. et al. Biochem Biophys Res Commun. 485(1):181-188, 2017), Nuclear Receptor Subfamily 4 Group A Member 2 (NR4A2; Komiya, T. Transl Lung Cancer Res. 6(5):600-610, 2017), Nuclear Receptor Subfamily 5 Group A Member 1 (NR5A1; Lewis, S. Endocrinology. 155(2):358-69,

2014), Octamer-Binding Protein 4 (OCT4, Villodre, E. et al. *Cancer Treat Rev.* 51:1-9, 2016), Paired Box 5 (PAX5; Shahjahani, M. *Med Oncol.* 32(1):360, 2015), Pancreatic And Duodenal Homeobox 1 (PDX1; Wu, J. *PLoS One.* 12(9):e0184984, 2017), Forkhead Box M1 (FOXM1; Gartel, A. *Cancer Res.* 77(12):3135-3139, 2017), Paired Related Homeobox 1 (PRRX1, Marchand, B. *Oncogene.* 2019 Jan 31. doi: 10.1038/s41388-019-0725-6), Runt Related Transcription Factor 1 (RUNX1; Sood, R. et al. *Blood.* 129(15):2070-2082, 2017), SRY-Box 2, 7, and 9 (SOX2, SOX7, SOX9; Grimm, D. et. al. *Semin Cancer Biol.* 2019 Mar 23. pii: S1044-579X(18)30141-X), TAL BHLH Transcription Factor 1, Erythroid Differentiation Factor (TAL1; Haider, Z. *Cancer Med.* 8(1):311-324, 2019), T-box 3 and 5 (TBX3, TBX5; DeBenedittis P, and Jiao K. *Biochem Biophys Res Commun.* 2011 Sep 9;412(4):513-7) Transcription Factor 12, 4, and 7L2 (TCF12, TCF4, TCF7L2; Quong MW et al. *Annu Rev Immunol.* 2002; 20:301-22. Epub 2001 Oct 4), Transcription Factor Dp-1 (TFDP1; Zhan et al. *Cell Signal.* 230:59-66, 2017), Transcription Factor Dp-2 (TFDP2; Rubin SM et al. *Cell.* 2005 Dec 16;123(6):1093-106), or a methyltransferase (including Dnmt1, Dnmt 3A and Dnmt 3B; Gnyszka, A et al. *Anticancer Res.* 33(8):2989-96, 2013).

In a particular embodiment, the DNA binding protein is a transcription factor implicated in cancer, such as one selected from the group consisting of: Pterin-4 Alpha-Carbinolamine Dehydratase 2 (TCF1), Groucho, estrogen receptor alpha (ER α), estrogen receptor beta (ER β), Aryl Hydrocarbon Receptor (AHR), Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT), Retinoic Acid Receptor (RAR), Retinoid X Receptor (RXR), Jun Proto-Oncogene, AP-1 Transcription Factor Subunit (JUN), Fos Proto-Oncogene, AP-1 Transcription Factor Subunit (FOS), Activating Transcription Factor 2 (ATF2), ETS Domain-Containing Protein Elk-1 (ELK1), DNA Damage Inducible Transcript 3 (DDIT3 / GADD153), ABL Proto-Oncogene 1, Non-Receptor Tyrosine Kinase (c-ABL), Nuclear Factor Kappa B Subunit 1 (NF κ B), Inhibitor Of Nuclear Factor Kappa B Kinase Subunit Beta (I κ B), RB1 (RB Transcriptional Corepressor 1), E2F Transcription Factor 1 (E2F), TATA-Box Binding Protein (TBP), Tumor Protein P53 (TP53) and P21, BAX, FAS, APO1, BAD, BCL2, GADD45, androgen receptor (AR), Rho GTPase Activating Protein 35 (ARHGAP35), AT-Rich Interaction Domain 5B (ARID5B), Achaete-Scute Family BHLH Transcription Factor 1 (ASCL1), ASH1 Like Histone Lysine Methyltransferase (ASH1L), Activating Transcription Factor 1 (ATF1), BCL11A, BAF Complex Component (BCL11A), BCL6 Transcription Repressor (BCL6), BCL2 Associated Transcription Factor 1 (BCLAF1), Basonuclin 2 (BNC2), POU Class 3 Homeobox 2 (BRN2), CCAAT or CCAAT Enhancer Binding Protein Beta (C/EBP β), Core-Binding Factor Subunit Beta (CBFB), Capicua Transcriptional Repressor (CIC), Clock Circadian Regulator (CLOCK), CSL, CCCTC-Binding Factor (CTCF), E74 Like ETS Transcription Factors (ELF1, ELF3, ELF4 and ELF5), Enhancer Of

Zeste 2 Polycomb Repressive Complex 2 Subunit (EZH2), Forkhead Box A1 (FOXA1), Forkhead Box A2 (FOXA2), Forkhead Box C2 (FOXC2), Forkhead Box P1 (FOXP1), Forkhead Box P3 (FOXP3), GATA Binding Protein 1, 2, 3, and 4 (GATA1, GATA2, GATA3, and GATA4), Hepatocyte Nuclear Factor 4 Alpha (HNF4), Interferon Regulatory Factor 2, 6 and 7 (IRF2, IRF6, and IRF7), Kruppel Like Factor 4 and 6 (KLF4, KLF6), MDS1 And EVI1 Complex Locus (MECOM), Myocyte Enhancer Factor 2A (MEF2), MYB Proto-Oncogene, Transcription Factor (MYB), MYC Proto-Oncogene, BHLH Transcription Factor (MYC), MYCN Proto-Oncogene, BHLH Transcription Factor (MYCN), Nuclear Factor Of Activated T Cells 4 (NFATC4), Neurogenin 3 (NGN3), NK3 Homeobox 1 (NKX3-1), Nuclear Factor, Erythroid 2 Like 2 (NR2F2), Nuclear Receptor Subfamily 4 Group A Member 2 (NR4A2), Nuclear Receptor Subfamily 5 Group A Member 1 (NR5A1), Octamer-Binding Protein 4 (OCT4), Paired Box 5 (PAX5), Pancreatic And Duodenal Homeobox 1 (PDX1), Forkhead Box M1 (FOXM1), Paired Related Homeobox 1 (PRRX1), Runt Related Transcription Factor 1 (RUNX1) SRY-Box 2, 7, and 9 (SOX2, SOX7, SOX9), TAL BHLH Transcription Factor 1, Erythroid Differentiation Factor (TAL1) T-box 3 and 5 (TBX3, TBX5) Transcription Factor 12, 4, and 7L2 (TCF12, TCF4, TCF7L2), Transcription Factor Dp-1 (TFDP1), Transcription Factor Dp-2 (TFDP2); or a methyltransferase (including Dnmt1, Dnmt 3A and Dnmt 3B).

In another embodiment, the DNA binding protein is a pioneer factor, such as one selected from the group consisting of: FOXA1, FOXA2, FOXA3, GATA factors, PU.1 (Zaret, K.S. and Carroll, J.S. *Genes Dev.* 25(21):2227-41, 2011), Zelda (Sun, Y. et. al. *Genome Res.* 25: 1703-1714, 2015), Pou5f3 (Iwafuchi-Doi, M. *Genes & Dev.* 28:2679-2692, 2014), Group B1 Sox and Sox2 (Kamachi, Y. and Kondoh, H. *Development.* 140:4129-4144, 2013), Oct3/4, Klf4, Ascl1, Pax7, p53 (Iwafuchi-Doi, M. *Wiley Interdiscip Rev Syst Biol Med.* 11(1):e1427, 2019), and CLOCK:BMAL1 (Menet, J.S. *Genes & Dev.* 28:8-13, 2014).

In another embodiment, the DNA binding protein is a pioneer factor, such as one selected from the group consisting of: FOXA1, FOXA2, FOXA3, GATA1, GATA2, GATA4, PU.1, Zelda, Pou5f3, Group B1 Sox and Sox2, Oct3/4, Klf4, Ascl1, Pax7, p53 and CLOCK:BMAL1.

In one embodiment, the pioneer factor is FOXA1.

In one embodiment, the DNA binding domain is from FOXA1 protein and comprises the sequence disclosed in SEQ ID NO: 10, which has the FOXA1 DBD based on Wang et al., (*Nucl Acid Res.* 46(11):5470-5486, 2018).

Exemplary FOXA1 protein sequences include those found at GenBank accession Nos.: EAW65844.1 and AAH33890.1; in UniProt under accession number: P55317-1; or that

identified as NP_004487.2. The NP_004487.2 sequence is also disclosed herein as SEQ ID NO 10. A splice variant (isoform 2) that lacks amino acids 1-33 is also known. This sequence is disclosed in SEQ ID NO: 16.

The polypeptide region of the fusion protein that represents the DNA binding domain of the DNA binding protein could be the full-length native protein sequence of the DNA binding protein or it could be a truncated portion of the full-length protein that comprises the DNA binding domain. For example, if the DNA binding protein portion of the fusion protein is the full length (or substantially full-length) native protein/polypeptide, it would more closely mimic the full-length protein and how it functions inside the cell. The use of the full-length DNA binding protein is however not required. It may be that the first part of the fusion protein (the polypeptide part comprising the DNA binding domain of the DNA binding protein) lacks one or more amino acids, such as at either terminus. It might be that the first part of the fusion protein only contains the DNA binding domain region. What is required, however, is that the first part of the fusion protein (the polypeptide part comprising the DNA binding domain of the DNA binding protein) must be capable of facilitating DNA binding of the fusion protein. The first part of the fusion protein (the polypeptide part comprising the DNA binding domain of the DNA binding protein) may therefore be the full-length polypeptide sequence of the DNA binding protein or a fragment thereof, which fragment may, compared to the full-length polypeptide sequence lack one or more amino acids, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more amino acids fewer than the full-length (wild-type) DNA binding protein. The DNA binding domain fragment may have a length (measured by number of consecutive amino acids) that is 99%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 50%, 45%, 40%, or 30% the length of the full-length (wild-type) DNA binding protein.

In one embodiment, the DNA-binding domain polypeptide is full-length FOXA1 protein sequence or one that lacks between 1 and 10 amino acids.

DNA binding proteins, including transcription factors have been well-characterised. Utilising conventional sequence alignment methods, the person of skill in the art is able to identify the DNA binding domain region of the DNA binding protein.

Exemplary DBDs of transcription factors that can be utilised as the DBD polypeptide of the fusion protein of the invention are disclosed in SEQ ID Nos: 10, 12, 13, 14, 15, and 16. Examples of encoding nucleotide sequences are in SEQ ID NOs: 7 and 8.

Nuclease

The second polypeptide part of the fusion protein used in the assay of the invention comprises a functionally active DNA endonuclease polypeptide. Thus, the second polypeptide comprises a nuclease domain. The functionally active DNA endonuclease polypeptide will lack its endogenous DNA binding domain. Typically, the nuclease has the ability to cleave double stranded DNA as a monomer. Suitably, the DNA endonuclease polypeptide is derived from a homing endonuclease. In particular embodiments, the DNA endonuclease is derived from a homing nuclease from the LAGLIDADG, HNH, His-Cys box, or GIY-YIG families of nucleases.

DNA endonuclease of the class homing endonucleases are well-known in the field of molecular cloning and genome engineering approaches. Homing endonucleases can generate a single- or double-strand break by recognizing specific DNA target sequence. Homing endonucleases usually recognize DNA target sites ranging from 12 to 45 base pairs in length and they are highly specific in their DNA cleavage function. For the assay described herein (sometimes referred to as endoscreen or endoscreen assay), the endonuclease used could be a member of the LAGLIDADG endonuclease family or HNH endonuclease family or Cis-Box family or the GIY-YIG endonuclease family (see WO2014/121222) (Pingoud et al Nucleic Acids Res., 42 (12) 7489-7527, 2014; Belfort and Bonocora, Meth. Mol. Biol., 1123, 1-26, 2014).

The GIY-YIG homing endonucleases that bind and cleave DNA as monomers are exemplified by, but not limited to, I-TevI (SEQ ID NO: 1 and 2), a DNA double-strand endonuclease encoded by the mobile td intron of T4 phage, I-BmoI (SEQ ID NO: 3) and I-TulI (SEQ ID NO: 4) or type II restriction endonuclease Eco31R (SEQ ID NO: 6). GIY-YIG homing endonucleases require a specific DNA sequence (in the case of I-TevI, the recognition sequence is CNNNG (SEQ ID NO: 22), where N can be any nucleotide) to generate a DSB.

In a particular embodiment, the endonuclease domain part of the fusion protein of the invention is derived from I-TevI.

WO 2014/121222 described chimeric endonucleases that can be used for gene editing. Certain of the chimeric endonucleases disclosed therein can be used or adapted for use in the methods of the present invention.

In one embodiment, the DNA endonuclease polypeptide is capable of creating single-stranded breaks in DNA.

In another embodiment, the DNA endonuclease polypeptide is capable of creating double-stranded breaks in DNA.

In one embodiment, the DNA endonuclease polypeptide is monomeric.

Examples of suitable monomeric nicking endonucleases include MthH and UbaLAI.

A few examples of site-specific endonucleases that cleave double-stranded DNA as monomers are, but are not limited to, MspI, MvaI and BclI.

The endonuclease domain for use in the invention may be derived from the DNA nicking HNH enzyme I-HmuI (SEQ ID NO: 5), which is structurally similar to GIY-YIG enzymes.

The type IIS restriction enzyme, Eco31R (SEQ ID NO:6), which functions as a monomer may also be used in the invention. It contains a C-terminal cleavage domain with an HNH motif, which binds to its DNA recognition sequence (GGTCTC) and makes a double-strand break (Jakubauskas et al. Biochemistry 47, 8546-8556, 2008).

It would be known to the skilled person that there are several other DNA endonucleases that can cause single- or double strand breaks at specific DNA recognition sequence associated with such endonucleases. Endonuclease domains derived from such endonucleases can be adapted in the design of the endoscreen technology (see e.g. Stoddard BL, Q Rev Biophys, 38(1), 49-95 Epub. 2005).

Monomeric endonucleases that cause double strand breaks include MvaI, I-TevI, and I-BmI. When utilising a DNA endonuclease polypeptide that creates a single stranded break it may be necessary to bring the host cell into contact with an effective amount of a compound that can inhibit the ability of the host cell to repair the single strand break; for example, by using a PARP inhibitor molecule. PARP-1 protein binds to single strand breaks, where it is activated to convert NAD⁺ into ADP-ribose polymers (PAR), and recruits XRCC1 to the site of damage. Base excision repair (BER) starts with removal of the damaged base, followed by separate recognition by AP-endonuclease (APE), which makes a single strand break incision. This acts as a substrate for repair involving PARP-1, and PARP-1 inhibited cells are hypersensitive to agents that cause base lesions (Mol Oncol. 2011 Aug;5(4):387-93). Therefore, inhibition of PARP may result in an increased sensitivity to the DNA damage induced by a DNA endonuclease polypeptide that creates a single stranded break.

In order to avoid the need to incorporate an additional agent to impede the cell's ability to repair the single-strand breaks, in certain circumstances it may be preferable to utilise a DNA endonuclease polypeptide that creates double-stranded breaks.

In particular embodiments, the DNA endonuclease polypeptide is selected from the group consisting of: a nuclease from the GIY-YIG family, or family of nucleases that comprises LAGLIDADG motif (SEQ ID NO: 23), or nuclease that comprises the HNH motif and nucleases that comprise the His-Cys box. WO2006/097854 discloses various types of DNA endonucleases that could be utilised in the present invention.

A suitable DNA endonuclease polypeptide is GIY or a functional variant thereof or a functional fragment of either. Merely by way of example, a suitable endonuclease polypeptide sequence may comprise a I-TevI endonuclease domain sequence whose sequence is shown in SEQ ID NO: 1 and 2.

a DNA double-strand endonuclease encoded by the mobile td intron of T4 phage), I-BmI (SEQ ID NO: 3) and I-TuIaI (SEQ ID NO: 4). GIY-YIG homing endonucleases require a specific DNA sequence (in the case of I-TevI, the recognition sequence is CNNNG (SEQ ID NO: 22), where N can be any nucleotide) to generate a DSB.

Alternatively, a suitable nuclease polypeptide sequence may only need a part of the sequence disclosed in SEQ ID NO: 1, such as that disclosed in SEQ ID NO: 2. It is known in the art that the I-TevI nuclease domain can be shortened to residues 1 – 201 and still retain very high levels of activity when fused to DNA binding domain regions (Kleinstiver, B.P. et. al. Proc. Natl. Acad. Sci. USA. 109 (21):8061-66, 2012).

In a particular embodiment, the “functionally active endonuclease” sequence of the fusion protein of the invention consists of or comprises SEQ ID NO: 2.

A functional endonuclease polypeptide of I-TevI consists of conserved sequence that forms the globular GIY-YIG domain characterized by a structurally conserved central three-stranded antiparallel β -sheet, with catalytic residues positioned to use a single metal ion to promote DNA hydrolysis. The GIY-YIG homing endonucleases that bind and cleave DNA as monomers are exemplified by, but not limited to, I-TevI (SEQ ID NO: 1), a DNA double-strand endonuclease encoded by the mobile td intron of T4 phage, I-BmI (SEQ ID NO: 3) and I-TuIaI (SEQ ID NO: 4). In particular embodiments, the nuclease is selected from the group consisting of: I-TevI, the DNA double-strand endonuclease encoded by the mobile td intron of T4 phage, I-BmI and I-TuIaI.

In other particular embodiments, the nuclease comprises a sequence that includes a sequence as disclosed in any of SEQ ID Nos 1 - 6.

A functional variant or fragment is a polypeptide sequence derived from a full-length functional endonuclease that lacks the endogenous DNA binding domain of the endonuclease but is still able to cleave nucleic acid when brought into contact with DNA. Suitably the size (e.g. length) of the functional endonuclease polypeptide is kept as small as

possible such that when fused to the DNA binding domain of the DNA binding protein (such as a transcription factor) it does not interfere (e.g. sterically interfere) with the ability of the fusion protein to bind DNA (i.e. transcription factor function).

Nucleases, such as endonucleases are well-characterised in the art. Utilising conventional sequence alignment methods, the person of skill in the art can identify the DNA binding domain region of the nuclease. This can then be omitted from the polypeptide that makes up the second part of the fusion protein.

In one embodiment, the DNA endonuclease polypeptide is GIY-YIG fragment or a functional variant thereof.

In other embodiments, the DNA endonuclease polypeptide is selected from the group consisting of: I-Bmol endonuclease (Edgell, D.R. et. al. Proc. Natl. Acad. Sci. USA. 98(14):7898-903, 2001); FokI endonuclease (Sugisaki, H. and Kanazawa S. Gene. 16(1-3):73-8, 1981) and I-SceI endonuclease (Plessis, A. et. al. Genetics. 130(3):451-60, 1992), or a functional variant or functional fragment of any thereof.

The full-length amino acid sequences of some exemplary nucleases are disclosed in SEQ ID NOs: 1 to 6. Removal of the endogenous DNA binding domain region would generate a polypeptide that could be utilised in the fusion protein of the invention.

An example of a suitable truncated version of an endonuclease that can be utilised in the fusion protein of the invention is SEQ ID NO: 2.

Fusion Protein

The invention utilises a fusion protein comprising (i) a DNA binding domain polypeptide (B) and (ii) DNA endonuclease polypeptide (D). The B and D portions can be arranged in either order, i.e. B-D or D-B. In one embodiment, the DNA binding domain polypeptide is fused directly or indirectly to the C-terminus of the DNA endonuclease polypeptide. In another embodiment, the DNA binding domain polypeptide is fused directly or indirectly to the N-terminus of the DNA endonuclease polypeptide. The fusion protein may also be referred to as a fusion polypeptide, a chimeric protein or a chimeric polypeptide. As used herein the B and D portions are fused when they are part of a continuous string of amino acids. The fusion protein retains the DNA binding property of the protein that the DNA binding domain (D) is derived from and the nuclease property of the protein that the nuclease domain (D) is derived from.

As used herein, the fusion protein may be identified by reference to the polypeptide components that it possesses. For example, it may be referred to as Endo-DBD fusion protein or Endo-TF (endonuclease fused to transcription factor (TF)), e.g. I-TevI-FOXA1. By directly fused is meant that the two parts are linked directly together, such as by standard peptide (amino acid-amino acid) bonding. By indirectly fused is meant that there is some additional linker moiety between the two parts. In a particular embodiment, there is a linker between the DNA binding polypeptide and the DNA endonuclease polypeptide.

The linker maybe any length, from 1 amino acid to 100 amino acids, however in particular embodiments, the linker is from 2-20 amino acids in length, such as 6 amino acids in length. The linker sequence could be a repeat of amino acid Glycine alone or a mix of Glycine and Serine. A few unlimiting examples are (G)_n, (GS)_n, (GGS)_n, (GGGS)_n, where G is glycine, S is serine and 'n' is the number of times the amino acid sequence in the parenthesis is repeated to give a maximum polypeptide linker length of 100 amino acids.

A suitable linker comprises an amino acid sequence selected from the group consisting of: GGSGGS (SEQ ID NO: 18), GGGS (SEQ ID NO: 19), GGGGS (SEQ ID NO: 20) and GGGGSGGGGS (SEQ ID NO: 21).

The endonuclease domain needs (at minimum) a specific cleavage site sequence (CNNNG), which may limit the number of DNA binding protein-binding sites that are subjected to DNA double strand breaks. The native nuclease linker is believed to make DNA contacts, potentially through specific DNA recognition sequence. The minimal nuclease domain for attachment to the native protein linker is from amino acids 1-97. The endonuclease domain prefers to fuse the DNA binding protein (e.g. transcription factor, TF) sequences at its C-terminal to mimic its native context. If the TF of interest binds as a dimer, then the Endo-TF fusion will potentially generate two DSBs on each side of the TF binding site (if there are the appropriate CNNNG motif). In a particular embodiment, the fusion protein of the invention comprises FOXA1 and I-TevI. Suitably, the FOXA1 portion is full-length FOXA1 protein. In particular embodiments the endonuclease portion is or comprises a I-TevI endonuclease domain sequence, such as one whose sequence is present within SEQ ID NO: 1 and 2 and the DNA binding domain (DBD) region is full-length FOXA1 protein, such as one whose sequence is shown in SEQ ID NO: 10 and 16 or full length or mutant estrogen receptor, such as one whose sequence is shown in SEQ ID NO: 12, 13 or 14. Optionally, the fusion protein has a Flag-Tag sequence, such as the one encoded by the sequence shown in SEQ ID NO: 24.

In one embodiment, the fusion protein comprises I-TevI endonuclease domain – FOXA1 DBD, and optionally a FlagTag sequence.

In one embodiment, the fusion protein comprises I-TevI endonuclease domain – ER DBD, and optionally a FlagTag sequence.

Inducible promoter

The method of the invention relies on inducible expression of the fusion protein, i.e. where the fusion protein is expressed from an inducible promoter. Thus, as used herein the phrase “expresses an inducible fusion protein” means the cell expresses the fusion protein when triggered to express the fusion protein, typically this arises when the cell is contacted with an agent that causes expression of a promoter which is normally inactive, i.e. not constitutively expressed.

Controlled expression of a gene of interest in cells and organisms is widely used in biomedical research. The most popular gene control systems are based on the bacterial regulator protein whose interaction with the operator sequence that is modulated by small drug molecules in a dose-dependent manner (Ramos et al., *Microbiol Mol Biol Rev.*, 69(2), 326-356, 2005). Such DNA-protein interactions mediated through small molecules can be employed as mammalian transcription control systems following the repression-based or activation-based genetic design of the expression module (Weber and Fussenegger, *Curr Opin Biotechnol* 18(5), 399-410, 2007).

The assay system of the present invention relies on detecting inhibition of the endonuclease action on DNA when the binding protein part of the fusion protein binds to DNA. Prior to contacting the cells with the test compound, it is essential to limit the amount of binding to cellular DNA of the fusion protein, otherwise cleavage of the DNA will arise before the test has been started.

One suitable way to achieve this is to ensure that the fusion protein is expressed in the cell from an inducible promoter. In this way, the host cell or cell population is allowed to grow, and the test compound can then be applied at the same time or just after the fusion protein is expressed following activation of the exogenous promoter.

Inducible promoters are well-known in the art (see e.g., Gossen M & Bujard H. *PNAS.*, 89(12):5547-51, 1992) and the person skilled in the art is able to select a suitable inducible promoter for expressing the fusion protein. Examples of suitable inducible promoters include: Tet-ON (Gossen, M. and Bujard, H. *Proc. Natl. Acad. Sci. USA.* 89(12):5547–5551, 1992; Gossen, M. et. al. *Science.* 268(5218):1766–1769, 1995; Kistner, A. et. al. *Proc Natl Acad Sci USA* 93:10933-38, 1996), lac (Liu et. al. *Proc. Natl. Acad.Sci. USA* 86:9951-9955, 1989), Ecdysone (No, D. et. al. *Proc Natl Acad Sci U S A.* 93(8): 3346–3351, 1996), Auxin-inducible (Conner, T.W. et. al. *Plant Mol Biol.* 15(4):623-32, 1990), hsp70-promoter induced (Wu, B.J. et. al. *Proc Natl Acad Sci USA.* 83(3):629-33, 1986), Cumate-system (Mullick, A.

BMC Biotechnol. 6:43, 2006) and Geneswitch (Wang, Y et. al. Proc. Natl. Acad. Sci. USA. 91:8180-8184, 1994).

Thus, in one embodiment, the fusion protein is capable of being expressed from an inducible promoter. In particular embodiments, the inducible promoter is selected from the group consisting of: Tet-ON, lac, Ecdysone, Auxin-inducible, hsp70-promoter induced, Cumate-system and Geneswitch™.

Inducible promoters are only active provided certain circumstances are met, which can vary depending on the inducible promoter used. Inducible promoters can be regulated by positive or negative control. For positive control the promoter is typically inactive because a specific protein required for transcription cannot bind unless it is in the presence of an inducer, which then allows the activator protein to bind the promoter. In contrast, negative control inducible promoter systems rely upon the promoter being in an inactive state as a result of a protein repressing the promoter, which is overcome by addition of an inducer, for example as a result of loss of binding of the repressor protein to the promoter. The selection of the appropriate inducible promoter should consider the background of the cell of interest. For example, a bacterial promoter should be selected for expression in a prokaryote, whereas for the various eukaryotic cell types (mammalian, yeast, plants) promoters specific to the relevant cell type should be selected. It is also important that the inducible promoter is under tight control, and does not express the protein under control erroneously, particularly where the expressed protein is toxic to the host cell. The level of transcriptional expression is also a consideration in selecting the correct inducible promoter, with high levels of induction preferred. Likewise, the time needed to induce the promoter should be considered, with a preference for systems wherein the time required to induce expression is short.

Depending on the inducible promoter used the inducer can vary from a chemical, elevation in temperature, or light. The tetracycline-ON (Tet-On) system is an example of a chemically inducible promoter. Addition of tetracycline (or its analogue doxycycline) to this system induces the promoter by activating the binding of reverse tetracycline-controlled trans-activator to tetracycline response elements in the inducible promoter.

Examples of inducible promoters are shown in the Table below.

Expression System	Promoter/Operator	Regulator/Inducer	Reference/supplier
Tet-On 3G	<i>PTRE3G (7x tetO)</i>	Tet-On 3G / Doxycycline	Gossen and Bujard (2002) Annu Rev Genet. 36:153-73. Clontech (https://www.takarabio.com)

T-Rex	<i>P_{CMV} (2x tetO)</i>	TetR / Tetracycline	Hillen W, Berens C. Annu Rev Microbiol. 48:345-69, 1994. Invitrogen (https://www.thermofisher.com)
Estrogen	<i>P_{Admodmin} Gal4</i>	Gal4BD-ER-LBD-VP16	Brassellmann, S. et al., Proc Natl Acad Sci U S A. 90(5):1657-61, 1993.
Ecdysone	<i>P_{3xSP1-mHSP} / 5x E-GRE</i>	VgRcR / Ponasterone A	No D. et al., Proc Natl Acad Sci U S A. 93(8):3346-51, 1996. Agilent (https://www.agilent.com)
GeneSwitch™	<i>P_{TKmin} 17 × Gal4</i>	Gal4 _{BD} -PR-LBD-p65-AD / Mifepristone	Wang et al., Proc Natl Acad Sci U S A. 91(17):8180-4, 1994. Invitrogen (https://www.thermofisher.com)
LacSwitch II	<i>P_{OPRSVI}</i>	Lac Repressor / IPTG	Labow et al., Mol Cell Biol. 10(7):3343-56, 1990. Stratagene/Agilent (https://www.takarabio.com)
Cumate activation system	<i>P_{CMVmin} CuO</i>	CymR / Cumate	Mullick et al., BMC Biotechnol. 6:43, 2006.

The above examples rely upon the addition of a chemical to induce promoter activation. However, alternative approaches to inducible promoters are also available, for example, the Cre-LoxP approach (Lox-STOP-Lox system with Cre-recombinase, LSL). In this system the gene of interest can be expressed under the control of a constitutive promoter, but a transcription termination sequence (SV40 late polyadenylation signal) is introduced between the promoter and the gene of interest, which in turn is flanked by two LoxP sequences. The presence of SV40-pA prevents the expression of the cytotoxic gene unless the stop signal is removed by recombination between the two LoxP sites mediated by the expression of Cre-recombinase (Nagy A, Genesis, 26(2), 99-109, 2000).

Alternatively, small-molecule displacement of a cryptic degron is known in the art as a mechanism for conditional control of protein degradation (Bonger, K.M. et al. Nat Chem Biol. 7(8):531-7, 2011). In this system the presence of a small molecule ligand of the degron is

used to induce the degradation of the protein, whereas in the absence of the small molecule the fusion protein is stable.

Alternatively, control of the fusion protein level can be achieved using F-box protein mediated regulation. In this system the degradation of the fusion protein can be induced by addition of the auxin indole-3-acetic acid (Nishimura, K. et. al. Nat. Methods. 6, 917–922, 2009).

Nucleic acid Construct

If not expressed from an inducible promoter, the fusion protein could be provided exogenously. However, conveniently the host cell is a recombinant cell that has been engineered to express the fusion protein from nucleic acid introduced into the cell.

The nucleic acid capable of expressing the fusion protein is typically provided in the form of linear DNA or on a plasmid vector or other nucleic acid construct.

In a particular embodiment, a nucleic acid construct capable of facilitating inducible expression of the fusion protein has been inserted into the host cell. By way of example, the nucleic acid construct inserted into the host cell is a plasmid or other episomal vector.

In one embodiment, the nucleic acid construct comprised nucleic acid encoding the fusion protein under the control of a suitable inducible promoter.

The nucleic acid construct may comprise other element such as a marker, a reporter gene/protein or a tag, or other elements that are found and used in expression vectors and the like. The marker could be, for example, a fluorescent marker, a luminescent marker or an antibiotic resistance gene whose purpose is to select cells that have taken up the construct.

A reporter gene/protein is typically the nucleic acid or encoded protein of a detectable marker, such as green fluorescent protein (GFP) or luciferase, which serves to indicate whether a particular gene (e.g. encoding the fusion protein of the invention) is expressed in the host cell. A reporter is usually a fluorescent protein or luminescent.

A tag is typically a peptide sequence which can either be attached to the expressed protein of interest or co-expressed from the same promoter, which can then be used to detect the presence of or quantify the amount of the target protein (e.g. fusion protein of the invention).

The detection of the tag can be done using immunological methods involving antibodies.

For example, an antibody raised against the tag can be used to detect the presence of the tag in the cell or culture medium. A tag can be fluorescent or non-fluorescent like, GFP, RFP, CFP, YFP, His, Myc, V5, FLAG etc. Tags such as FLAG, Myc, V5, MBP should be fused to the protein of interest if they are to be detected immunologically.

In one embodiment, the nucleic acid construct is also capable of expressing a marker and/or reporter protein and/or a tag (e.g. FLAG tag). Certain of these elements may be expressed

from the same inducible promoter as the fusion polypeptide to allow quantitation of the amount of expressed fusion protein. Certain of these elements may be useful as an internal control to verify that the fusion protein has been expressed and may also be used to quantify the amount of expression.

The person skilled in the art is able to select and use marker and/or reporter protein and/or tag sequences (e.g. FLAG tag) as required.

Under certain circumstances a reporter gene/protein may also be a tag.

In particular embodiments, the tag is selected from the group consisting of but not limited to: Cellulose Binding Domain (CBD), Dihydrofolate reductase (DHFR), Calmodulin binding protein (CBP), FLAG, SUMO, S-tag, Glutathione S-Transferase (GST), Hemagglutinin A (HA), Histidine (His), Herpes Simplex Virus (HSV), Maltose-Binding Protein (MBP), c-Myc, Protein A, Protein G, Streptavidin, T7, V5, Vesicular Stomatitis Virus Glycoprotein (VSV-G), Yeast 2-hybrid tags (B42, GAL4, LexA, VP16); fluorescent reporter proteins such as Green Fluorescent Protein (GFP), Red Fluorescent Protein (RFP), mCherry, Cyan Fluorescent Protein (CFP), Blue Fluorescent Protein (BFP), Yellow Fluorescent Protein (YFP); luminescent reporter proteins such as Luciferase, NanoLuc.

In particular embodiments, the reporter gene/protein is selected from the group consisting of but not limited to: a fluorescent reporter protein, such as Green Fluorescent Protein (GFP), Red Fluorescent Protein (RFP), mCherry, Cyan Fluorescent Protein (CFP), Blue Fluorescent Protein (BFP) and Yellow Fluorescent Protein (YFP) and a luminescent reporter protein, such as Luciferase or NanoLuc.

In particular embodiments, the marker is selected from the group consisting of but not limited to: a selectable antibiotic resistance marker, such as Chloramphenicol Acetyl Transferase (CAT), beta lactamase, blasticidin deaminase, neomycin phosphotransferase, hygromycin B phosphotransferase, puromycin-N-acetyltransferase. All such markers are widely available from multiple commercial suppliers and can be used as the native sequence or humanised forms.

In the Examples the inventors have used synthetic humanised monster GFP (hMGFP) sequence from which 96% of the transcription factor binding sites have been removed.

In a particular embodiment this reporter protein sequence is used.

In the examples the inventors have also used FLAG-tags. In a particular embodiment the nucleic acid construct that expresses the fusion protein of the invention also expresses a FLAG-tag.

Transfection

The nucleic acid construct for use in the invention must be introduced into the host cell.

This can be carried out by a number of techniques well known to the person skilled in the art.

In particular embodiments, the nucleic acid construct has been inserted into the cell via transfection, transduction, electroporation or transformation. In particular, introduction of the nucleic acid construct of the invention into a eukaryotic cell may use a viral or a plasmid-based system. The plasmid system may be maintained episomally or may be incorporated into the host cell or into an artificial chromosome. Incorporation may be either by random or targeted integration of one or more copies at single or multiple loci. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation and transduction using bacteriophage.

As used herein, “transfection” is the method of introducing nucleic acid (such as plasmid or linear DNA) into eukaryotic cells. A transfectant is a cell which has had exogenous nucleic acid introduced therein. The introduced nucleic acid can be said to be transiently introduced (or expressed, as appropriate) wherein the nucleic acid resides in the cell and after a number of rounds of division might be lost. Alternatively, it might be stably integrated, wherein the nucleic acid is incorporated into the host cellular nucleic acid (e.g. genomic DNA) and is more permanently integrated and is passed on to future generations of cells. As used herein “transformation” is the method of introducing nucleic acid (such as plasmid or linear DNA) into prokaryotic cells (such as *E.coli*). A transformant is a prokaryotic cell that comprises exogenously introduced DNA.

In order to allow nucleic acid (or indeed other agents such as chemicals or drugs) to be introduced into a cell, the permeability of the cell membrane is typically increased using chemical or physical methods (electroporation, sonoporation, etc.). Chemical methods involve treating the cells with chemicals (such as calcium phosphate or calcium chloride) that permeabilise the cell membrane. Nucleic acid can then more easily enter the cell, where it can then move from the cytoplasm into the nucleus and either reside there or integrate into genomic DNA.

Electroporation is the use of high-voltage electric shocks to introduce DNA into cells. This can be used with most cell types and yields a high frequency of both stable and transient transfectant cells.

Another physical transfection method is sonoporation, which uses sound (typically ultrasonic frequencies) to permeabilise the cell membrane so as to allow introduction of DNA into the cell.

In one embodiment, the nucleic acid construct is transformed or transfected into the host cell. In one embodiment the cell is a transient transfectant. In another embodiment the cell is a stable transfectant.

In one embodiment, the nucleic acid encoding the fusion protein and promoter elements necessary for transcribing the nucleic acid encoding the fusion protein is stably integrated.

The introduction may be followed by causing or allowing expression from the nucleic acid, e.g. by culturing host cells under conditions for expression of the gene.

Cell culture

The transfected cells capable of expressing the fusion protein of the invention can be grown under appropriate conditions (temperature, time etc.) for such cells and in an appropriate culture medium.

Any cell culture medium that supports cell growth and maintenance under the conditions of the invention may be used. Typically, the medium contains water, an osmolality regulator, a buffer, an energy source, amino acids, an inorganic or recombinant iron source, one or more synthetic or recombinant growth factors, vitamins, and cofactors.

Commercially available media such as Ham's F12, (Sigma), Minimal Essential Medium (MEM, Sigma), RPMI-1640 (Sigma), and Dulbecco's Modified Eagle's Medium (DMEM, Sigma) are suitable for culturing the host cells.

One skilled in the art will recognize which cell media, inoculation media, etc. is appropriate to culture a particular cell, e.g., animal cells (e.g., CHO cells). For example, one of skill in the art will be able to suitably select for a particular culture the amount of glucose and other nutrients, such as glutamine, iron, trace elements, and the like, as well as other culture variables, such as, e.g., the amount of foaming, osmolality, etc. (see, e.g., Mather, J. P., et al. (1999) "Culture media, animal cells, large scale production," Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis, and Bioseparation, Vol. 2:777-785.

The culture conditions, such as temperature, pH, and the like, to be used with the host cell selected for expression will be apparent to the person skilled in the art.

Measuring the effects on the host cell (e.g. cell viability)

The fitness of the cells or the viability of the cells in the assay of the present invention can be measured using several different approaches. As the Endo-DBD fusion protein, for example Endo-TF, binds to its recognition sequence on cellular DNA and cause DNA double strand break, the response of cells to such genomic insult could be diverse. Such extensive DNA damage could lead to apoptosis and cell death, in which case assaying for Caspase 3/7 activity could be a good read out using Caspase-Glo. Cell death can also lead to reduced ATP levels in the total cell population which can be measured by assays such as CellTitre-Glo. Based on the genetic background of the cell line of choice and its DNA damage response, the cells in an assay of the present invention could lead to a DNA damage-mediated cell cycle arrest, where by no considerable cell death is observed. In such a scenario, measuring the total ATP levels will not give a meaningful read-out as such arrested

cells will have high levels of ATP. However, assaying for DNA damage biomarkers such as Rad51, γ H2AX or phospho-KAP1 (S824) would give a more meaningful readout.

Changes in DNA damage biomarkers can be measured using immune-assays such as ELISA and immune-fluorescence technologies.

Cell proliferation can be measured using technologies such as IncuCyte® live cell imaging, which will encompass both cell death-mediated and cell cycle-arrested reduction in cell growth (Sittampalam, G.S. et. al, editors. Assay Guidance Manual. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK53196>)

Screening systems

The screen can be conducted using one host cell or a clonal population of host cells with one test compound. However, the system can also be scaled up so that many different test compounds can be assessed simultaneously, such as for high throughput screening. For example, this can be done by running a multitude of individual tests in parallel, for example, using multiwell plates; or, by combining a group of test compounds, such as 3, 5, 10 compounds etc. from a larger panel of compounds (e.g. 1000) and testing these together. Then by altering the combination of compounds from the larger panel in parallel tests it is possible, from all the positive and negative reactions detected, to deduce which individual compound from the original panel (e.g. 1000) is a putative inhibitor. This compound can then be tested individually for its effect in the assay to verify that it is a putative inhibitor of the DNA binding protein. In this way, rather than having to run a reaction for each compound in the panel, for example 1000 parallel test reactions, a much smaller number (e.g. 30-50) of reactions could be carried out. In one embodiment, a multiplex screen of potential inhibitors is carried out (e.g. a screen of at least 100 test compounds/potential inhibitors, at least 1000 potential inhibitors, or at least 10,000 potential inhibitors), with each test compound typically screened by itself. Putative DNA binding protein inhibitor compounds identified in this way can then be tested to verify that the compound is an actual inhibitor of the DNA binding domain, rather than a false positive, such as one that interferes directly with the nuclease being used in the assay.

According to a second aspect of the invention there is provided an inhibitor of a DNA binding protein identified in accordance with the first aspect of the invention.

Pharmaceutical compositions

The method according to the first aspect of the invention will identify compounds that are putative or actual inhibitors of the DNA binding protein. Such molecules, or molecules derived therefore can be used to treat diseases mediated by or associated with binding of the DNA binding protein to its target DNA; such as cells that over express the DNA binding protein. In order to treat a subject (e.g. human or animal) with the DNA binding protein inhibitor compound it is typically necessary to administer the compound in a suitable pharmaceutical composition.

Thus, the DNA binding protein inhibitors will typically be admixed with one or more pharmaceutically acceptable excipients. The choice of excipient can be selected with regard to the intended route of administration and standard pharmaceutical practice.

For some embodiments of the present invention, the pharmaceutical composition will comprise a compound that has been screened by the assay according to the first aspect of the invention as described herein.

According to a fourth aspect of the invention there is provided a pharmaceutical composition comprising an inhibitor of a DNA binding protein in accordance with the second aspect of the invention.

According to a fifth aspect of the invention there is provided method of preparing a pharmaceutical composition comprising an inhibitor of a DNA binding protein comprising: (i) selecting a compound that is an inhibitor of a DNA binding protein according to the first aspect of the invention; and, (ii) admixing the inhibitor with one or more pharmaceutically acceptable excipients.

A DNA binding protein inhibitor identified according to the present invention can be incorporated into pharmaceutical compositions suitable for administration, for example, in accordance with the methods of treatment or medical uses described herein. Such compositions typically comprise the agent/compound and one or more pharmaceutically acceptable excipients. The term "pharmaceutically-acceptable excipient" as used herein means one or more compatible solid or liquid fillers, diluents or encapsulating substances that are suitable for administration into a human. The term "excipient" denotes an organic or inorganic ingredient, natural or synthetic, with which the active ingredient is combined to facilitate the application. Types of suitable excipient are salts, buffering agents, wetting agents, emulsifiers, preservatives, compatible carriers, diluents, carriers, vehicles, supplementary immune potentiating agents such as adjuvants and cytokines that are well known in the art and are available from commercial sources for use in pharmaceutical preparations (Remington: The Science and Practice of Pharmacy with Facts and Comparisons: Drugfacts Plus, 20th Ed. Mack Publishing; Kibbe et al., (2000) Handbook of Pharmaceutical Excipients, 3rd Ed., Pharmaceutical Press; and Ansel et al., (2004)

Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th Ed., Lippencott Williams and Wilkins). Optionally, the pharmaceutical compositions contain one or more other therapeutic agents or compounds. Suitable pharmaceutically acceptable excipients are relatively inert and can facilitate, for example, stabilisation, administration, processing or delivery of the active compound/agent into preparations that are optimised for delivery to the body, and preferably directly to the site of action.

The pharmaceutical compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use.

When administered, the pharmaceutical compositions of the present invention are administered in pharmaceutically acceptable preparations/compositions. Such preparations may routinely contain one or more pharmaceutically acceptable “excipients”.

Administration may be topical, i.e., substance is applied directly where its action is desired, enteral or oral, i.e., substance is given via the digestive tract, parenteral, i.e., substance is given by other routes than the digestive tract such as by injection. Large biologic molecules are typically administered by injection.

Pharmaceutical compositions for parenteral administration (e.g. by injection), include aqueous or non-aqueous, isotonic, pyrogen-free, sterile liquids (e.g. solutions, suspensions), in which the active ingredient is dissolved, suspended, or otherwise provided (e.g. in a liposome or other microparticulate). Such liquids may additionally contain one or more pharmaceutically acceptable carriers, such as anti-oxidants, buffers, stabilisers, preservatives, suspending agents, and solutes that render the formulation isotonic with the blood (or other relevant bodily fluid) of the intended patient. In particular embodiments, the composition may be lyophilised to provide a powdered form that is ready for reconstitution as and when needed. When reconstituted from lyophilised powder the aqueous liquid may be further diluted prior to administration. For example, diluted into an infusion bag containing 0.9% sodium chloride injection, USP, or equivalent, to achieve the desired dose for administration. In particular embodiments, such administration can be via intravenous infusion using an IV apparatus.

In one aspect, the active agent of the invention (e.g. compound identified as an inhibitor of DNA binding protein according to the first aspect of the invention) and optionally another therapeutic or prophylactic agent are formulated in accordance with routine procedures as pharmaceutical compositions adapted for intravenous administration to human beings. Typically, the active agents for intravenous administration are solutions in sterile isotonic

aqueous buffer. Where necessary, the compositions can also include a solubilizing agent. Compositions for intravenous administration can optionally include a local anaesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule. Where the active compound is to be administered by infusion, it can be dispensed, for example, with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the active compound is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients can be mixed prior to administration.

Compositions for oral delivery can be in the form of tablets, lozenges, aqueous or oily suspensions, granules, powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions can contain one or more optional agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavouring agents such as peppermint, oil of wintergreen, or cherry; colouring agents; and preserving agents, to provide a pharmaceutically palatable preparation. A time delay material such as glycerol monostearate or glycerol stearate can also be used. Oral compositions can include standard vehicles such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

Compositions for use in accordance with the present invention can be formulated in conventional manner using one or more physiologically acceptable excipients. Thus, the active agent (DNA binding protein inhibitor) and optionally another therapeutic or prophylactic agent and their physiologically acceptable salts and solvates can be formulated into pharmaceutical compositions for administration by inhalation or insufflation (either through the mouth or the nose) or oral, parenteral or mucosal (such as buccal, vaginal, rectal, sublingual) administration. In one aspect, local or systemic parenteral administration is used.

For oral administration, the compositions can take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets can be coated by methods well known in the art. Liquid preparations for oral administration can take the form of, for example, solutions, syrups or suspensions, or they

can be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations can be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations can also contain buffer salts, flavouring, colouring and sweetening agents as appropriate.

The pharmaceutical compositions of the invention are for administration in an effective amount. An "effective amount" is the amount of a composition that alone, or together with further doses, produces the desired response.

Typically, a physician will determine the actual dosage which will be most suitable for an individual subject. The specific dose level and frequency of dosage for any particular patient may be varied and will depend upon a variety of factors including the activity of the specific compound employed, the metabolic stability and length of action of that compound, the age, body weight, general health, sex, diet, mode and time of administration, rate of excretion, drug combination, the severity of the particular condition, and the individual undergoing therapy. In addition, the appropriate dosage for administration to a nominal patient with a particular disease will likely have been determined following extensive clinical trials by the innovator pharmaceutical company developing the drug.

In certain embodiments, the compound/agent that inhibits the DNA binding protein can be administered as a pharmaceutical composition in which the pharmaceutical composition comprises between 0.1-1mg, 1-10 mg, 10-50mg, 50-100mg, 100-500mg, or 500mg to 5g of the active compound/agent.

In particular embodiments, the DNA binding protein inhibitor will be administered at approximately 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 mg/Kg body weight per dose. Other embodiments comprise the administration of the DNA binding protein inhibitor at about 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900 or 2000 mg/Kg body weight dose. Using the teaching herein, one of skill in the art can determine the effective dose and dosing schedule/regime of the DNA binding protein inhibitor based on preclinical and clinical studies and standard medical and biochemical measurements and techniques.

The pharmaceutical composition could be for veterinary (*i.e.* animal use) or for human usage.

Medical uses

According to a sixth aspect of the invention there is provided the inhibitor of a DNA binding protein identified in accordance with the first aspect of the invention or pharmaceutical composition according to the fourth aspect of the invention for use in therapy, e.g. in a method of treatment of the human or animal body.

According to a seventh aspect of the invention there is provided a method of treatment of an individual in need of treatment with an inhibitor of a DNA binding protein comprising administering to said individual a therapeutically effective amount of a pharmaceutical composition comprising an inhibitor of a DNA binding protein, such as in accordance with the fourth aspect of the invention.

According to another aspect of the invention there is provided a pharmaceutical composition comprising an inhibitor of a DNA binding protein identified in accordance with the first aspect of the invention, for use in therapy, particularly for use in a method of treatment of the human or animal body.

According to another aspect of the invention there is provided an inhibitor of a DNA binding protein identified in accordance with the first aspect of the invention or a pharmaceutical composition comprising said inhibitor, for use in the manufacture of a medicament for use in therapy.

According to another aspect of the invention there is provided a method of treating a subject in need of treatment with an inhibitor of a DNA binding protein comprising identifying an inhibitor or putative inhibitor of a DNA binding protein in accordance with the first aspect of the invention. Optionally formulating the inhibitor by admixing with one or more pharmaceutically acceptable excipients and administering an effective amount of said inhibitor or a pharmaceutical composition comprising said inhibitor to a patient in need thereof.

In a particular embodiment, the DNA binding protein is implicated in cancer and the medical uses as described herein are for use in treating said cancer for example the treatment of breast cancer patients with an inhibitor of the DNA binding protein oestrogen receptor, or treatment of prostate cancer patients with an inhibitor of the DNA binding protein androgen receptor.

Kits

The materials for use in the present invention can be provided in a kit. Such a kit may comprise one or more containers, each with one or more reagents (optionally in concentrated form) utilised in the method according to the first aspect of the invention, including a cell (or a clonal population thereof) that expresses or is capable of expressing a

fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide. The kit may also comprise one or more controls, such as a compound that is capable of inhibiting the DNA binding protein and one that is not.

A set of instructions will also typically be included.

Throughout the description and claims of this specification, the words “comprise” and “contain” and variations of them mean “including but not limited to”, and they are not intended to (and do not) exclude other moieties, additives, components, integers or steps.

Throughout the description and claims of this specification, the singular encompasses the plural unless the context otherwise requires. In particular, where the indefinite article is used, the specification is to be understood as contemplating plurality as well as singularity, unless the context requires otherwise.

Features, integers, characteristics, embodiments described in conjunction with a particular aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example described herein unless incompatible therewith.

The patent, scientific and technical literature referred to herein establish knowledge that was available to those skilled in the art at the time of filing. The entire disclosures of the issued patents or published patent applications, and other publications, including sequence accession numbers, that are cited herein are hereby incorporated by reference to the same extent as if each was specifically and individually indicated to be incorporated by reference. In the case of any inconsistencies, the present disclosure will prevail.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Sambrook *et al.*, ed. (1989) *Molecular Cloning A Laboratory Manual* (2nd ed.; Cold Spring Harbor Laboratory Press); D. N. Glover ed., (1985) *DNA Cloning*, Volumes I and II; and e.g. Ausubel *et al.*, (1989) *Current Protocols in Molecular Biology* (John Wiley and Sons, Baltimore, Md.).

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. For example, Singleton and Sainsbury, *Dictionary of Microbiology and Molecular Biology*, 2d Ed., John Wiley and Sons, NY (1994); and Hale and Marham, *The Harper Collins Dictionary of Biology*, Harper Perennial, NY (1991) provide those of skill in the art with a general dictionary of many of the terms used in the invention. Unless otherwise

indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art.

DESCRIPTION OF THE FIGURES

Figure 1 shows a schematic diagram representing the composition of the 1428bp endo cassette synthesised by GeneArt Gene synthesis (Invitrogen), cloned as a Sall-BamHI restriction fragment in to the GeneArt vector pMK-RQ.

Figure 2 shows a vector map of pTRE3G-Endo response plasmid prepared by sub-cloning the GeneArt synthesised, approximately 1.4kb Endo cassette in to the pTRE3G response vector (Clontech #631168) as a Sall-BamHI fragment.

Figure 3 shows a vector map of pTRE3G-Endo response plasmids into which the transcription factor coding sequences were inserted. Panel A shows Estrogen receptor alpha (approximately 1.8kb). Panel B shows a FOXA1 full length (approximately 1.4kb). Panel C shows FOXA1 with the DNA binding domain (DBD) deleted (approximately 1.3kb), which was used as a negative control. The FOXA1 DNA binding domain deletion mutant had 156bp of coding sequence or 52 amino acids in the DBD deleted, which made it incapable of binding to DNA.

Figure 4 shows a vector map of pCMV-Tet3G regulator plasmid (Clontech #631168) that encodes for the Tet-On 3G transactivator protein.

Figure 5 shows a Western blot analysis of the U2OS cells stably expressing Tet3G transactivator protein. A protein band of expected 30kDa was expressed in the cells that were transfected with the pCMV-Tet3G expression vector and the corresponding band was absent in the parental U2OS cells that were not transfected.

Figure 6 shows Western analyses to confirm the induced expression of the transgene in U2OS/Tet3G cells that are double-stable for Endo alone or Endo-ER. Addition of doxycycline (24hrs) resulted in the expression of a protein band of approximately 29kDa for Endo alone (Endo-FLAG, Fig 6A) and approximately 90kDa for Endo-ER (Endo-ER-FLAG, Fig 6B) that were detected by the FLAG antibody and the corresponding bands were absent in the vehicle control treated cells. Parental U2OS cells were used as negative control. p150/Glued was used as protein loading control.

Figure 7 shows a graph representing the induced expression of Endo-ER, but not Endo alone, in U2OS/Tet3G cells resulted in reduced cell viability. Data plotted are mean of 3 technical replicates $\pm$ standard error of mean.

Figure 8 shows titration of doxycycline to obtain optimal expression of Endo-ER. Increasing concentration of Doxycycline (0 to 100ng/ml) was added to U2OS/Tet3G cells harbouring Endo alone or Endo-ER and cell viability was assessed after 72hrs of incubation using CellTitre-Glo reagent. The doxycycline concentrations tested are marked on the X-axis and the corresponding cell viability is represented as relative luminescence unit (RLU) on the Y-axis. Data plotted are mean of 4 technical replicates $\pm$ standard error of mean. Percent reduction in viability for any particular doxycycline concentration was calculated as the difference in viability percentage between Endo-alone and Endo-ER cells.

Figure 9 shows induced expression of Endo-ER in U2OS/Tet3G cells showed altered cell morphology and reduced cell growth. Bright field microscopy images of U2OS/Tet3G cells expressing Endo alone or Endo-ER. The images were acquired 72hrs after the addition or omission of 20ng/ml doxycycline. Induction of Endo-ER has resulted in very few attached cells on the surface of the cell culture dish compared to Endo alone expressing cells.

Figure 10 shows Western gels of Endo alone or Endo-ER upon induction with 20ng/ml doxycycline in U2OS/Tet3G cells at 0, 2, 4, 6, 8 and 24hr time points. Expression of Endo alone or Endo-ER were detected using FLAG antibody. γ H2AX (Cell Signaling #9718) was used as the DNA damage biomarker. p150(Glued) was used as loading control. The position of the expected Endo alone or Endo-ER protein bands are marked by arrows. Both the proteins are expressed to detectable levels by Western blot at around 4hrs post-induction.

Figure 11 shows Western gels demonstrating that fulvestrant treatment reduced the level of DNA damage response biomarkers upon induced expression of Endo-ER (proof of concept). The expression of Endo alone or Endo-ER were induced by the addition of 20ng/ml doxycycline. For the rescue experiments, 1000nM of fulvestrant was also added along with doxycycline. γ H2AX (Cell Signaling #9718) was used as the DNA damage biomarker. p150(Glued) was used as loading control. The position of the expected Endo alone or Endo-ER protein bands are marked by arrows. Addition of fulvestrant resulted in the degradation of Endo-ER, but not Endo alone, which rescued the level of DNA damage as seen by the γ H2AX DNA damage biomarker.

Figure 12 shows graphs demonstrating that fulvestrant (Faslodex or ICI-182780) treatment significantly regained the cellular proliferation that was affected by the induced expression of Endo-ER (proof of concept). The expression of Endo alone or Endo-ER were induced by the

addition of 20ng/ml doxycycline. For the rescue experiments, 1000nM of fulvestrant was added together with doxycycline. Addition of fulvestrant resulted in regaining cell proliferation to the level that was obtained for Endo alone in the presence of doxycycline only or doxycycline together with fulvestrant shown in panel A. Swarm plot of the data obtained for the 24hr timepoint showing significant rescue of cell proliferation in Endo-ER cells upon fulvestrant treatment, shown in panel B. Data plotted are mean of 8 technical replicates $\pm$ standard error of mean for figure 12A and 24 technical replicates $\pm$ standard error of mean for figure 12B.

Figure 13 shows the fusion of FOXA1 transcription factor to endonuclease exhibited similar reduction in cell proliferation in U2OS/Tet3G cells, whereas induced expression of Endo-FOXA1-DBDdel had less of an effect in reducing viability of the cells

The invention will now be further described with reference to the following non-limiting Examples, and the figures described above.

EXAMPLES

Preparation of the Endo-TF vector construct

The endo screening DNA cassette was designed with the capability of introducing any DNA sequence that codes for a DNA binding domain, including but not limited to transcription factors, by using Gibson assembly technology (New England Biolabs, USA). A 1428bp DNA sequence was synthesised by GeneArt Gene synthesis (Invitrogen) that contained the coding sequence for I-Tev1 endonuclease domain followed by a six amino acid linker GGSGGS (SEQ ID NO: 18) encoded by the sequence GGCGGATCAGGCGGAAGC (SEQ ID NO: 17) which contained the recognition sequence for the blunt-end cutter restriction enzyme Afe1 followed by a FLAG tag (SEQ ID NO: 24), P2A sequence (SEQ ID NO: 25) for ribosome skipping and humanised Monster GFP sequence (SEQ ID NO: 26; Promega) after which stop codons (SEQ ID NO: 27) were introduced in all three reading frames. The entire synthesised sequence was cloned in to the vector pMK-RQ (GeneArt, Invitrogen) as a Sall-BamH1 restriction fragment. The schematic composition of the synthesised endo screening cassette is depicted in Figure 1. The synthesised sequence of the endo cassette is shown in SEQ ID NO: 28.

The GeneArt synthesised Endo cassette was excised out of pMK-RQ (SEQ ID NO: 29) as a Sall-BamHI restriction fragment of approximately 1.4kb and was then sub-cloned in to the response plasmid pTRE3G (Clontech #631168) that was double-digested with the same restriction enzymes. (Figure 2).

For the introduction of coding sequences of DNA binding proteins such as transcription factors (exemplified by, but not limited to, ER and FOXA1), such protein coding DNA fragments were PCR amplified using Phusion DNA polymerase (New England Biolabs, USA; # E0553) with PCR primers that rendered the PCR products compatible for Gibson assembly (New England Biolabs, USA; # E2611) in to the pTRE3G endo vector (SEQ ID NO: 30; Figure 2) that was linearised by restriction digestion (AfeI restriction enzymes, New England Biolabs).

ESR1 (gene encoding for ER alpha) (SEQ ID NO: 7), *FOXA1* (SEQ ID NO: 8) and *FOXA1 DBDdel* (SEQ ID NO: 9) ORFs (*ESR1* #RC213277 and *FOXA1* #RC206045 were supplied by ORIGENE; *FOXA1 DBDdel* was a gift from Carroll lab, Cambridge, UK) were used as template for the PCR amplification of the respective coding sequences. The forward and reverse PCR primer sequences used for the amplification of ER alpha coding sequence were SEQ ID NO: 31 (GGCGGATCAGGCGGAAGCACCATGACCCTCCACACCAAAGCATC) and SEQ ID NO: 32 (GTCGTCCTTGTAGTCAGCGACCGTGGCAGGGAAACCCTCTGC) and those for FOXA1 were SEQ ID NO: 33 (GGCGGATCAGGCGGAAGC*TTAGGAACTGTGAAGATGGAAGG*) and SEQ ID NO: 34 (GTCGTCCTTGTAGTCAGCGGAAGTGT*TTAGGACGGGTCTGG*) [The primer sequence underlined function as overlapping regions for the Gibson assembly whereas the sequence in italics are gene-specific sequence].

An approximately 1.9kb PCR product for *ESR1*, 1.4kb for *FOXA1* and 1.3kb for *FOXA1*-DNA binding domain deletion (*DBDdel*) were obtained.

The plasmid nomenclature used below is vector background-endonuclease fusion partner-DNA binding protein fusion partner. For example, “pTRE3G-Endo-FOXA1” refers to a plasmid comprising the pTRE3G vector, expressing a fusion of I-Tev1 endonuclease with the FOXA1 full length protein comprising of DNA binding domain. Where *DBDdel* is used this indicates the DNA binding protein expressed has a deletion in the DNA binding domain. Where *DBD* is used this refers to any full-length DNA binding protein such as, but not limited to transcription factors or the minimal fragment thereof that is capable of binding to DNA.

The vector maps of the final plasmid constructs that were prepared by Gibson assembly are represented in Figure 3; pTRE3G-Endo-ER (Figure 3A; SEQ ID NO: 39), pTRE3G-Endo-FOXA1 (Figure 3B; SEQ ID NO: 40) and pTRE3G-Endo-FOXA1-*DBDdel* (Figure 3C; SEQ ID NO: 41). Double restriction digestion of the above constructs with *Sall* and *BamHI* (NEB, USA) followed by agarose gel electrophoresis produced an expected vector backbone fragment of approximately 3.4kb as well as an approximately 3.2kb fragment for Endo-ER, 2.9kb for Endo-FOXA1 and 2.7kb for FOXA1-*DBDdel* (SEQ ID NO: 9). All the above constructs were sequenced end-to-end (sequencing primer SEQ ID NOs 35, 36, 37 and 38).

Tet3G transactivator-mediated doxycycline-induced expression of the sequence present downstream of the TRE3G promoter results in the production of a single mRNA comprising the Endo sequence all the way to the humanised monster GFP (hMGFP) sequence. Upon translation, the DNA binding domain coding sequence that is introduced by Gibson assembly into the AfeI digested Endo cassette will be expressed as a fusion protein with the Endo domain fused at its N-terminal and the FLAG-tag at its C-terminal. The P2A sequence present between the FLAG-tag and hMGFP results in ribosome skipping which results in the hMGFP being separated from the rest of the polypeptide and released as a free molecule. This results in the Endo-DBD-FLAG tag and hMGFP to be expressed to equimolar levels in individual cells. The FLAG-tag can be used for the immuno-detection (Western, Immuno-fluorescence, ELISA, but not limited to) of any DBD that is fused to the Endo domain or the Endo domain alone using commercially available FLAG antibodies. This excludes the requirement of production or purchase of antibodies specific to DBDs that are of interest to the user of the endoscreen technology. The synthetic hMGFP gene expresses a 26kDa protein (Promega, USA) with improved fluorescence intensity compared to the native gene. This gene has been codon optimized and cleared of most consensus transcription factor binding sites to ensure reliability and high expression levels. The expression of hMGFP can assist in determining the level and kinetics of Endo-DBD expression in single cells and at the population level, as well as across other endoscreen enabled cell lines using fluorescence technologies. The person of skill in the art will appreciate that other linkers, reporter proteins or markers could be used.

Preparation of stable host cell line for endoscreen

Human osteosarcoma cell line, U2OS, was used to carry out the proof-of-concept experiments for the endoscreen. U2OS cells were obtained from ATCC (#HTB-96) and were grown in DMEM (Thermo Fisher Scientific, #41965-039) supplemented with 10% FBS (Thermo Fisher Scientific, #10500-064) at 37°C with 5% CO₂. Trypsin-EDTA (0.05%) without phenol red (ThermoFisher Scientific, #15400-054) was used for dissociating the cells to prepare the cell suspension. Vi-CELL XR (Beckman Coulter) was used to determine viable cell density. The pCMV-Tet3G regulator plasmid (Figure 4, Clontech #631168) that encodes for the Tet-On 3G transactivator protein was linearised by Scal restriction digestion (NEB, USA, #R3122) and the approximately 7.2kb linear vector was purified using QIAquick PCR purification kit (Qiagen, #28104) following manufacturer's protocol. One day before transfection, U2OS cells were seeded in 24-well Nunc plates (ThermoFisher Scientific, #122475) at a cell density of 75000 cells per well in 1ml of medium. Prior to transfection, the medium was replaced by 0.5ml of fresh medium. Viomer-Red (Lipocalyx GmbH, #VR-01LB) transfection reagent was used to introduce exogenous DNA in to the cells following the

manufacturer's protocol. 500ng of linearised pCMV-Tet3G vector complexed with the transfection reagent was added to individual wells. Standard methods of gene delivery were used, (ROSSI and BECKMAN, Edited by FRIEDMANN, 2007; *Gene Transfer: Delivery and Expression of DNA and RNA, A Laboratory Manual* 2007, University of California, San Diego, Research Institute of the City of Hope, Duarte, California).

48hr after transfection, the cells were dissociated into single cells and transferred into a 10cm Nunc plate (ThermoFisher Scientific, #150350) in a total volume of 10ml medium. Next day, after the cells had attached, the medium was replaced with fresh medium containing 1mg/ml Geneticin (Gibco, #10131035) and this was continued every 3 days for up to 2 weeks until large single colonies of Tet3G stable cells were obtained. 24 such colonies were picked in to 96 well plates (ThermoFisher Scientific, #167008) and were maintained and scaled up by standard cell culture techniques in medium containing 0.2mg/ml Geneticin (Longo et al., *Meth Enzymol*, 536, 165-172, 2014).

To confirm the expression of the Tet3G transactivator protein from the constitutively active CMV promoter, individual clones were grown in 6-well Nunc plates (ThermoFisher Scientific, #140675) to a density of 1×10^6 cells per well. Cells in individual wells were washed with PBS, lysed directly in 100 μ l of 2X SDS loading buffer with β -mercaptoethanol, the lysed cells were transferred in to 1.5ml microtubes, sonicated for 2 mins (30 secs on/30 secs off) in Bioruptor plus, incubated at 80°C for 10minutes and 20 μ l of the resultant sample was loaded per well of a SDS-polyacrylamide gel (ThermoFisher Scientific, NW04125).

Western analysis, using well-known standard techniques, was carried out using antibodies to TetR (Clontech #631132) and p150/Glued (BD Biosciences #610473) as loading control. The result in figure 5 shows the expression of an expected band of approximately 30kDa of the Tet3G protein, which was absent in the non-transfected parental U2OS cells. Though in this case the Tet3G transactivator was expressed under the constitutive CMV promoter. It will be appreciated that this could also be expressed from other promoters such as EF1 alpha, but not limited to, which makes it suitable for long-term expression in a wide range of cell types including hematopoietic cells and stem cells.

Retroviral (for dividing mammalian cells) and lentiviral (for both dividing and non-dividing cells) vectors can also be used for the delivery of the Tet3G transactivator to the cells. The viral mode of delivery is particularly useful for those cells that are difficult to transfect by conventional methods such as lipofection.

The next step was to introduce the vector bearing the endo cassette under the TRE3G inducible promoter in to the U2OS/Tet3G stable cells. Two versions of such vectors

(described previously, Figure 2 and Figure 3) were used for making the subsequent double-stable cells. As a non-limiting example, U2OS/Tet3G cells were made double stable for the induced expression of endo alone (lacking the DNA binding domain, Figure 2) or Endo-Estrogen Receptor alpha (Figure 3A). Importantly, Tetracycline-free FBS (Clontech, USA #631105) was used in the medium for culturing the cells from this step forward. For the preparation of the double stable cells, the corresponding vectors were linearised by Scal restriction digestion (approximately 4.8kb for pTRE3G-Endo alone and 6.6kb for pTRE3G-Endo-ER), the digested vectors were purified and transfected in to U2OS/Tet3G cells as described above. A separate linear Puromycin selection marker (Clontech, USA #631626) under the control of SV40 promoter and the SV40 polyadenylation signal was co-transfected with the endo expression construct for the selection of the double-stable cells. Other selection markers could also be used. Separate linear selection marker avoids promoter interference with the basal expression of the gene of interest and thus results in higher fold induction of the gene of interest in the selected clones. 700ng of linearised pTRE3G-Endo alone or pTRE3G-Endo-ER vector together with 50ng of linear Puromycin selection marker complexed with the transfection reagent was added to U2OS/Tet3G cells individual wells of a 24 well plate. Following the method described above, clonal isolation of double-stable cells was carried out under antibiotic selection with 0.2mg/ml Geneticin (Gibco, #10131035) and 500ng/ml Puromycin (Sigma #P9620).

The endo cassette under the TRE3G inducible promoter can also be delivered to cells using retroviral (for dividing mammalian cells) and lentiviral (for both dividing and non-dividing cells) vectors. The viral mode of delivery is particularly useful for those cells that are difficult to transfect by conventional methods such as lipofection.

In order to check for the induced expression of the integrated endo constructs (endo alone or endo-ER) (ER refers to full length estrogen receptor alpha) and to identify the optimally expressing U2OS clones, Western analysis was carried out on 24 clones each for U2OS/Tet3G endo alone or endo-ER double-stable cells. Each clone was seeded in duplicate wells of a 6-well plate at a density of 0.75×10^6 cells per well the day before induction. The cells were treated with 100ng/ml doxycycline (Sigma #D9891) as the inducer or DMSO as vehicle control for 24hrs and the cells were collected for western analysis as described above. The Western blots were probed with antibodies to FLAG-tag (Sigma #F1804) that detects the expression of Endo alone or Endo-ER and p150/Glued (BD Biosciences #610473) was used as loading control. As shown in figure 6, addition of doxycycline resulted in the expression of a protein band of approximately 29kDa for Endo alone (Endo-FLAG, Figure 6A) and approximately 90kDa for Endo-ER (Endo-ER-FLAG, Figure 6B) and the corresponding bands were absent in the vehicle control treated cells.

Induced expression of Endo-ER in U2OS/Tet3G cells compromised cell viability

U2OS/Tet3G cells containing inducible Endo alone or Endo-ER were seeded in Corning 96-well clear bottom black plates (Fisher scientific #10530753) at a density of 3000 cells per well in DMEM without phenol red (Thermo Fisher Scientific, #31053-028) supplemented with 2mM L-Glutamine (Thermo Fisher Scientific, #25030-081), 1mM Sodium pyruvate (Thermo Fisher Scientific, #11360070) and 10% Tet-free FBS (Clontech #631105) and incubated at 37°C with 5% CO₂ overnight. The next day, the medium was replaced with fresh medium containing 250nM final concentration of IncuCyte Cytotox Red reagent (IncuCyte #4632) with 100ng/ml doxycycline or DMSO vehicle control. The cells were imaged for 48hrs on IncuCyte live cell analysis system (3 locations per well every 3hrs). The diminished plasma membrane integrity of unhealthy cells would result in increased uptake of Cytotox Red reagent yielding between 100-1000-fold increase in red fluorescence up on binding to DNA. The red fluorescent objects (unhealthy cells) were quantified using IncuCyte integrated analysis software.

Upon induction of Endo-ER with 100ng/ml doxycycline, nearly 60% of the cells showed increased uptake of the Cytotox reagent compared to uninduced Endo-ER cells as well as induced or uninduced Endo-alone cells (Figure 7).

This result clearly demonstrated that the DNA binding deficient endonuclease domain of I-TevI used here, when fused to other DNA binding proteins such as transcription factors exemplified by ER, but not limited thereto, can result in DNA cleavage mediated through the DNA binding by the transcription factor to its own recognition sequence.

Titration of doxycycline for optimal expression of Endo-ER in U2OS/Tet3G cells

U2OS/Tet3G cells containing inducible Endo alone or Endo-ER were seeded in Corning 96-well clear bottom black plates (Fisher scientific #10530753) at a density of 3500 cells per well in DMEM (Thermo Fisher Scientific, #41965-039) supplemented with 10% Tet-free FBS (Clontech #631105) and incubated at 37°C with 5% CO₂ overnight. The next day, the medium was replaced with fresh medium containing increasing concentration of doxycycline (0 to 100ng/ml) and the incubation was continued for 72hrs. Cell viability was measured using CellTiter-Glo (CTG) Luminescent Cell Viability Assay reagent (Promega, catalogue number G7570). After equilibrating cells to room temperature for 1 hour, 100µl of CTG reagent was added to each well, and the plate shaken on an orbital rotator at 150rpm for 1 minute. Ten minutes after addition of CTG reagent, luminescence was measured using a PHERAstar FS microplate reader (BMG Labtech, catalogue number 1128717). A doxycycline concentration of 20ng/ml gave the best differential in viability between Endo

alone and Endo-ER cells by CellTitre-Glo assay (Figure 8). Bright field microscopy images showed markedly reduced growth of U2OS/Tet3G cells expressing Endo-ER compared to Endo alone cells (Figure 9). Promega CellTitre-Glo assay protocol was used.

Time-course of induced expression of Endo alone or Endo-ER in U2OS/Tet3G cells

Endo alone or Endo-ER bearing U2OS/Tet3G cells were grown in 6-well plates as described above and the expression of respective proteins were induced by the addition of 20ng/ml doxycycline. Cells were harvested at 0, 2, 4, 6, 8 and 24hr time points for Western analysis as described above. Expression of Endo alone or Endo-ER were detected using FLAG antibody. Protein bands corresponding to both Endo alone and Endo-ER were expressed to detectable levels at around 4hrs after induction with doxycycline (Figure 10). The increased expression of Endo-ER through the time-course was also reflected in the increased presence of the signal for the DNA damage biomarker γ H2AX (Cell Signaling #9718) (Firsanov et al., Genome Stability, 37, 635-649, 2016)

Selective Estrogen Receptor degrader (SERD) drug Fulvestrant dramatically reduces DNA damage response biomarker levels in Endo-ER-expressed U2OS/Tet3G cells

Approximately 70% of breast cancers are ER+ which respond well to hormone therapies. Fulvestrant (Faslodex or ICI-182780) is the first selective estrogen receptor downregulator or degrader (SERD) made available for clinical practice (Rocca et al., Cancer Manag Res., 10:3083-3099, 2018).

An endoscreen of the invention was designed to identify inhibitors of DNA binding proteins such as transcription factors which prevents them from localisation to its DNA recognition sites, thereby preventing DNA damage and subsequent cell death leading to increased cell survival. As a proof-of-principle we treated U2OS/Tet3G cells that were induced to express Endo alone or Endo-ER with 20ng/ml doxycycline together with 1000nM fulvestrant (Selleck Chemicals #S1191) for 24hrs. The cells were harvested for Western analysis as described earlier. The FLAG antibody detected the expression of both Endo alone and Endo-ER when doxycycline alone was added to the cells (Figure 11). High level of expression of Endo alone did not cause significant increase in DNA damage in the cells as shown by the low level of γ H2AX signal. This confirms the inability of the Endo domain used in this assay to bind to cellular DNA. On the contrary, expression of Endo-ER resulted in significant increase in the γ H2AX signal. Addition of fulvestrant resulted in considerable degradation of Endo-ER, but the level of Endo alone remained unchanged. The elimination of Endo-ER by the potent

SERD fulvestrant was also reflected in the reduction of the γ H2AX signal to basal levels as seen in the Endo alone control cells (Figure 11).

In order to assess whether the DNA damage caused by the expression of Endo-ER and its rescue by fulvestrant was also reflected in the growth kinetics of the cells, a proliferation assay was conducted using IncuCyte live cell analysis system. U2OS/Tet3G cells containing inducible Endo alone or Endo-ER were seeded in Corning 96-well clear bottom black plates (Fisher scientific #10530753) at a density of 3500 cells per well in DMEM (Thermo Fisher Scientific, #41965-039) supplemented with 10% Tet-free FBS (Clontech #631105) and 2.5mM Thymidine (Sigma #T1895) (Banfalvi G. (ed) *Cell Cycle Synchronization. Methods in Molecular Biology*, vol 76, 2011. Humana Press) and incubated at 37°C with 5% CO₂ for 24hrs. The next day, the cells were released from the thymidine block for 5hrs by gently washing and replacing with fresh medium. After 5hrs of release from the thymidine block, the medium was again replaced with fresh medium containing 20ng/ml doxycycline alone or 1000nM fulvestrant alone or both together. The total time available for release of cells from the thymidine block would be approximately 8hrs before the cells start to express the Endo proteins (Figure 10). The cells were subsequently imaged for 24hrs on IncuCyte live cell analysis system (3 locations per well every 3hrs). The confluence percentage as a reflection of proliferation was quantified using IncuCyte integrated analysis software.

Induced expression of Endo-ER resulted in significantly reduced proliferation of U2OS/Tet3G cells (Figure 12). Addition of fulvestrant along with doxycycline inducer regained the proliferation to the level comparable with that of Endo alone that were exposed to doxycycline only or together with fulvestrant. These results demonstrate that fulvestrant is a DNA binding protein inhibitor (in this case through directed degradation of Endo-ER) and that this assay can be used to identify inhibitors of DNA binding proteins and confirms the feasibility of the endoscreen assay in drug discovery screening assays.

This system could therefore be used to screen for inhibitors of ER.

Fusion of FOXA1 transcription factor to endonuclease exhibited similar reduction in cell proliferation in U2OS/Tet3G cells as Endo-ER

Molecular and genetic studies have shown that FOXA1 is abnormally expressed in a number of cancer types, including acute myeloid leukemia (AML), lung, oesophageal, thyroid, breast and prostate cancers (Yang and Yu. *Genes Dis.* 2(2),144–151, 2015). Genome-wide mapping experiments have revealed the critical role of FOXA1 in permitting chromatin accessibility for ER to interact with cis-regulatory regions (Carroll et al., *Cell.* 122(1), 33-43.2,

2005 and Carroll et al., Nat Genet. 38(11), 1289-97, 2006). FOXA1 is the critical determinant of the Androgen Receptor transcriptional programme in prostate cancer (Yang, Y.A. and Yu, J. Genes Dis. 2(2):144-151, 2015). Analyses of human prostate cancer specimens have revealed that FOXA1 is overexpressed in metastatic as well as castration-resistant prostate cancer (CRPC) patients (Gerhardt et al., Am J Pathol., 180:848–861, 2012). FOXA1 gene is mutated in cancer, as reported in the scientific literature as well as in The Cancer Genome Atlas (TCGA). Collectively cancer that requires the presence of FOXA1 protein for cell proliferation or survival can be termed FOXA1-dependent cancers. Despite the considerable opportunity FOXA1 has not yet been directly drugged. FOXA1 has no ligand and is non-catalytic and is therefore not amenable to classic drug discovery approaches. Accordingly, there is an unmet need in the art for new treatments for FOXA1-dependent cancers.

In order to confirm whether a second transcription factor such as FOXA1 can be adapted in the endoscreen technology, a transient transfection approach was used with inducible endo vectors that contained full length FOXA1 coding sequence (Bingle and Gowan, Biochim Biophys Acta., 1307(1) :17-20, 1996; SEQ ID NO: 8) or a mutant version that lacked the DNA binding domain (gift from Carroll lab, Cambridge UK, unpublished; SEQ ID NO: 9) and therefore unable to bind to its recognition sequences on the genome (Figures 3B and 3C). One day before transfection, U2OS/Tet3G cells were seeded in 24-well Nunc plates (ThermoFisher Scientific, #122475) at a cell density of 20000 cells per well in 1ml of medium. Prior to transfection, the medium was replaced by 0.5ml of fresh medium. Viromer-Red (Lipocalyx GmbH, #VR-01LB) transfection reagent was used to introduce exogenous DNA in to the cells following manufacturer's protocol. 750ng of pTRE3G-Endo-FOXA1 (Figure 3B; SEQ ID NO: 40) or pTRE3G-Endo-FOXA1-DBDdel (Figure 3C; SEQ ID NO: 41) vector complexed with the transfection reagent was added to individual wells.

Five hours post-transfection, the cells were washed with fresh medium. After adding 1ml of fresh medium with or without 20ng/ml doxycycline, cell growth was monitored for 72hrs on IncuCyte live cell analysis system (16 locations per well every 3hrs). The confluence percentage as a reflection of proliferation was quantified using IncuCyte integrated analysis software.

As seen in the case of Endo-ER, induced expression of Endo-FOXA1 resulted in significant reduction in the proliferation of U2OS/tet3G cells (Figure 13). Expression of DNA binding deficient version of FOXA1 (FOXA1-DBDdel) did not show considerable reduction in the proliferation of cells. This result reiterates the fact that the endoscreen can be used not only for ER but also for other DNA binding proteins such as the broad family of transcription factors.

This system could therefore be used to screen for inhibitors of FOXA1.

Improved target identification by combining endoscreen with CRISPR/Cas9 whole genome screening

The endoscreen assay has been designed as a 'plug and play' assay, where any DNA binding protein (DBP, e.g. transcription factor (TF)) of interest in any disease can be inserted into the inducible vector and expressed in the host cell line. This technology represents an improved way to screen for inhibitors of DBPs binding based on fusion of the DBP of interest to an endonuclease domain which generates double strand breaks leading to cell death when the DBP binds to its DNA recognition sites. An inhibitor of DBP function which prevents its localisation to its DNA sites, would prevent cell death and cell survival would be the readout. Importantly, a positive signal in the assay can only result from DBP dissociation from the chromatin, a critical endpoint that ensures that unbiased identification of an inhibitor represents a compound that results in global dissociation of the specific DBP from DNA.

Furthermore, the gain of function signal of the technology, renders this assay format ideal for whole genome CRISPR/Cas9 gene knock-out screening (Shalem et al., Science, 343 (6166), 84-87, 2014; Ran et al., Nat Protoc., 8(11), 2281-2308, 2013).

For example, if CRISPR knock-out of a gene (such as, but not limited to, a kinase) prevents the localisation of Endo-TF to its DNA sites, this would prevent cell death resulting in cell survival as the readout. The ability to perform whole-genome functional studies in parallel with compound screening will greatly facilitate target identification. It could be envisaged that the CRISPR/Cas9 screen is used as the primary screen to identify druggable TF targets that can be brought forward into standard drug discovery, targeted biochemical or fragment-based screening approaches (*CRISPR-Cas: A Laboratory Manual* 2016, Edited by Jennifer Doudna, University of California, Berkeley; Prashant Mali, University of California, San Diego).

Potential utilities of endoscreen

This utility of endoscreen is to expand the spectrum of disease-relevant drug-targets for therapeutic intervention. A surfeit of biomolecular targets has been implicated in disease pathophysiology, however, many are considered undruggable and of these a large proportion are DNA binding proteins and whose biological activity requires DNA binding. For example, in oncology FOXA1, MYC, MYB, and nuclear factor- κ B (NF- κ B) are all highly validated as biologically meaningful targets yet remain undrugged. This is largely due to

these being difficult targets against which to apply conventional drug discovery approaches, often due to lack of catalytic or ligand dependent activity, large protein–protein interaction interfaces or their lack of deep protein pockets.

Endoscreen provides a new unbiased cell-based approach to identifying chemical and biological reagents that can cause the loss of DNA binding of the DNA binding protein of interest. This will inform new ways of inhibiting the biological function of the DNA binding protein for therapeutic purposes. Endoscreen may be used as a primary discovery or screening assay, or as an assay orthogonal to other discovery methods.

SEQUENCE LISTING FEATURES:

SEQ ID NO:1 P13299|TEV1_BPT4 Intron-associated endonuclease 1 OS=Enterobacteria phage T4; Full length: 245 amino acids

SEQ ID NO: 2 I-TevI endonuclease domain peptide sequence (201 amino acids)

SEQ ID NO:3 Q9ANR6_BACMO Intron encoded BmoI OS=Bacillus mojavensis; Full length: 266 amino acids; Nuclease domain: 1-92

SEQ ID NO:4 I-Tulal endonuclease protein sequence; Full length: 245 amino acids; Nuclease domain: 1-114

SEQ ID NO:5 P34081 I-HmuI endonuclease protein sequence; Full length: 174 amino acids

SEQ ID NO: 6 Eco31R type II restriction endonuclease protein sequence

Q8RNY7_ECOLX Alw26I/Eco31I/Esp3I family type II restriction endonuclease

SEQ ID NO: 7 Human ER alpha wild-type full length coding nucleotide sequence

SEQ ID NO: 8 Human FOXA1 wild-type full length coding nucleotide sequence

SEQ ID NO: 9 Human FOXA1 coding nucleotide sequence with DNA binding domain deleted

SEQ ID NO: 10 Human FOXA1 wild-type protein sequence (NP_004487.2)

SEQ ID NO: 11 Human FOXA1 protein sequence without DNA binding domain (DBD)

SEQ ID NO: 12 Human ER alpha wild-type protein sequence (NP_001116214.1)

SEQ ID NO: 13 Human ER alpha protein sequence with D5538G mutation

SEQ ID NO: 14 Human ER alpha protein sequence with Y537S mutation

SEQ ID NO: 15 Human Androgen receptor transcript Variant 3 protein sequence (AR-V7), NP_001334990.1

SEQ ID NO: 16 Human FOXA1 isoform 2 protein sequence with amino acids 1-33 missing

SEQ ID NO: 17 Linker nucleotide sequence used in the endo construct; Encodes 6 amino acids with the last serine contributed by 5'-half of Afe1 restriction endonuclease recognition sequence)

SEQ ID NO: 18 Linker peptide sequence used in the endo construct; Translation of SEQ ID NO 17

SEQ ID NO: 22 I-TevI endonuclease recognition sequence (where N is any nucleotide)

SEQ ID NO: 23 Motif defining homing endonuclease class

SEQ ID NO: 24 FLAG-tag coding sequence (from Sigma)

SEQ ID NO: 25 P2A peptide coding sequence for ribosome skipping

SEQ ID NO: 26 Humanised MonsterGFP (hMGFP) coding sequence (from Promega)

SEQ ID NO: 27 Nucleotide sequence with STOP codons in all three reading frames

SEQ ID NO: 28 Endo Cassette nucleotide sequence as SalI-BamHI restriction endonuclease fragment; synthesized by GeneArt

SEQ ID NO: 29 pMK-RQ Endo cassette; Supplied by GeneArt after custom Synthesis of endo cassette and cloning in to pMQ-RQ vector

SEQ ID NO: 30 pTRE3G Endo cassette

SEQ ID NO: 31 ER-Gibson_F PCR primer sequence

SEQ ID NO: 32 ER-Gibson R PCR primer sequence

SEQ ID NO: 33 FOXA1-Gibson_F PCR primer sequence

SEQ ID NO: 34 FOXA1-Gibson_R PCR primer sequence

SEQ ID NO: 35 TRE3G-seqF sequencing primer

SEQ ID NO: 36 GIY-seqF sequencing primer

SEQ ID NO: 37 hMGFP-seqR sequencing primer

SEQ ID NO: 38 TRE3G-seqR sequencing primer

SEQ ID NO: 39 pTRE3G Endo ER nucleotide sequence

SEQ ID NO: 40 pTRE3G Endo FOXA1 nucleotide sequence

SEQ ID NO: 41 pTRE3G Endo FOXA1 delDBD nucleotide sequence

Claims:

1. A method of screening for a putative inhibitor of a DNA binding protein comprising:
 - a) bringing into contact (i) a host cell that expresses an inducible fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide; and, (ii) a test compound, under conditions where the fusion protein, in the absence of the test inhibitor compound, is capable of binding to and cleaving host cellular DNA; and
 - b) determining whether the test compound reduces the ability of the fusion protein to cleave cellular DNA,

wherein if the test compound inhibits the ability of the fusion protein to cleave cellular DNA the test compound is a putative inhibitor of the DNA binding protein.

2. The method according to claim 1, wherein the ability of the test compound to inhibit the ability of the fusion protein to cleave cellular DNA is determined indirectly by measuring a host cell phenotype, such cell number, cell growth, cell proliferation, cell viability, cellular DNA damage or cell cycle phase distribution.
3. The method according to claim 2, wherein the ability of the fusion protein to cleave cellular DNA, cellular DNA damage, cell viability, cell proliferation, or cell cycle phase distribution arrest when in the presence of the test compound is compared to suitable reference or control values or control cells.
4. The method according to any one of claims 1, 2 or 3, wherein the inhibitor is one that interferes with the DNA binding of a DNA binding protein.
5. The method according to any one of claims 1 to 4, wherein the inhibitor is one that directly interferes with the DNA binding of a DNA binding protein.
6. The method according to any one of claims 1 to 4, wherein the inhibitor is one that indirectly interferes with the DNA binding of a DNA binding protein.
7. The method according to any one of claims 1 to 4, wherein the inhibitor is one that inhibits pre-processing of a DNA binding protein.
8. The method according to any one of claims 1 to 7, wherein the ability of the test compound to inhibit the ability of the fusion protein to cleave cellular DNA determined in step (b) is or can be compared to:
 - (i) a reference cell viability or cell cycle phase separation;
 - (ii) cell viability or cell cycle arrest when in the absence of the test inhibitor compound;

- (iii) cell viability or cell cycle arrest when in the presence of a control compound which is known to not inhibit the DNA binding protein;
- (iv) cell viability or cell cycle arrest where the nuclease is not activated; or
- (v) cell viability or cell cycle arrest where the DNA binding domain of said fusion protein has been mutated such that it cannot bind DNA,

wherein an increase in cell number or cell viability, or reduction in cell cycle arrest or reduction in DNA damage in the presence of the test inhibitor compound compared to any of (i) – (v) in the absence of the test compound is indicative that the test inhibitor compound is a putative inhibitor of the DNA binding protein.

9. The method according to any one of claims 1 to 8, wherein cell viability or cell cycle arrest is assessed by measuring ATP consumption or cell confluence, wherein an increase in ATP consumption or cell confluence indicates that the compound is an inhibitor of the DNA binding protein.

10. The method according any one of claims 1 to 9, wherein cellular DNA damage is determined by measuring the amount of gammaH2AX or RAD51 foci produced by the cells, wherein a decrease in gammaH2AX or RAD51 foci production in the cells indicates that the compound is an inhibitor of the DNA binding protein.

11. The method according to any one of claims 1 to 10, wherein the compound is a small molecule compound or a nucleic acid containing molecule.

12. The method according to claim 11, wherein the nucleic acid containing molecule is an antisense oligonucleotide, a guide RNA (gRNA) or an RNAi molecule, such as siRNA.

13. The method according to any one of claims 1 to 12, wherein the host cells are eukaryotic cells.

14. The method according to any one of claims 1 to 13, wherein the host cells are mammalian cells, cell lines or immortalized cells

15. The method according to any one of claims 1 to 14, wherein a compound identified as an inhibitor of the DNA binding protein is tested to confirm that it is not an inhibitor of the DNA endonuclease polypeptide.

16. The method according to any one of claims 1 to 15, wherein the DNA binding protein is a transcription factor.

17. The method according to claim 16, wherein the DNA binding protein is a transcription factor implicated in a disease, such as in cancer.

18. The method according to any one of claims 1 to 17, wherein the DNA binding protein is a transcription factor implicated in cancer selected from the group consisting of: Pterin-4 Alpha-Carbinolamine Dehydratase 2 (TCF1), Groucho, estrogen receptor alpha (ER α), estrogen receptor beta (ER β), Aryl Hydrocarbon Receptor (AHR), Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT), Retinoic Acid Receptor (RAR), Retinoid X Receptor

(RXR), Jun Proto-Oncogene, AP-1 Transcription Factor Subunit (JUN), Fos Proto-Oncogene, AP-1 Transcription Factor Subunit (FOS), Activating Transcription Factor 2 (ATF2), ETS Domain-Containing Protein Elk-1 (ELK1), DNA Damage Inducible Transcript 3 (DDIT3 / GADD153), ABL Proto-Oncogene 1, Non-Receptor Tyrosine Kinase (c-ABL), Nuclear Factor Kappa B Subunit 1 (NFκB), Inhibitor Of Nuclear Factor Kappa B Kinase Subunit Beta (IκB), RB1 (RB Transcriptional Corepressor 1), E2F Transcription Factor 1 (E2F), TATA-Box Binding Protein (TBP), Tumor Protein P53 (TP53) and P21, BAX, FAS, APO1, BAD, BCL2, GADD45, androgen receptor (AR), Rho GTPase Activating Protein 35 (ARHGAP35), AT-Rich Interaction Domain 5B (ARID5B), Achaete-Scute Family BHLH Transcription Factor 1 (ASCL1), ASH1 Like Histone Lysine Methyltransferase (ASH1L), Activating Transcription Factor 1 (ATF1), BCL11A, BAF Complex Component (BCL11A), BCL6 Transcription Repressor (BCL6), BCL2 Associated Transcription Factor 1 (BCLAF1), Basonuclin 2 (BNC2), POU Class 3 Homeobox 2 (BRN2), CCAAT or CCAAT Enhancer Binding Protein Beta (C/EBPβ), Core-Binding Factor Subunit Beta (CBFB), Capicua Transcriptional Repressor (CIC), Clock Circadian Regulator (CLOCK), CSL, CCCTC-Binding Factor (CTCF), E74 Like ETS Transcription Factors (ELF1, ELF3, ELF4 and ELF5), Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit (EZH2), Forkhead Box A1 (FOXA1), Forkhead Box A2 (FOXA2), Forkhead Box C2 (FOXC2), Forkhead Box P1 (FOXP1), Forkhead Box P3 (FOXP3), GATA Binding Protein 1, 2, 3, and 4 (GATA1, GATA2, GATA3, and GATA4), Hepatocyte Nuclear Factor 4 Alpha (HNF4), Interferon Regulatory Factor 2, 6 and 7 (IRF2, IRF6, and IRF7), Kruppel Like Factor 4 and 6 (KLF4, KLF6), MDS1 And EVI1 Complex Locus (MECOM), Myocyte Enhancer Factor 2A (MEF2), MYB Proto-Oncogene, Transcription Factor (MYB), MYC Proto-Oncogene, BHLH Transcription Factor (MYC), MYCN Proto-Oncogene, BHLH Transcription Factor (MYCN), Nuclear Factor Of Activated T Cells 4 (NFATC4), Neurogenin 3 (NGN3), NK3 Homeobox 1 (NKX3-1), Nuclear Factor, Erythroid 2 Like 2 (NR2F2), Nuclear Receptor Subfamily 4 Group A Member 2 (NR4A2), Nuclear Receptor Subfamily 5 Group A Member 1 (NR5A1), Octamer-Binding Protein 4 (OCT4), Paired Box 5 (PAX5), Pancreatic And Duodenal Homeobox 1 (PDX1), Forkhead Box M1 (FOXM1), Paired Related Homeobox 1 (PRRX1), Runt Related Transcription Factor 1 (RUNX1) SRY-Box 2, 7, and 9 (SOX2, SOX7, SOX9), TAL BHLH Transcription Factor 1, Erythroid Differentiation Factor (TAL1) T-box 3 and 5 (TBX3, TBX5) Transcription Factor 12, 4, and 7L2 (TCF12, TCF4, TCF7L2), Transcription Factor Dp-1 (TFDP1), Transcription Factor Dp-2 (TFDP2) and a methyltransferase (including Dnmt1, Dnmt 3A and Dnmt 3B).

19. The method according to any of any one of claims 1 to 18, wherein the DNA binding protein is a pioneer factor.

20. The method according to claim 19, wherein the pioneer factor is selected from the group consisting of FOXA1, FOXA2, FOXA3, GATA1, GATA2, GATA4, PU.1, Zelda, Pou5f3, Group B1 Sox and Sox2, Oct3/4, Klf4, Ascl1, Pax7, p53 and CLOCK:BMAL1.
21. The method according to any one of claims 1 to 20, wherein the DNA binding domain is from FOXA1 protein and comprises the sequence disclosed in SEQ ID NO: 10.
22. The method according to any one of claims 1 to 21, wherein the DNA binding domain is from ER protein and comprises the sequence disclosed in SEQ ID NO: 12.
23. The method according to any one of claims 1 to 22, wherein the DNA endonuclease polypeptide lacks its endogenous DNA binding domain.
24. The method according to any one of claims 1 to 23, wherein the DNA endonuclease polypeptide is capable of creating single-stranded breaks in DNA.
25. The method according to any one of claims 1 to 23, wherein the DNA endonuclease polypeptide is capable of creating double-stranded breaks in DNA.
26. The method according to claim 25, wherein the DNA endonuclease polypeptide is selected from the group consisting of: GIY-YIG, LAGLIDADG, HNH and His-Cys box.
27. The method according to claim 26, wherein the DNA endonuclease polypeptide is GIY, a functional variant thereof or a functional fragment of either.
28. The method according to claim 27, wherein the DNA endonuclease polypeptide is GIY-YIG fragment
29. The method according to claim 25, wherein the DNA endonuclease polypeptide is selected from the group consisting of: I-TevI endonuclease, I-BmI endonuclease, FOK1 endonuclease and I-SceI endonuclease, or a functional variant or functional fragment of any thereof.
30. The method according to any one of claims 1 to 29, wherein the DNA binding domain polypeptide is fused directly or indirectly to the C-terminus of the DNA endonuclease polypeptide
31. The method according to any one of claims 1 to 29, wherein the DNA binding domain polypeptide is fused directly or indirectly to the N-terminus of the DNA endonuclease polypeptide
32. The method according to any one of claims 1 to 29, wherein there is a linker between the DNA binding polypeptide and the DNA endonuclease polypeptide.
33. The method according to claim 32, wherein the linker is from 2-20 amino acids in length, such as 6 amino acids in length.
34. The method according to claim 32 or claim 33, wherein the linker comprises the sequence GSGGS.

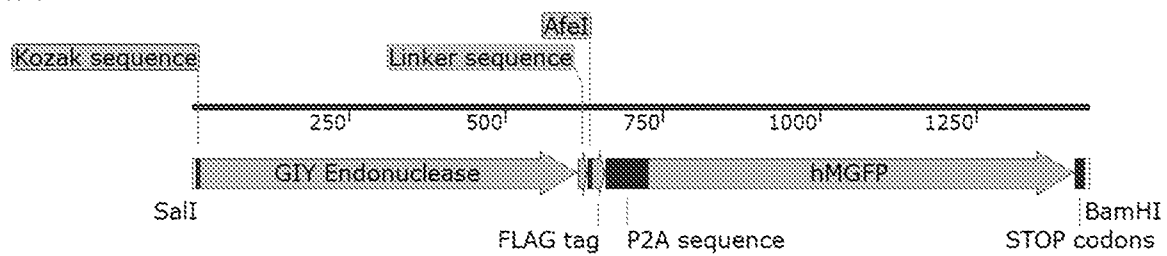
35. The method according to claim 29, wherein the fusion protein comprises an I-TevI endonuclease polypeptide fused to human FoxA1 polypeptide, optionally also comprising a FlagTag sequence.
36. The method according to claim 29, wherein the fusion protein comprises an I-TevI endonuclease polypeptide fused to human ER polypeptide, optionally also comprising a FlagTag sequence.
37. The method according to any one of claims 1 to 36, wherein the fusion protein is expressed from an inducible promoter selected from the group consisting of: Tet-ON, lac, Ecdysone, Auxin-inducible, hsp70-promoter induced, Cumate-system, Geneswitch.
38. The method according to any one of claims 1 to 37, wherein a nucleic acid construct capable of facilitating expression of the fusion protein has been inserted into the host cell.
39. The method according to claim 38, wherein the nucleic acid construct inserted into the host cell is a plasmid or other episomal vector.
40. The method according to claim 38 or claim 39, wherein the nucleic acid construct has been inserted into the cell via transfection, transduction, electroporation or transformation.
41. The method according to claim 38, wherein the nucleic acid construct is also capable of expressing a marker and/or reporter protein and or a tag sequence.
42. The method according to claim 41, wherein the marker is selected from the group consisting of a fluorescent marker, a luminescent marker and an antibiotic resistance gene.
43. The method according to claim 42, wherein the reporter protein is selected from the group consisting of: a fluorescent reporter protein such as Green Fluorescent Protein (GFP), Red Fluorescent Protein (RFP), mCherry, Cyan Fluorescent Protein (CFP), Blue Fluorescent Protein (BFP) and Yellow Fluorescent Protein (YFP) and a luminescent reporter protein, such as Luciferase or NanoLuc.
44. The method according to claim 43, wherein the tag sequence is selected from the group consisting of but not limited to: Cellulose Binding Domain (CBD), Dihydrofolate reductase (DHFR), Calmodulin binding protein (CBP), FLAG, SUMO, S-tag, Glutathione S-Transferase (GST), Hemagglutinin A (HA), Histidine (His), Herpes Simplex Virus (HSV), Maltose-Binding Protein (MBP), c-Myc, Protein A, Protein G, Streptavidin, T7, V5, Vesicular Stomatitis Virus Glycoprotein (VSV-G), Yeast 2-hybrid tags (B42, GAL4, LexA, VP16); fluorescent reporter proteins (such as Green Fluorescent Protein (GFP), Red Fluorescent Protein (RFP), mCherry, Cyan Fluorescent Protein (CFP), Blue Fluorescent Protein (BFP), Yellow Fluorescent Protein (YFP); luminescent reporter proteins such as Luciferase, NanoLuc.
45. The method according to any one of claims 41, 43 or 44, wherein the reporter protein or tag sequence is expressed from the same inducible promoter as the fusion protein.

46. The method according to claim 45, wherein the reporter protein or tag sequence is attached to the fusion protein.
47. The method according to claim 45, wherein the reporter protein or tag sequence is expressed separately from the fusion protein.
48. A method of preparing a pharmaceutical composition comprising an inhibitor of a DNA binding protein comprising: (i) selecting a compound that is an inhibitor of a DNA binding protein according to any one of claims 1 to 47; and, (ii) admixing the compound with one or more pharmaceutically acceptable excipients.
49. A host cell capable of expressing a fusion protein comprising a DNA binding domain from a DNA binding protein and a functionally active DNA endonuclease polypeptide in accordance with any one of claims 1 to 47.

FIGURES:

Figure 1

1A



1B

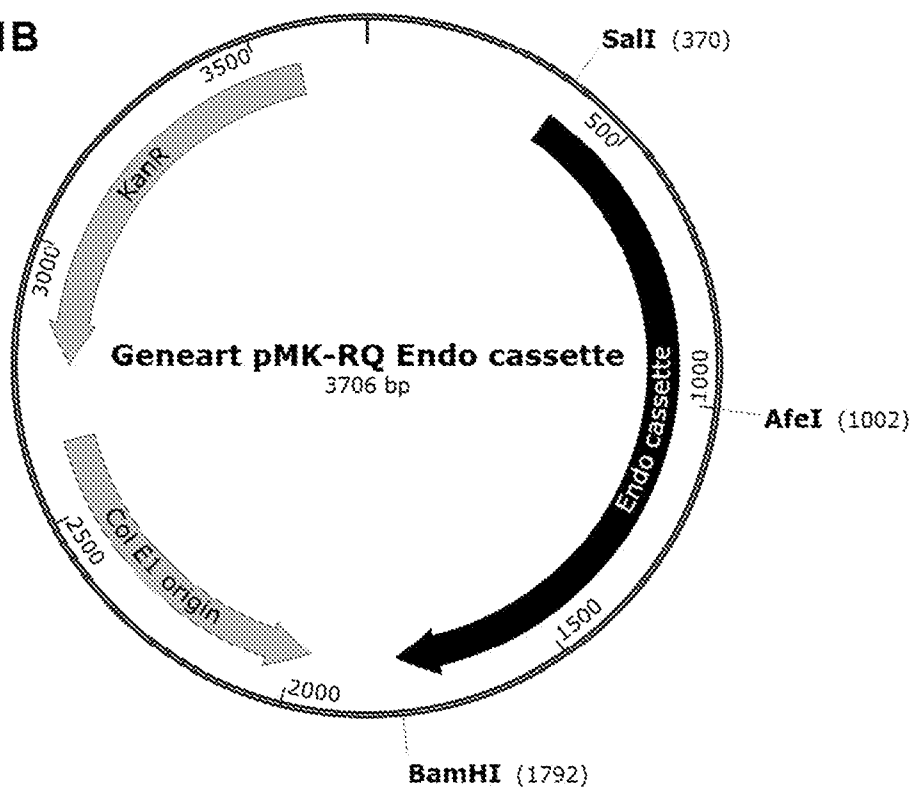


Figure 2

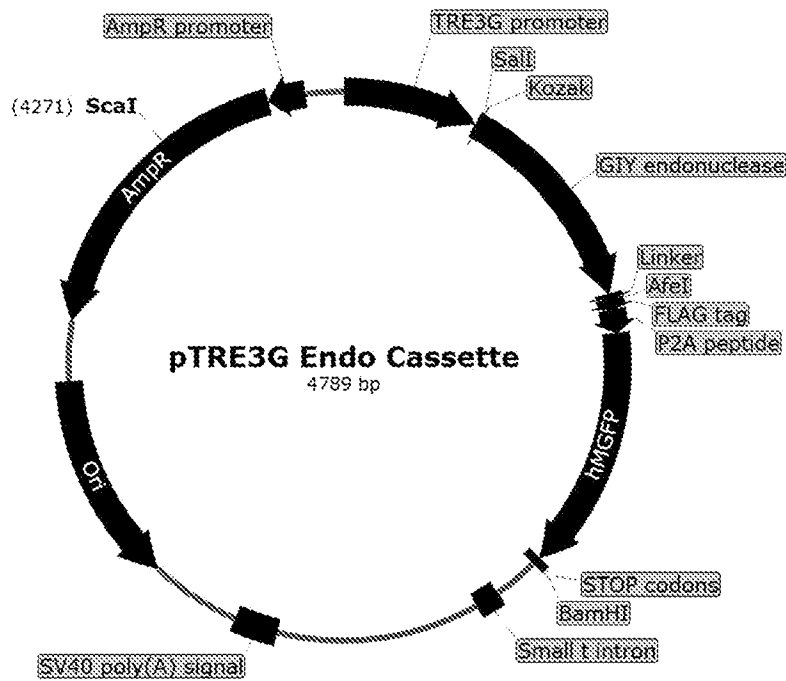


Figure 3

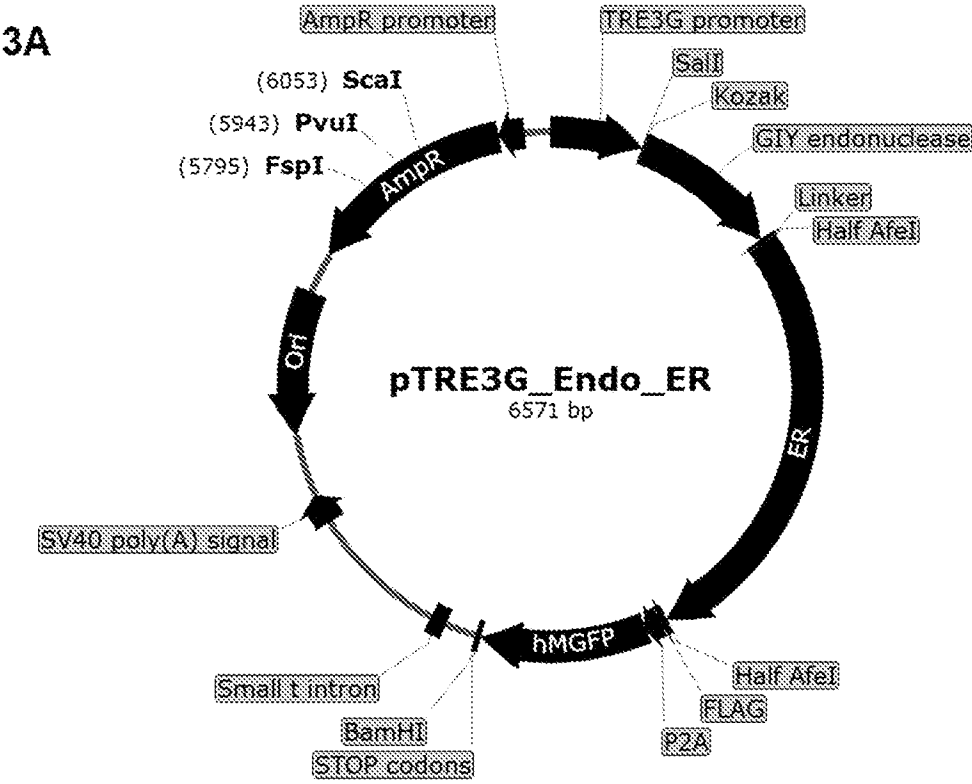


Figure 3 (contd)

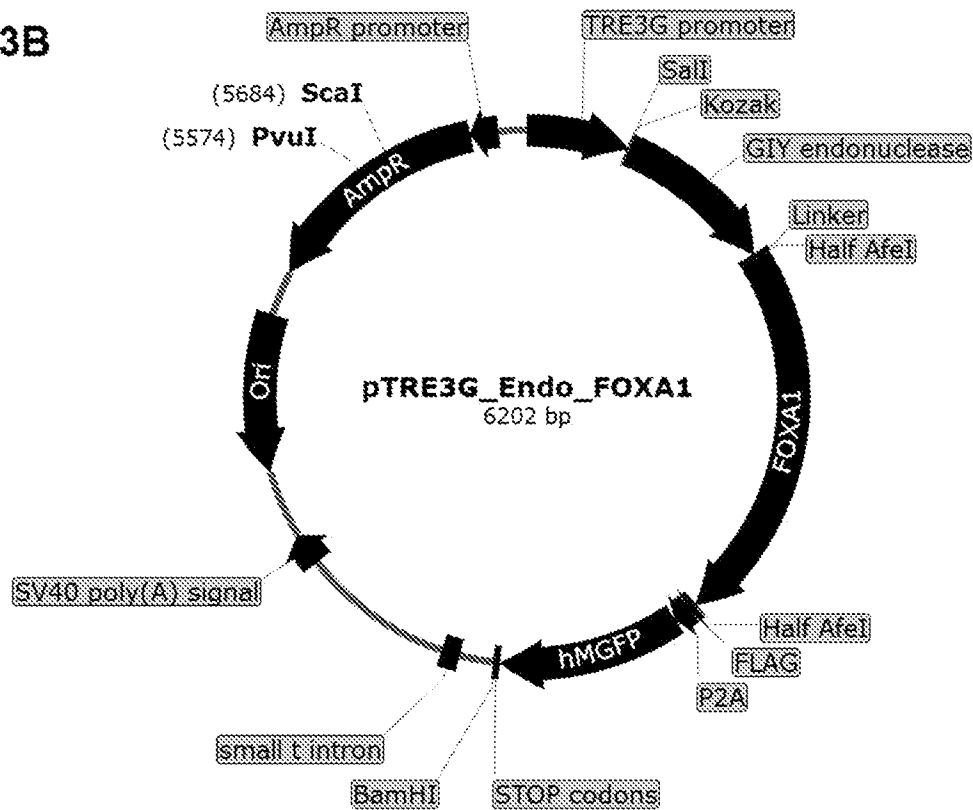
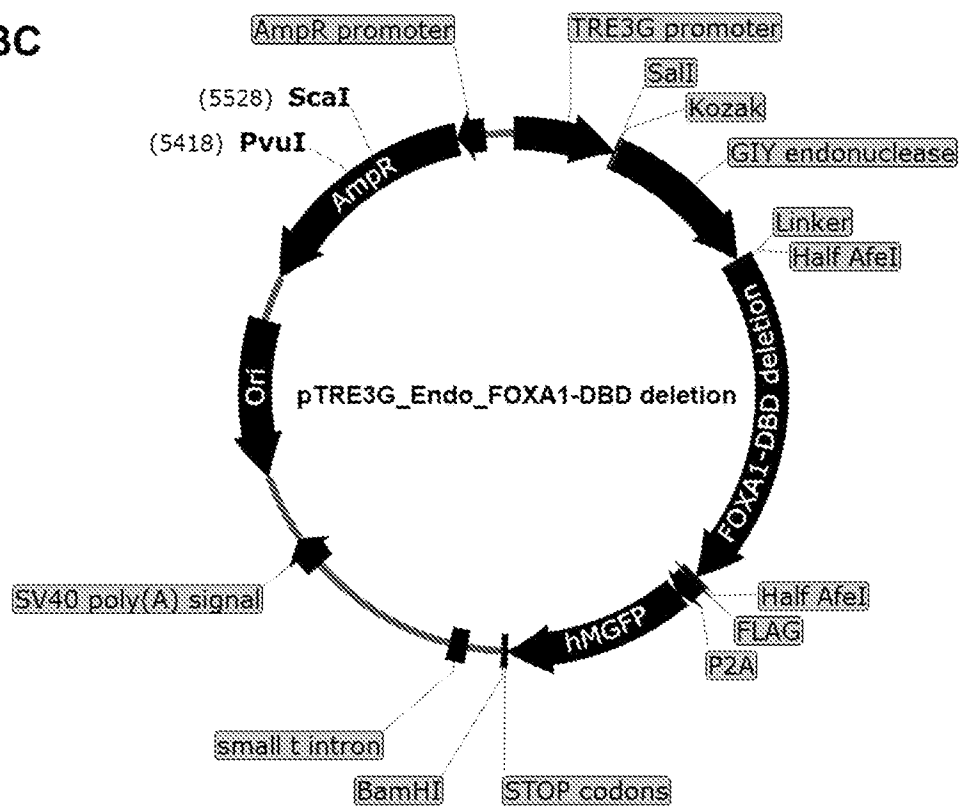
3B**3C**

Figure 4

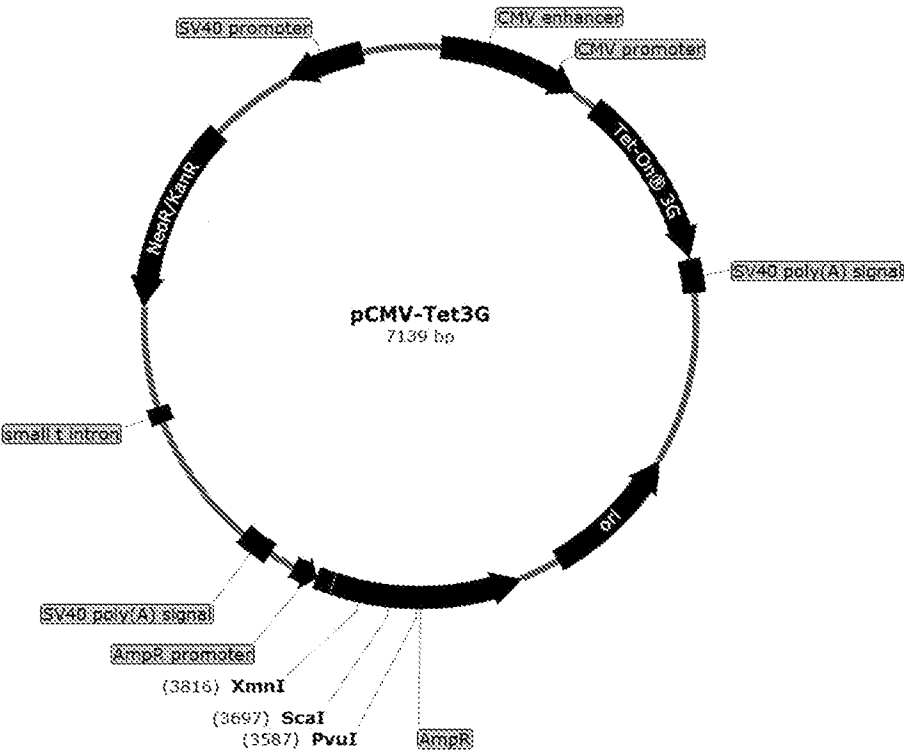


Figure 5

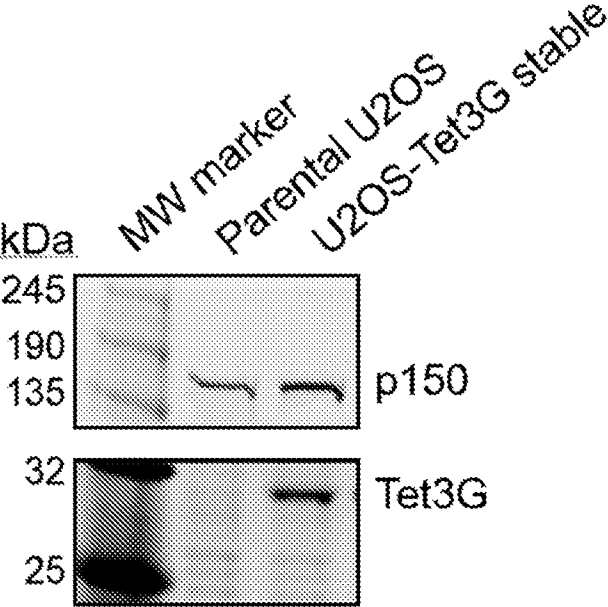


Figure 6

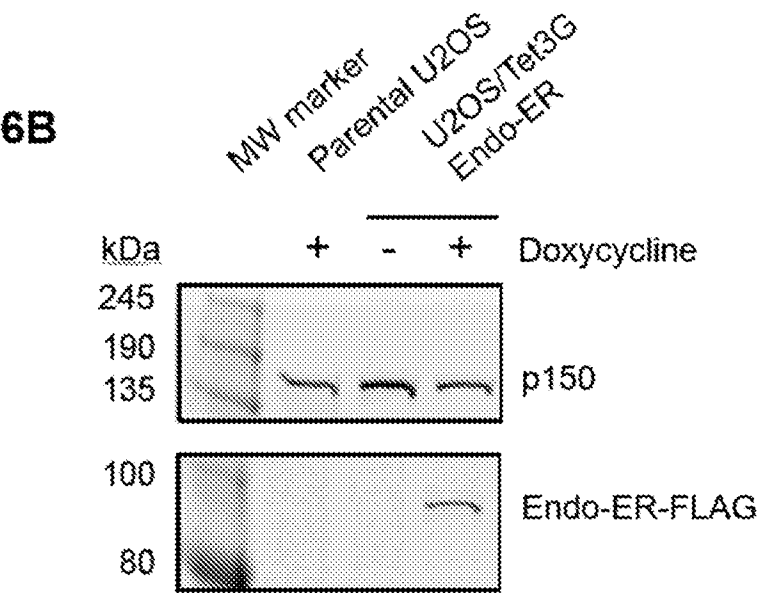
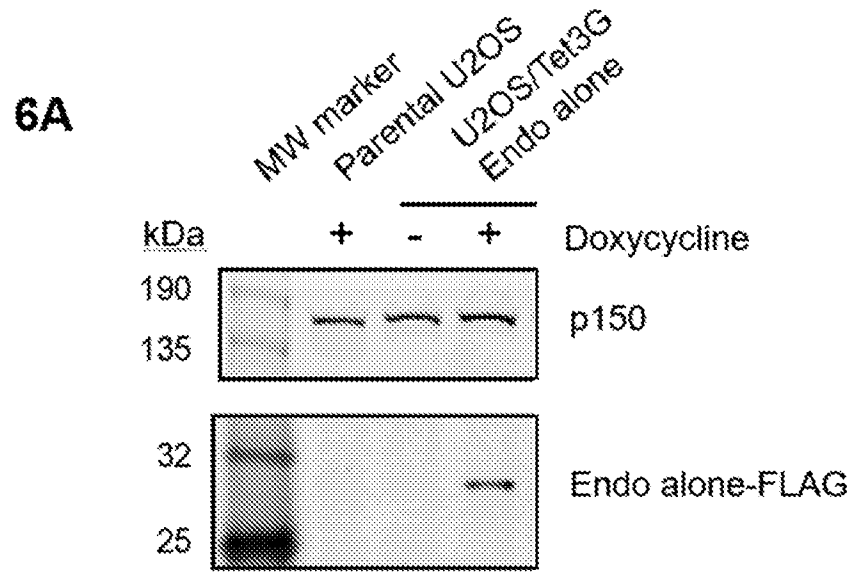


Figure 7

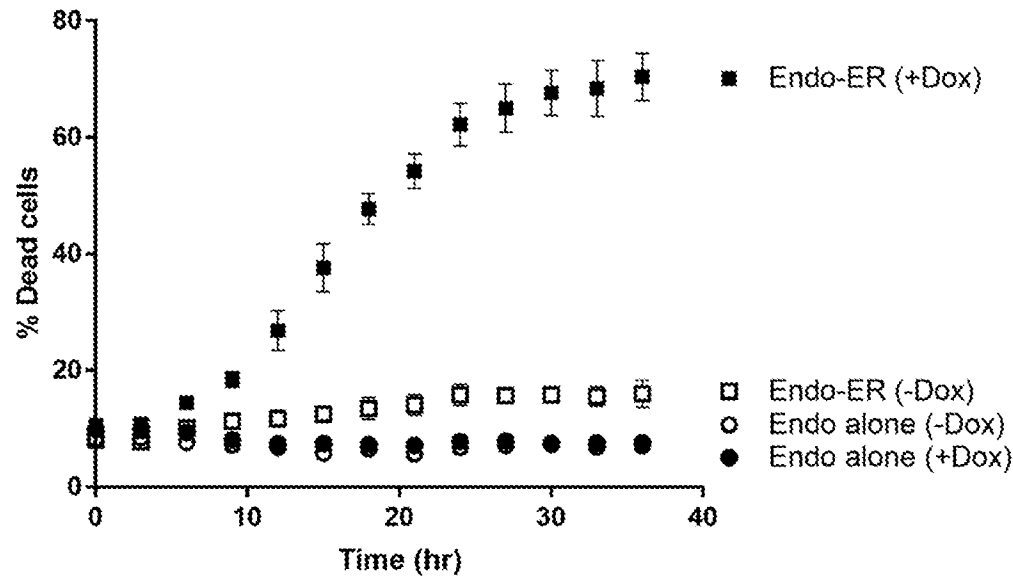


Figure 8

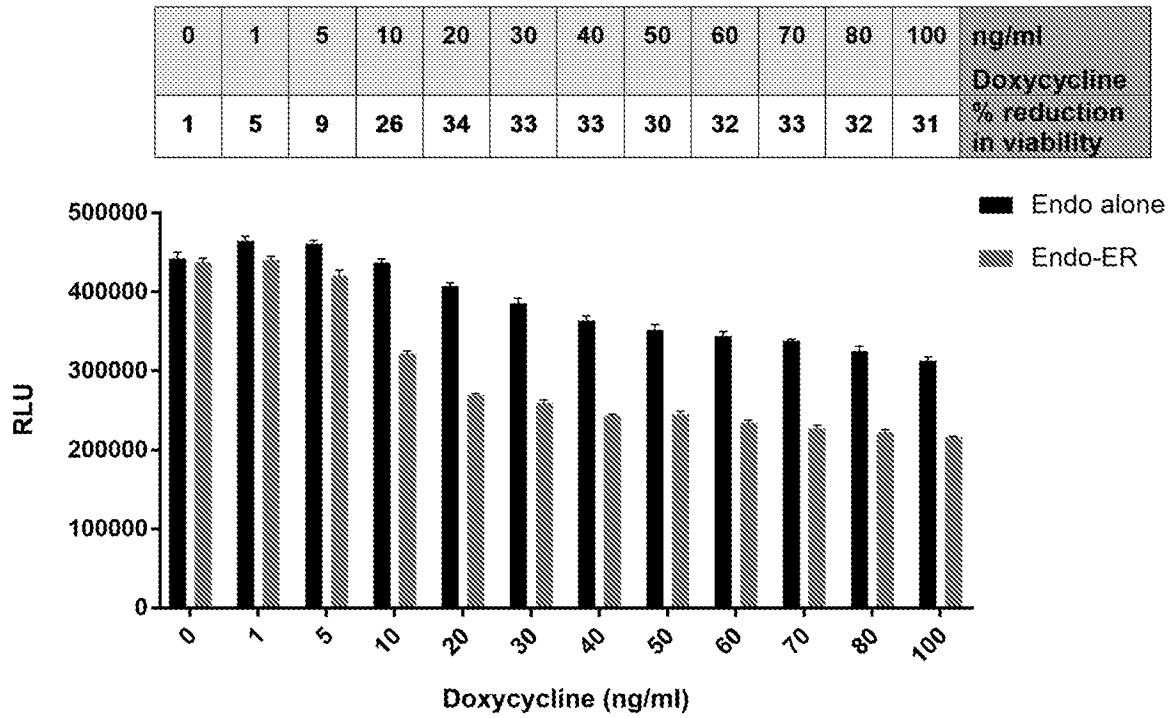


Figure 9

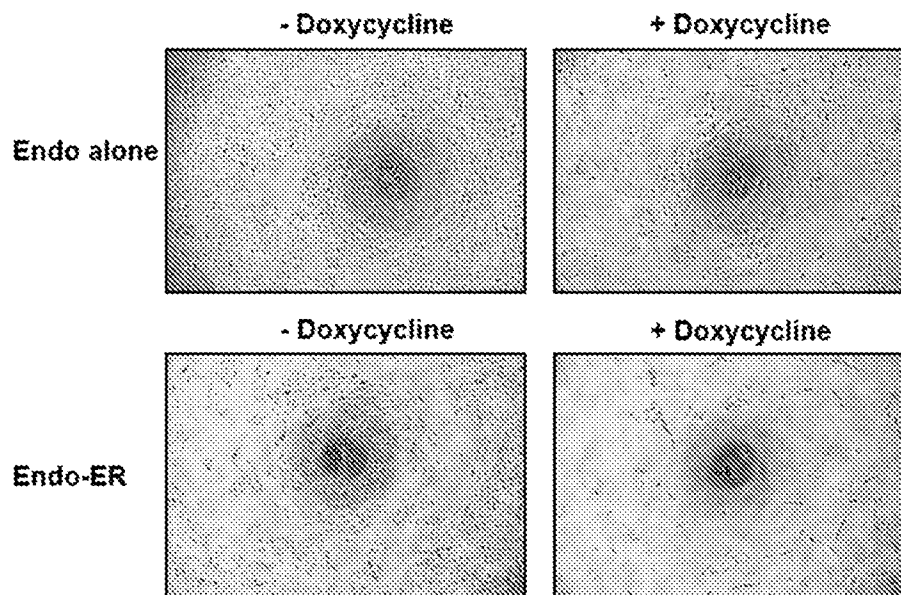


Figure 10

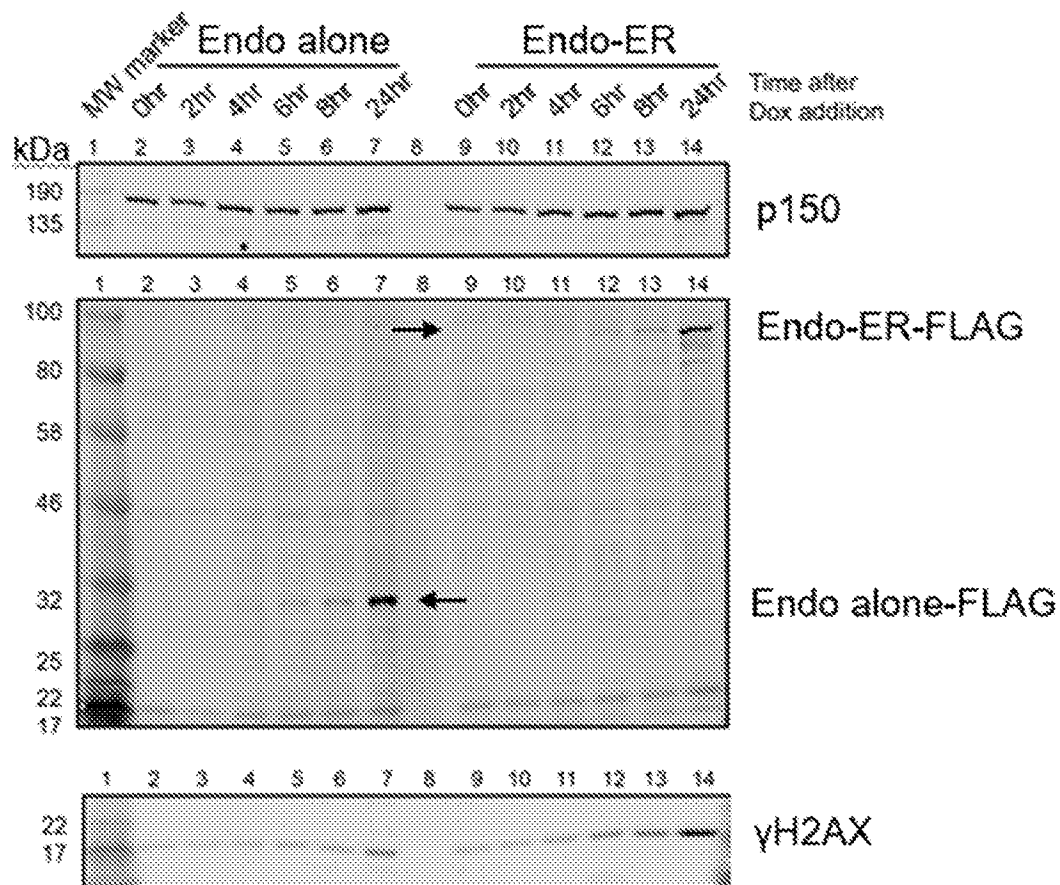


Figure 11

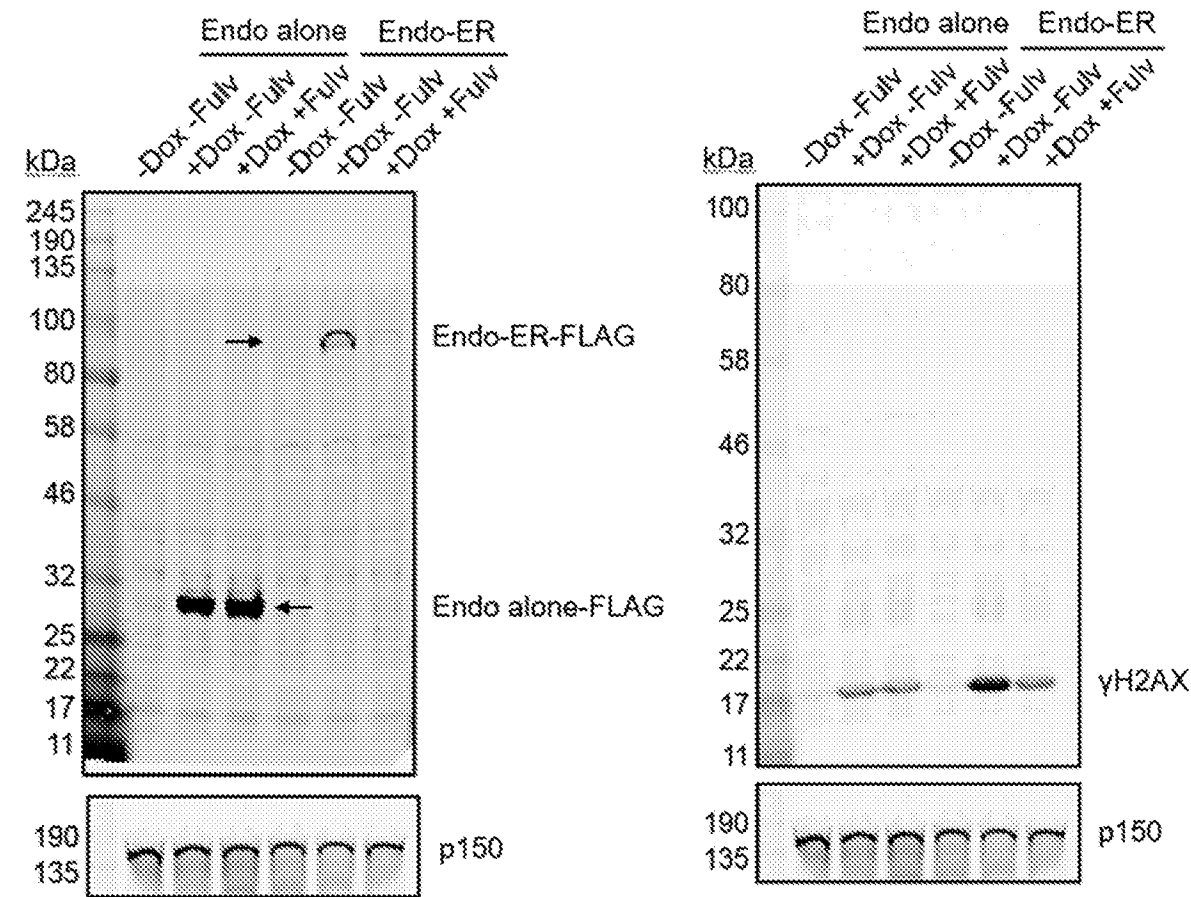


Figure 12

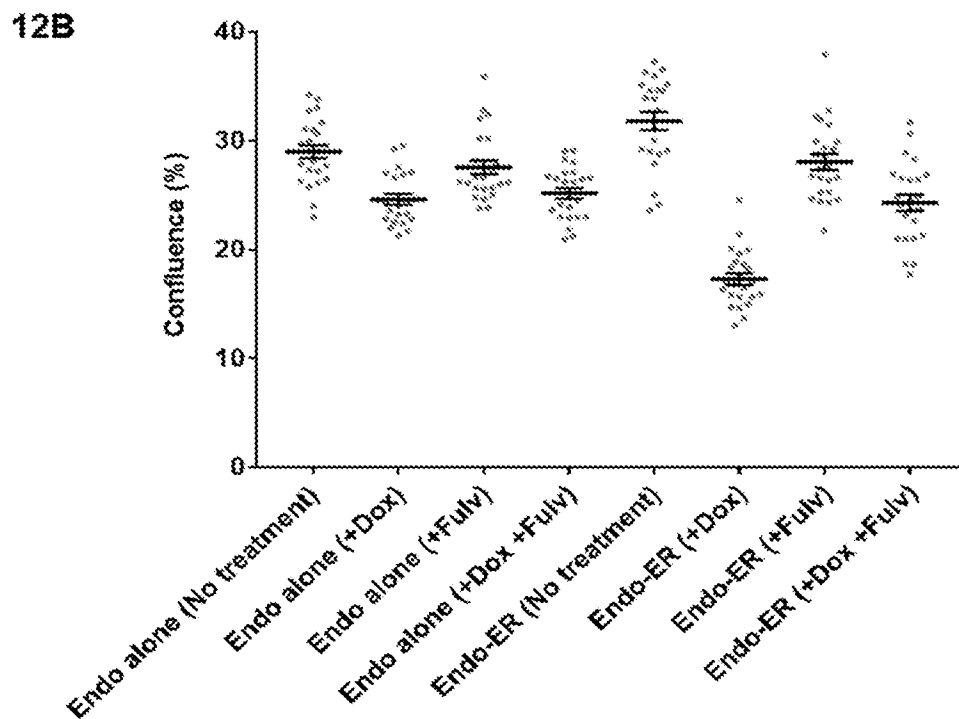
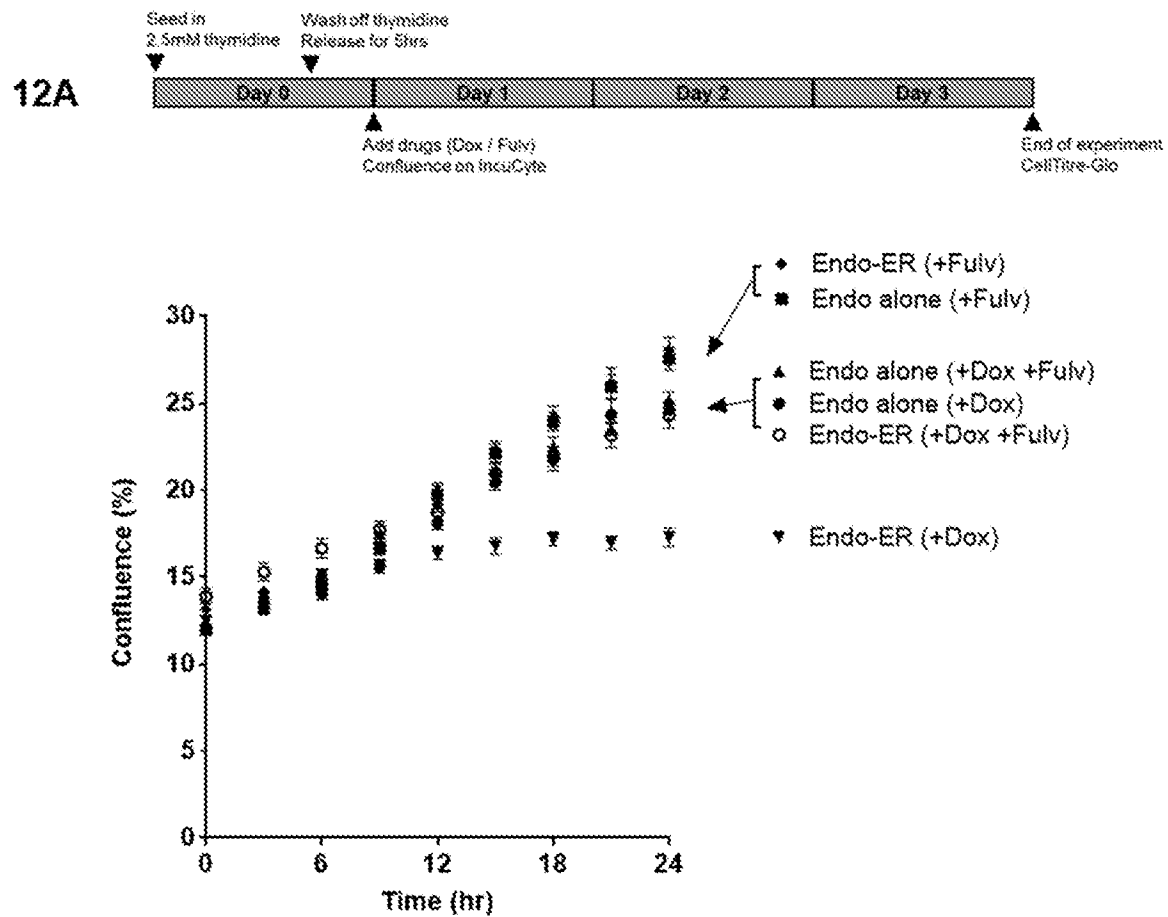
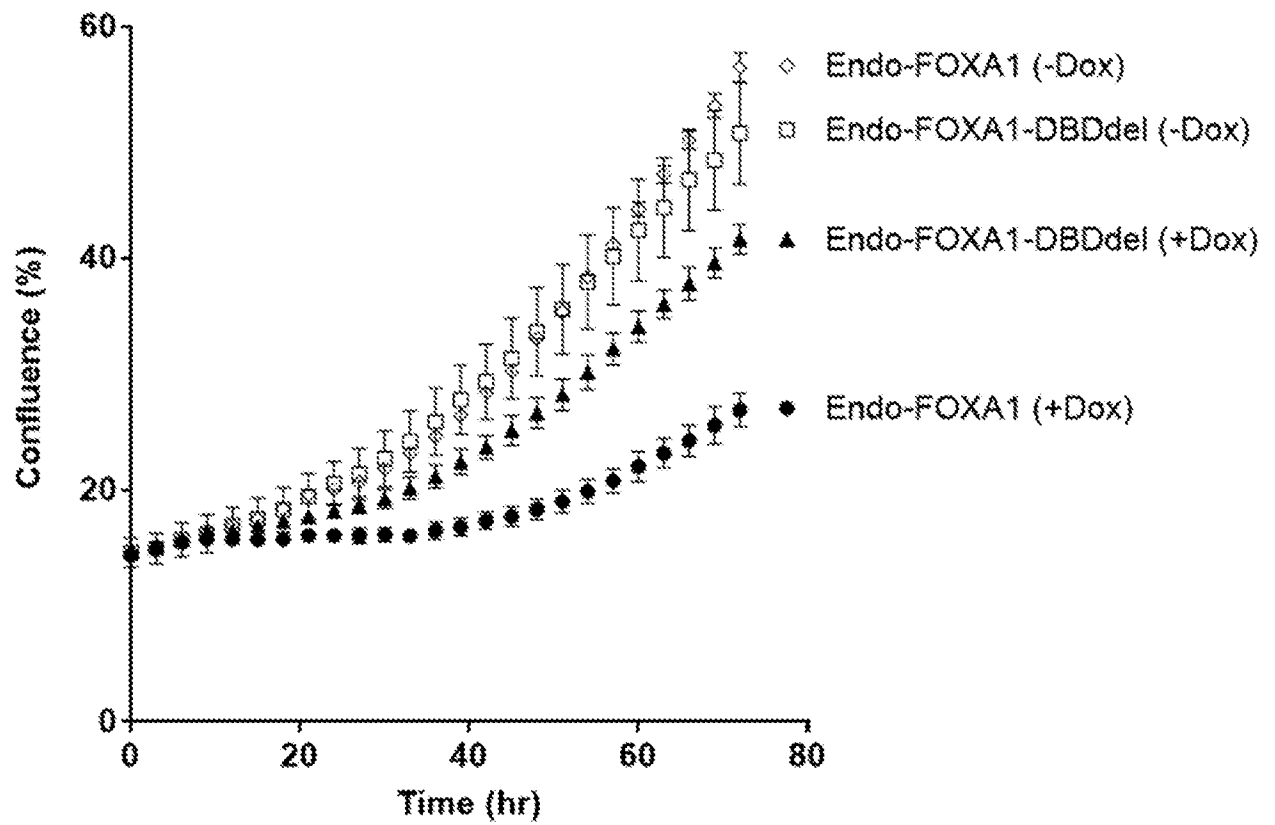


Figure 13



INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2020/051559

A. CLASSIFICATION OF SUBJECT MATTER

INV. G01N33/50 C12N9/22
ADD.

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)

G01N C12N C07K

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

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

EPO-Internal, WPI Data, BIOSIS, CHEM ABS Data, Sequence Search, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2014/121222 A1 (UNIV WESTERN ONTARIO [CA]; EDGELL DAVID R [CA] ET AL.) 7 August 2014 (2014-08-07) cited in the application	49
A	paragraphs [0061], [0079] - [0106] paragraphs [0110] - [0114] -----	1-47
X	WO 2011/064751 A1 (BASF PLANT SCIENCE CO GMBH [DE]; HLUBEK ANDREA [DE] ET AL.) 3 June 2011 (2011-06-03)	49
A	page 2, line 7 page 6, line 10 - page 30, line 30 -----	1-47
X	WO 2013/045480 A1 (UNIV GIESSEN JUSTUS LIEBIG [DE]) 4 April 2013 (2013-04-04)	49
A	page 3, lines 1-7 -----	1-47
	-/--	



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of 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

7 September 2020

Date of mailing of the international search report

15/09/2020

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Petri, Bernhard

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2020/051559

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HANSHUO ZHANG ET AL: "Engineered TAL Effector modulators for the large-scale gain-of-function screening", NUCLEIC ACIDS RESEARCH, vol. 42, no. 14, 17 June 2014 (2014-06-17), pages e114-e114, XP055724173, ISSN: 0305-1048, DOI: 10.1093/nar/gku535 page 3, left-hand column, paragraph 1 - page 6, right-hand column, paragraph 1 -----	1-47
A	CHENGJIAN MAO ET AL: "A New Small Molecule Inhibitor of Estrogen Receptor [alpha] Binding to Estrogen Response Elements Blocks Estrogen-dependent Growth of Cancer Cells", JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 283, no. 19, 9 May 2008 (2008-05-09), pages 12819-12830, XP055724339, US ISSN: 0021-9258, DOI: 10.1074/jbc.M709936200 page 12823, left-hand column, paragraph 1-3; figure 2 -----	1-47
A	NICOLE ALONSO ET AL: "A rapid and sensitive high-throughput screening method to identify compounds targeting protein-nucleic acids interactions", NUCLEIC ACIDS RESEARCH, vol. 43, no. 8, 30 April 2015 (2015-04-30), pages e52-e52, XP55724284, ISSN: 0305-1048, DOI: 10.1093/nar/gkv069 abstract figure 1 -----	1-47
A,P	LI QINGLIANG ET AL: "Inhibition of p53 DNA binding by a small molecule protects mice from radiation toxicity", ONCOGENE, NATURE PUBLISHING GROUP UK, LONDON, vol. 39, no. 29, 17 June 2020 (2020-06-17), pages 5187-5200, XP037194884, ISSN: 0950-9232, DOI: 10.1038/S41388-020-1344-Y [retrieved on 2020-06-17] page 5188, right-hand column, paragraph 2 abstract -----	1-47

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2020/051559

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2014121222	A1	07-08-2014	NONE
WO 2011064751	A1	03-06-2011	AU 2010325564 A1 12-07-2012
		BR 112012012444 A2	22-09-2015
		CA 2781835 A1	03-06-2011
		CN 102762726 A	31-10-2012
		DE 112010004584 T5	29-11-2012
		EP 2504430 A1	03-10-2012
		JP 2013511979 A	11-04-2013
		US 2012324603 A1	20-12-2012
		WO 2011064751 A1	03-06-2011
		ZA 201204697 B	25-09-2013
WO 2013045480	A1	04-04-2013	EP 2573173 A1 27-03-2013
		WO 2013045480 A1	04-04-2013