



A high-throughput platform to select for regulators of crispr-cas associated activity

Vazquez-Uribe, Ruben; Van Der Helm, Eric; Sommer, Morten Otto Alexander; Lee, Sang Woo

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Vazquez-Uribe, R., Van Der Helm, E., Sommer, M. O. A., & Lee, S. W. (2019). A high-throughput platform to select for regulators of crispr-cas associated activity. (Patent No. WO2019185750).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



(51) International Patent Classification:

C12N 15/10 (2006.01)

(21) International Application Number:

PCT/EP20 19/057785

(22) International Filing Date:

27 March 2019 (27.03.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

18164235.6 27 March 2018 (27.03.2018) EP

182093 15.3 29 November 2018 (29. 11.2018) EP

(71) Applicant: DANMARKS TEKNISKE UNIVERSITET
[DK/DK]; Anker Engelunds Vej 101 A, 2800 Kgs. Lyngby
(DK).

(72) Inventors: VAZQUEZ-URIBE, Ruben; Arnestedet 5,
2720 Vanlose (DK). VAN DER HELM, Eric; Istedgade
75, 1th, 1650 Kobenhavn V (DK). SOMMER, Morten
Otto Alexander; Duntzfelts Alle 22, 2900 Hellerup (DK).
LEE, Sang Woo; Buddingevej 281, 2860 Soborg (DK).

(74) Agent: GUARDIAN IP CONSULTING I/S; Diplomvej,
Building 381, 2800 Kgs. Lyngby (DK).

(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, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, 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,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

(54) Title: A HIGH-THROUGHPUT PLATFORM TO SELECT FOR REGULATORS OF CRISPR-CAS ASSOCIATED ACTIVITY

(57) Abstract: The invention relates to a method for in vivo detection of small molecules or polypeptides that act as regulators of CRISPR-Cas associated activity, in particular the method is for use in the screening of molecule libraries and non-host DNA libraries. The method employs cells comprising genes expressing CRISPR-Cas associated activity and a reporter protein to allow the detection of said regulators. The invention further relates to cells comprising genes expressing of CRISPR-Cas associated activity and a reporter protein for use in said in vivo screening and detection methods.



TITLE: A high-throughput platform to select for regulators of CRISPR-Cas associated activity

Technical field of the invention

5 The invention relates to a method for *in vivo* detection of small molecules or polypeptides that act as regulators of CRISPR-Cas associated activity, in particular the method is for use in the screening of molecule libraries and non-host DNA libraries. The method employs cells comprising genes expressing CRISPR-Cas associated activity and a reporter protein to allow the detection of said regulators. The invention further relates to cells comprising genes expressing of CRISPR-Cas associated activity
10 and a reporter protein for use in the *in vivo* screening and detection methods.

Background of the invention

15 CRISPR-Cas systems originate from bacteria where they function as adaptive immune systems that cleave foreign DNA or RNA in a sequence specific manner. Components of type II CRISPR-Cas systems, such as Cas9 from *Streptococcus pyogenes*, have been exploited for precise and programmable gene editing. Given the significant impact of CRISPR-Cas systems, efforts have been made to discover new systems in cultured and recently uncultured bacterial species. CRISPR-Cas technologies represent effective
20 biotechnological tools that broadly impact life sciences. Applications range from treating human genetic diseases to eliminating pathogens using gene-drives. Beyond genome editing, CRISPR systems have several other applications, including programmable transcriptional regulators and ultra-sensitive detection devices. However, there is increasing concern about the safety and precision of CRISPR-Cas. Indeed, the 2016 worldwide threat assessment of U.S. intelligence listed genome
25 editing technologies as a potential weapon of mass destruction and proliferation. Accordingly, more research and tools are needed to control the CRISPR-Cas technology in order to reduce potential risks of genome editing, or to contain their spread and potential harm to other organisms. A promising method to control CRISPR activity is via anti-CRISPR (ACR) proteins. These naturally occurring inhibitors can potentially inactivate CRISPR systems when
30 needed, or decrease off-target effects, without significantly affecting on-target efficiency. ACRs were initially discovered for type I CRISPR systems. Subsequent bioinformatic mining identified ACR activity against type II-C Cas9 from *Neisseria meningitidis* and type II-A Cas9 from *Streptococcus pyogenes* (spCas9). More recently an additional ACR against spCas9 was identified by cloning and testing multiple genes
35 from a phage that was able to escape CRISPR-based immunity from *Streptococcus*

thermophilus, reflecting the biological arms race occurring between phages and CRISPR systems. Given the abundance of CRISPR systems in bacteria as well as the abundance of uncharacterized phages, it is likely that we currently have only revealed a minute proportion of ACRs in the environment. Discovery of novel genes with this functionality is, however, complicated both by a lack of abundant reference ACR families as well as the limited availability of phage genomic information. Accordingly, there exists a need for better methods and genetic tools for identifying further ACRs, both those found in nature as well as small molecules exhibiting the functional properties of ACRs.

Summary of the invention

A first embodiment of the invention provides a method for in vivo detection of regulators of CRISPR-Cas associated activity comprising:

a. providing cells capable of inducible CRISPR-Cas associated activity wherein said cells are prepared on a defined medium devoid of one or more agents for inducing expression of said activity, and either

I. contacting said cells with a molecule or library of molecules; or

II. introducing a self-replicating library of non-host DNA fragments into said cells;

b. culturing the cells obtained from step (a) (I) or (II) on said defined culture medium for a period sufficient for uptake or expression of a candidate regulator of CRISPR-Cas associated activity;

c. culturing the cells obtained from step (b) on said defined culture medium wherein said medium is supplemented with one or more agents to induce expression of said CRISPR-Cas associated activity, and

d. detecting expression of a reporter gene present in cells obtained from step (c), wherein said cells comprise:

i. a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said first gene or first gene cluster is operably linked a promoter having a measured strength of >0.65 on the Anderson scale and whose expression is inducible by at least one of said one or more agents;

ii. a reporter gene operably linked to a promoter;

iii. a nucleic acid molecule operably linked to a promoter and whose expression is inducible by at least one of said one or more agents,

wherein said nucleic acid molecule encodes one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said reporter gene or its transcript; and

wherein the polypeptide or polypeptide complex having RNA-guided endonuclease activity is capable of inactivating said reporter gene or its transcript.

A second embodiment of the invention provides a biological cell comprising :

- 5 i. a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said first gene or first gene cluster is operably linked to a promoter having a measured strength of >0.65 on the Anderson scale and whose expression is inducible by at least one agent;
- 10 ii. a reporter gene operably linked to a promoter,
- iii. a nucleic acid molecule operably linked to a promoter and whose expression is inducible by at least one agent,

wherein said nucleic acid molecule encodes one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said reporter gene or its transcript; and

15

wherein the polypeptide or polypeptide complex having RNA-guided endonuclease activity is capable of inactivating said reporter gene or its transcript.

20 A third embodiment of the invention provides for the use of a biological cell according to the invention, for detecting a regulator of CRISPR-Cas associated activity.

A fourth embodiment of the invention provides a biological cell library of non-host DNA fragments, wherein each cell of the library is a cell according to the invention, and wherein each cell further comprises a self-replicating genetic element comprising

- 25 a) a non-host DNA fragment, and
- b) a gene encoding a selectable marker polypeptide and operably linked to a promoter.

30 A fifth embodiment of the invention provides for a use of a biological cell library of non-host DNA fragments according to the invention, for detecting a regulator of CRISPR-Cas associated activity.

Description of the invention

Figures:

35 **Figure 1** is a cartoon illustrating the method of the invention ; comprising providing a biological cell comprising a gene or gene cluster encoding a protein or protein complex having RNA-guided endonuclease activity, a nucleic acid molecule encoding chimeric gRNA (sgRNA) or a CRISPR array, as well as a reporter gene encoding a selection marker, and then either contacting the cell with a molecule, or introducing a self-

replicating library of non-host DNA into the cell, such as to allow detection of molecules or expressed polypeptides that inhibit inactivation of the reporter gene or its transcript by the RNA-guided endonuclease.

5 **Figure 2** is a cartoon illustrating the pCasens3 and pDual3 plasmids and interaction with a non-peptide molecule with anti-CRISPR activity. Plasmid pCasens3 (to the left) comprises : the *Streptococcus pyogenes* Cas9 gene (spCas9) [SEQ ID No.:150] cloned into the backbone of the plasmid pSEVA47. Additionally, a sigma70 constitutive promoter (P23100) [SEQ ID No.:148], adjacent to a DNA molecule encoding a theophylline translational riboswitch [SEQ ID No.:149] were inserted upstream of the spCas9 gene in pSEVA47, in order to control its expression. pSEVA47 comprises a low copy number origin of replication, SC101 [SEQ ID No.:155]; and the antibiotic resistance gene aadA [SEQ ID No.:153] under the control of its native (constitutive) promoter and ribosomal binding site, conferring resistance against Spectinomycin.

10

15 Plasmid pDual3 (to the right) comprises a chloramphenicol resistance gene (CmR) [SEQ ID No.:164] operably linked to an upstream constitutive promoter and RBS; and a DNA sequence encoding a sgRNA [SEQ ID No.:161] whose transcription/expression is regulated by the inducible pBAD promoter [SEQ ID No.: 158], located upstream of the respective sequences. The plasmid further comprises an araC L-arabinose sensor gene [SEQ ID No.:159] that regulates the pBAD promoter in response to the metabolite, L-arabinose. pDual3 comprises a low copy number origin of replication, p15A [SEQ ID No.:163].

20

Figure 3 is a cartoon illustrating (A) the negative control plasmid comprising a gene [SEQ ID No.:140] encoding Green Fluorescent Protein (GFP) under the control of a constitutive Ptet promoter [SEQ ID No.:139]; and (B) positive control plasmid comprising a gene [SEQ ID No.:146] encoding AcrIIA2 protein under the control of a constitutive Ptet promoter [SEQ ID No.:139]. Both negative and positive control plasmids further comprise a kanamycin resistance gene [SEQ ID No.:144] under the control of its native (constitutive) promoter and ribosomal binding site and a colE1 origin of replication [SEQ ID No.:143].

25

30

Figure 4 is a histogram showing the number of colony forming units (CFU.mL⁻¹) of *E. coli* cells comprising the pDual3 plasmid in combination with 3 alternative versions of the pCasens3 plasmid; following cultivation on selective media comprising chloramphenicol, where the sgRNA and Cas9 genes of the genetic circuit are induced as indicated with arabinose and theophylline respectively. The 3 forms of pCasens3 comprise a Cas9 gene operably linked to either a medium strength promoter of <0.50 on the Anderson scale (ProC); the medium strength promoter (ProC) together with a

35

theophylline riboswitch 5' of the cas9 gene; or a strong promoter (BBa J23100 [SEQ ID No.:172]) together with a theophylline riboswitch 5' of the cas9 gene.

Figure 5 shows (A) a cartoon illustrating the molecular mechanism whereby in *E. coli* cells comprising the Genetic Circuit (*E. coli*-GC), the genetic circuit is induced to express the cas9 gene and sgRNA, the formation of a Cas9-sgRNA complex, and the endonuclease-mediated cleavage of the reporter gene (Cat), leading to cell death when exposed to chloramphenicol. The cartoon further shows that co-expression of an *acrIIA2* gene in the *E. coli*-GC cells can inhibit Cas9-sgRNA inactivation of the Cat gene, leading to cell survival under selective conditions. (B) is a histogram showing the number of colony forming units (CFU.ml⁻¹) of *E. coli*-GC cells comprising a negative control plasmid (expressing GFP) or the positive control plasmid (expressing *AcrIIA2*) following cultivation on selective media, where the genetic circuit is either induced or non-induced.

Figure 6 is a cartoon illustrating (A) a metagenomics DNA expression plasmid comprising a fragment of (a library of) metagenomics DNA fragments cloned downstream of a constitutive Ptet promoter [SEQ ID No.:139]; a kanamycin resistance gene under the control of its native (constitutive) promoter and ribosomal binding site; together with the genetic circuit comprising plasmids pCasens3 and pDual3; (B) a positive control plasmid comprising a gene encoding *AcrIIA2* protein and the genetic circuit plasmids pCasens3 and pDual3.

Figure 7 is a cartoon showing use of host cells comprising the genetic circuit to screen a metagenomics DNA expression library for anti-CRISPR proteins.

Figure 8 is a histogram showing the number of colony forming units (CFU.ml⁻¹) of *E. coli*-GC cells comprising selected metagenomics inserts, designated AC11 - AC42; as compared to a negative control plasmid (expressing GFP) or the positive control plasmid (expressing *AcrIIA2*) following cultivation on selective media, where the genetic circuit is induced.

Figure 9 is a histogram showing the number of colony forming units (CFU.ml⁻¹) of *E. coli*-GC cells comprising plasmids pTypeI-B-Cas in combination with pDual3; following sequential cultivation in growth media that are either non-inducing or inducing for TypeI-B-Cas expression; and then on either non-selective or selective media comprising chloramphenicol, wherein TypeI-B-Cas expression was induced using 0 - 100µM IPTG.

Figure 10 is a histogram showing the number of colony forming units (CFU.ml⁻¹) of *E. coli*-GC cells comprising plasmids pTypell-A-Cas in combination with pDual3; following sequential cultivation in growth media that are non-inducing or inducing for Typell-A-Cas expression; and then on either non-selective or selective media comprising chloramphicol, wherein Typell-A-Cas expression was induced using 0 - 100µM IPTG.

Figure 11 is a histogram showing the number of colony forming units (CFU.ml⁻¹) of *E. coli*-GC cells comprising plasmids pTypeV-A-Cas in combination with pDual3; following sequential cultivation in growth media that are non-inducing or inducing for TypeV-A-Cas expression; and then on either non-selective or selective media comprising chloramphicol, wherein TypeV-A-Cas expression was induced using 2mM theophylline.

Figure 12 is a cartoon showing a method for functional screening for anti-CRISPR molecules using biological cells comprising the genetic circuit according to the invention that is suitable for screening: a metagenomics library encoding candidate anti-CRISPR proteins; an isolated candidate anti-CRISPR protein; and an isolated anti-CRISPR small molecule. The method comprises 6 steps:

1) Cultivation of biological cells, comprising the genetic circuit according to the invention, on a defined growth medium. The medium comprises nutrients essential for growth and when required one or more agents selective for maintenance of genes of the genetic circuit including the reporter gene, but devoid of agents required for, or capable of, inducing expression of the genetic circuit (i.e. the genes or gene clusters encoding an RNA-guided endonuclease and its cognate RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said reporter gene or its transcript).

2) An aliquot of the cell culture produced in step 1) is inoculated into said defined growth medium to produce a cell suspension; but wherein said medium is both devoid of agents required for, or capable of, inducing expression of said genetic circuit; and devoid of agents selective for maintenance of genes of the genetic circuit including the reporter gene;

3) cells in the cell suspension produced in step 2) are either:

i) transformed with a metagenomic library to produce a transformed cell suspension, or

ii) contacted with a candidate anti-CRISPR small molecule or candidate anti-CRISPR protein by their addition to said cell suspension;

4) the cell suspension resulting from:

- step 3) i) is incubated to allow expression of a candidate anti-CRISPR protein encoded by a DNA molecule in the metagenomic library; or from

- step 3) ii) is incubated to allow intracellular uptake of the candidate anti-CRISPR small molecule or candidate anti-CRISPR protein;

5) Agent(s) capable of inducing expression of said genetic circuit are added to the cell suspension(s) resulting from step 4; and the suspensions are further incubated under conditions suitable for inducing expression of genes of the genetic circuit;

6) Cells derived from the cell suspension(s) resulting from step 5) are plated on agar plates composed of a defined media comprising nutrients essential for growth and supplemented with agents required for, or capable of, inducing expression of said genetic circuit and as required agents selective for maintenance of genes of the genetic circuit including the reporter gene. The plates are incubated under suitable conditions for a time sufficient to detect viable biological cells.

Figure 13 is a histogram showing the number of colony forming units (CFU.ml⁻¹) of *E. coli*-GC cells comprising a CRISPR-Cas genetic circuit consisting of either: (1) pCasens3 (SP(II-A)) and pDUAL3; (2) pTypeV-A-Cas (FN(V-A)) and pDual3; (3) pTypeII-A-Cas (STI(II-A)) and pDual3; or (4) pTypeI-B-Cas (CD(I-B)) and pDual3; following incubation on liquid medium comprising either M9 minimal medium or yeast extract medium (2xYT); each medium with or without the addition of agents to induce expression of the respective CRISPR-Cas genetic circuit, and then respectively plated on agar medium comprising either M9 minimal medium or yeast extract (2xYT); each medium comprising chloramphenicol, and with or without addition of said inducing agents.

Figure 14 is a histogram showing the the number of colony forming units (CFU.ml⁻¹) of *E. coli* cells comprising the pDual3 plasmid in combination with 3 alternative versions of the pCasens3 plasmid following culture in the presence of chloramphenicol and in the presence, or absence, of arabinose and theophylline to simultaneously induce expression of the sgRNA and Cas9 genes respectively. The 3 forms of pCasens3 plasmid comprise a Cas9 gene operably linked to either a weak promoter (J23116 [SEQ ID No.:176]) shown in the left-hand bar; medium promoter (J23111 [SEQ ID No.:177]) shown in the central bar; or strong promoter (J23100 [SEQ ID No.:172]) shown in the right-hand bar; having increasing promoter strengths of 0.16; 0.58 and 1.00 on the Anderson scale, each combined with a theophylline riboswitch 5' of the cas9 gene.

Abbreviations and definition of terms:

Agent: is a small molecule capable of inducing expression of one or more genes operably linked to a promoter in the genetic circuit of the invention.

Amino acid sequence identity : The term "sequence identity" as used herein, indicates a quantitative measure of the degree of homology between two amino acid sequences of substantially equal length. The two sequences to be compared must be aligned to give a best possible fit, by means of the insertion of gaps or alternatively, truncation at the ends of the protein sequences. The sequence identity can be calculated as $((N_{ref} - N_{dif}) / 100) / (N_{ref})$, wherein N_{dif} is the total number of non-identical residues in the two sequences when aligned and wherein N_{ref} is the number of residues in one of the sequences. Sequence identity can alternatively be calculated by the BLAST program e.g. the BLASTP program (Pearson W.R and D.J. Lipman (1988)) (www.ncbi.nlm.nih.gov/cgi-bin/BLAST). In one embodiment of the invention, alignment is performed with the sequence alignment method ClustalW with default parameters as described by Thompson J., et al 1994, available at <http://www2.ebi.ac.uk/clustalw/>.

Preferably, the numbers of substitutions, insertions, additions or deletions of one or more amino acid residues in the polypeptide as compared to its comparator polypeptide is limited, i.e. no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 substitutions, no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 insertions, no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 additions, and no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 deletions. Preferably the substitutions are conservative amino acid substitutions : limited to exchanges within members of group 1: Glycine, Alanine, Valine, Leucine, Isoleucine; group 2: Serine, Cysteine, Selenocysteine, Threonine, Methionine; group 3: proline; group 4: Phenylalanine, Tyrosine, Tryptophan; Group 5: Aspartate, Glutamate, Asparagine, Glutamine.

Cis element: is a sequence of DNA that is capable of conferring inducible expression of a gene that is functionally linked to a promoter; where the cis element may induce expression by regulating transcription of the gene or by regulating translation of transcript product of the gene.

Cognate: is used to refer to interacting pairs of functional entities; more specifically that each protein or protein cluster having RNA-guided endonuclease activity has a cognate mat-crRNA, and in the case of Type II CRISPR-Cas systems both a cognate mat-crRNA and cognate tracrRNA, that are required for their activity.

CRISPR system: is a Clustered Regularly Interspaced Short Palindromic Repeats is bacterial immune system that cleave DNA or RNA in a sequence specific manner, mediated by a RNA guided nuclease.

***E. coli* TopIO** : is *E. coli* having chromosomal Genotype mcrA, A(mrr-hsdRMS-mcrBC), Phi80lacZ(del)M15, AlacX74, deoR, recAI, araD139, A(ara-leu)7697, galU, galK, rpsL(SmR), endAI, nupG

gi number : (genInfo identifier) is a unique integer which identifies a particular sequence, independent of the database source, which is assigned by NCBI to all

sequences processed into Entrez, including nucleotide sequences from DDBJ/EMBL/GenBank, protein sequences from SWISS-PROT, PIR and many others.

Heterologous gene and heterologous DNA molecule: have a different genetic origin from the recombinant biological cell in which they are expressed; and this also applies to the protein or transcript that they encode. The nucleotide sequence of the heterologous gene or heterologous DNA molecule may be optimized (e.g. codon optimization) with respect to the recombinant biological cell in which they are expressed. Heterologous gene and heterologous DNA molecule may be located on (and therefore be a part of) the chromosome or a episome of the recombinant cell, and may be inserted into this location by recombinant DNA cloning.

Promoter: is a region of DNA that initiates transcription of a particular gene. Promoters are located near the transcription start sites of genes, on the same strand and upstream on the DNA (towards the 5' region of the sense strand). A promoter that is "functionally linked" to a gene is capable of driving expression of said gene. A promoter may be drive expression constitutively or only in response to an inducer.

Promoter activity: is the measured strength i.e. measured relative activity of a promoter to drive expression of a reporter gene encoding Red Fluorescent Protein (RFP) in *E. coli*. Measurement is performed on cells of *E. coli* strain TGI [K-12 supE thi-1 A(lac-proAB) A(mcrB-hsdSM)5, (rK-m K-)] comprising the plasmid J61002 (as described in http://parts.igem.org/Part:BBa_J61002) in which the tested promoter linked to the RFP reported gene is cloned, and where the cells are grown in LB media to saturation. The range of measured promoter strengths is relative to the Anderson Promoter Collection Scale (Anderson scale) of 0.00 to 1.00, wherein promoter J23100 [SEQ ID No.: 172] has the highest measured relative strength of 1.00; and the promoter J23112 [SEQ ID No.: 171] has a measured relative strength of 0.00, corresponding to non-detectable promoter activity (as described by <http://parts.iaem.org/Promoters/Catalog/Anderson>). As defined herein, a strong promoter is one having a measured strength of >0.65 on this Anderson scale.

RBS: Ribosomal Binding Site is a sequence of nucleotides upstream of the start codon of an mRNA transcript that is responsible for the recruitment of a ribosome during the initiation of protein translation.

Small molecule library: library of small organic molecules; preferably having a mol mass of less than 900 daltons; for example libraries available from Asinex (www.asinex.com).

Detailed description of the invention

A first embodiment of the invention provides a method for screening a molecule or library of molecules (in particular small molecules) or polypeptides for a candidate

molecule or polypeptide capable of regulating CRISPR-Cas associated activity. The regulator may be an inhibitor or an enhancer of CRISPR-Cas associated activity. The method is based on an *in vivo* assay employing cells; where the cells comprise genes encoding and capable of expressing the essential components required for CRISPR-Cas associated activity (Figure 1). The method of the invention has broad application, since it can be used for screening for regulators of all types of CRISPR-Cas systems having CRISPR-Cas associated activity.

While CRISPR-Cas systems, encoded by a CRISPR locus, find their origin in adaptive immune systems in a wide range of archaea and bacteria, they share a common functional property of cleaving a reporter DNA molecule or its transcript; and thereby allowing genetic material to be added, removed, or altered at particular locations in the genome of a biological cell or in a DNA molecule. In general terms, a CRISPR locus consists of a CRISPR array, comprising short variable DNA sequences (called spacer sequence(s)) interspersed by short palindromic repeat sequences (called repeat sequence(s)), this being flanked by diverse cognate cas genes. The CRISPR-Cas systems have been classified into 2 classes; whereby members of Class 1 comprise a multi-subunit subunit protein complex, each subunit encoded by a gene in the CRISPR locus; while members of Class 2 comprise a single multidomain protein encoded by a single gene. Class 1, includes type I, III and IV CRISPR-Cas systems; while Class 2, includes type II, V and VI CRISPR-Cas systems. Thus, in total there are six types, which have further been classified into over 28 subtypes; based on their protein structure, signature Cas genes and operon arrangements (Koonin, E. V. et al., 2017).

More specifically: the loci comprising Class 1, type I CRISPR-Cas systems share in common a signature gene *cas3* (or its variant *cas3'*) encoding a nuclease that both unwinds and cleaves target double-stranded DNA (dsDNA) and RNA-DNA duplexes. Type I systems comprise seven subtypes, I-A to I-F, and I-U. The loci comprising Class 1, type III CRISPR-Cas systems all contain the signature gene *cas10*, as well genes encoding *Cas5* gene and several paralogues of *Cas7* that co-transcriptionally target RNA and DNA. Type III systems comprise four subtypes, III-A to III-D. The loci comprising Class 1, type IV CRISPR-Cas system comprise two subtypes, IV-A and IV-B; where the *Csfl* can serve as as signature gene.

The loci for the Class 2, type II CRISPR-Cas systems, comprise the signature *Cas9* gene encoding a multidomain protein that combines the functions of a crRNA-effector complex and cleaves target DNA. All Type II loci are characterised by a nucleic acid sequence encoding a tracrRNA, which is partially complementary to the repeats within the respective CRISPR array. Type II systems comprise three subtypes, II-A to II-C. Class 2, type V CRISPR-Cas systems comprise the signature gene *Cpfl*, also encoding

a single multidomain protein. Unlike Cas9, the Cpf1 gene is commonly located outside a CRISPR-Cas locus. Type V systems are currently divided into five subtypes, V-A to V-E.

5 For the purpose of expression of members of the Class 2 type II CRISPR-Cas systems in a cell, it is sufficient to express the signature Cas gene encoding the multidomain protein having RNA guided-endonuclease activity. Expression of all other Class 1 and 2 CRISPR-Cas systems can be achieved by co-expressing the several genes encoding the
10 Cas proteins that constitute the multi-subunit protein complex having RNA guided-endonuclease activity.

When expression of the CRISPR locus is activated, transcription of the CRISPR array yields a pre-crRNA, that is processed into individual mature crRNAs (mat-crRNA). Each mat-crRNA sequence, aided by its cognate single or multi-subunit Cas protein(s)
15 (effector complex), functions as a guide to specifically target the effector complex to cleave a reporter nucleic acid molecule (DNA or RNA). Processing of the pre-crRNA is mediated by either an endonuclease subunit of the effector complex; or in the case of type II CRISPR-cas system by a combination of a RNase III protein together with a trans-activating CRISPR RNA (tracrRNA); where tracrRNA serves to recruit the
20 components of the type II effector complex.

For the purpose of expressing of any member of the Class 1 or 2 CRISPR-Cas systems in a cell, an individual mat-crRNA comprising at least one spacer sequence, flanked at least at one end (preferably at both ends) by a palindromic repeat sequence, is
25 sufficient to specifically guide the cognate effector complex to cleave a reporter nucleic acid molecule (DNA or RNA). Alternatively, a complete CRISPR array (pre-crRNA) can be expressed. The reporter nucleic acid molecule comprises a sequence that is complementary to the spacer sequence. A tracrRNA is additionally necessary of the expression of a type II CRISPR-cas system. Synthetic DNA molecules can conveniently
30 be engineered to express a RNA transcript comprising a spacer sequence flanked, at least at one end, by a palindromic repeat sequence (i.e. a mat-RNA), for use in the expression of the Class 1 or 2 CRISPR-Cas systems. However, in the case a type II CRISPR-cas system, the RNA transcript of the synthetic DNA molecule can comprise a a mat-RNA fused with a tracrRNA sequence, also called single-guide RNA (sgRNA).

35 A second embodiment of the invention provides cells for use in screening molecules or polypeptides (or a library thereof) for a candidate molecule or polypeptide capable of regulating CRISPR-Cas associated activity (Figure 1). The cells, employed in the method of screening, comprise a genetic circuit designed for detecting molecules or

polypeptides for regulators (inhibitors or enhancers) of CRISPR-Cas associated activity. The genetic circuit is engineered to operate in a host biological cell and provide an *in vivo* based screen for molecules or polypeptides having anti-CRISPR associated activity. In general terms, the genetic circuit comprises genes or cluster of

5 genes encoding a RNA-guided CRISPR-Cas system and its cognate RNA(s); wherein said CRISPR-Cas system comprises either a multi-subunit subunit protein complex (Class 1), each subunit being encoded by a gene derivable from (or present in) the corresponding native CRISPR locus; or a single multidomain protein (Class 2) encoded

10 by a single gene. The genetic circuit further comprises a reporter gene that functions as a marker for CRISPR activity. Each gene or gene cluster of the genetic circuit, present in a cell, is functionally-linked to a promoter and RBSs. The biological cell may further comprise a cis element (DNA) in combination with the promoter for regulating the expression of one or more of the genes or gene clusters of the genetic circuit. Each

15 of said one or more of said genes or gene clusters of the genetic circuit may be located on an episome or on the cell's chromosome. Preferably, the genes and DNA molecules encoding the components of the Class 1 or 2 CRISPR-Cas systems and their respective promoters are cloned into a suitable host cell; more preferably integrated into plasmids. These genes and DNA molecules may be heterologous in origin with respect to the host cell; and are preferably heterologous with respect to their cognate

20 promoter. The host biological cell (recombinant host cell) is preferably a bacterial cell.

When the genetic circuit encoding RNA-guided CRISPR-Cas system and its cognate RNA(s) is expressed in the biological cell, the cognate RNA serves to guide the system to a reporter gene, such that its endonuclease activity cleaves and inactivates the

25 reporter gene or its transcript. Loss of expression of the inactivated reporter gene is detectable in a cell comprising the genetic circuit. Correspondingly, a molecule or polypeptide that is capable of preventing inactivation of the reporter gene in a cell comprising the genetic circuit is detectable, either by means of detection of a product of the reporter gene itself, or by the survival and growth of cells that continue to

30 express the reporter gene under selective growth conditions. Cells comprising the genetic circuit are both suitable for screening molecules entering the cell, as well as polypeptides that are recombinantly expressed within the cell. Non-host cell DNA libraries encoding polypeptides can be introduced into recombinant cells comprising the genetic circuit and rapidly screened for regulators of CRISPR-Cas associated

35 activity.

I Method for *in vivo* detection of small molecules or polypeptides that act as regulators of CRISPR-Cas associated activity

The invention according to the first embodiment provides a method for *in vivo* detection of regulators of CRISPR-Cas associated activity preferably comprising the steps of:

5

- a. providing recombinant biological cells having inducible CRISPR-Cas associated activity wherein said cells are prepared on a defined medium devoid of one or more agents for inducing expression of said activity, and either
 - I. contacting each of one or more aliquots of said recombinant biological cells with a small molecule or an individual molecule of a small molecule library; or
 - II. introducing a self-replicating library of non-host DNA fragments into said recombinant biological cells;
- b. culturing the cells obtained from step (a) (I) or (II) on said defined culture medium for a period sufficient for uptake or expression of said a candidate regulator of CRISPR-Cas associated activity;
- c. culturing the cells obtained from step (b) on said defined culture medium wherein said medium is supplemented with one or more agents to induce expression of said CRISPR-Cas associated activity; and
- d. detecting expression of a reporter gene present in cells obtained from step (c),

10

15

20

wherein said recombinant biological cells comprise:

25

30

35

- i. a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said gene or gene cluster is operably linked a first inducible promoter having a measured strength of >0.65 on the Anderson scale, or a combination of a first promoter having a measured strength of >0.65 on the Anderson scale and a first cis element conferring inducible expression, wherein said inducible expression is induced by at least one of said one or more agents;
- ii. a reporter gene operably linked to a constitutive promoter;
- iii. a nucleic acid molecule encoding any one of:
 - o a pre-crRNA molecule (encoded by a DNA molecule comprising a CRISPR array) wherein said pre-crRNA comprises a spacer sequence complementary to a nucleic acid sequence of at least 15 nucleotides of at least one strand of said reporter gene or its transcript;
 - o a mat-crRNA molecule comprising at least one spacer sequence flanked by at least one palindromic repeat sequence, wherein the

spacer sequence is complementary to a nucleic acid sequence of at least 15 nucleotides of at least one strand of said reporter gene or its transcript;

5 o the pre-crRNA molecule and a trans-activating crRNA; or

 o the mat-crRNA molecule and a trans-activating crRNA (or a fusion thereof),

wherein said nucleic acid molecule is operably linked a second inducible promoter having a measured strength of >0.65 on the Anderson Scale, or a combination of a second promoter and second cis element conferring inducible expression, wherein said inducible expression is induced by at least one of said one or more agents;

10

and

wherein the palindromic repeat sequence and the trans-activating crRNA are cognate with respect to the polypeptide or polypeptide complex having RNA-guided endonuclease activity; and

15

wherein said RNA-guided endonuclease activity is capable of inactivating said heterologous reporter gene or its transcript.

Preferably, said reporter gene is comprised on a plasmid devoid of genes capable of expressing CRISPR-Cas associated activity.

20

The first gene or first gene cluster encoding a polypeptide or polypeptide complex, respectively having RNA-guided endonuclease activity, may be selected from any gene or gene cluster encoding cas protein(s) having CRISPR associated activity. Examples of cas proteins or cas protein complexes, suitable for expression in recombinant cells of the invention include those belonging to the Type 1, Type II, Type III; Type IV, Type V, and Type VI and their subtypes. In most cases the first gene or first gene cluster and its respective encoded polypeptide or polypeptide complex will be heterologous in origin with respect to both the promoter to which it is operably linked, and to the recombinant biological cell in which it is expressed. Examples of the components of cas proteins or cas protein complexes of the various RNA-guided CRISPR-Cas system are listed in Table 1 below together with their DNA coding sequences and respective amino acid sequences:

25

30

Table 1

Type - Sub-type	Strain designation & Genome ID	Locus id	Cas protein or protein complex sequences		Cognate palindromic repeat SEQ ID No.:	tracrRNA SEQ ID No.:
			Protein id	SEQ ID No.:		

I A	<i>Archaeoglobus fulgidus</i> DSM 4304 NC_000917.1	AF1870 AF1871 AF1872 AF1873 AF1874 AF1875 AF1876 AF1877 AF1878 AF1879	WP_010879363.1 WP_010879364.1 WP_010879365.1 WP_010879366.1 WP_010879367.1 WP_010879368.1 WP_010879369.1 WP_010879370.1 WP_010879371.1 WP_010879372.1	SEQ ID No: 1 SEQ ID No: 2 SEQ ID No: 3 SEQ ID No: 4 SEQ ID No: 5 SEQ ID No: 6 SEQ ID No: 7 SEQ ID No: 8 SEQ ID No: 9 SEQ ID No: 10	SEQ ID: 109	N/A
I B	<i>Clostridium difficile</i> 630 NC_009089.1	CD630_29770* CD630_29820 CD630_29810 CD630_29800 CD630_29780 CD630_29790	YP_001089495.1 YP_001089500.1 YP_001089499.1 YP_001089498.1 YP_001089496.1 YP_001089497.1	SEQ ID No: 11 SEQ ID No: 12 SEQ ID No: 13 SEQ ID No: 14 SEQ ID No: 15 SEQ ID No: 16	SEQ ID: 110	N/A
I C	C-125 DNA <i>Bacillus halodurans</i> NC_002570.2	BH0336 BH0337 BH0338 BH0339 BH0340 BH0341 BH0342	WP_010896517 WP_010896518.1 WP_010896519.1 WP_010896520.1 BAB04059.1 WP_010896522.1 WP_010896523.1	SEQ ID No: 17 SEQ ID No: 18 SEQ ID No: 19 SEQ ID No: 20 SEQ ID No: 21 SEQ ID No: 22 SEQ ID No: 23	SEQ ID: 111	N/A
I D	<i>Cyanothece</i> sp. 8802 NC_013161.1	Cyan8802_0527 Cyan8802_0526 Cyan8802_0525 Cyan8802_0524 Cyan8802_0523 Cyan8802_0522 Cyan8802_0521 Cyan8802_0520	WP_015783167.1 WP_015783166.1 WP_015783165.1 WP_015783164.1 WP_015783163.1 WP_015783162.1 WP_015783161.1 WP_015783160.1	SEQ ID No: 24 SEQ ID No: 31 SEQ ID No: 30 SEQ ID No: 29 SEQ ID No: 28 SEQ ID No: 27 SEQ ID No: 26 SEQ ID No: 25	SEQ ID: 112	N/A
I E	<i>Escherichia coli</i> K-12 NC_000913.3	ygcB ygbF	NP_417241.1 NP_417234.2	SEQ ID No: 32 SEQ ID No: 33	SEQ ID: 113	N/A
I F	<i>Yersinia</i> <i>pseudotuberculosis</i> YPIII NC_010465.1	YPK_1644 YPK_1645 YPK_1646 YPK_1647 YPK_1648	WP_072080408.1 WP_012303962.1 WP_012303963.1 WP_012303964.1 WP_011192624.1	SEQ ID No: 34 SEQ ID No: 35 SEQ ID No: 36 SEQ ID No: 37 SEQ ID No: 38	SEQ ID: 114	N/A

		YPK_1649	WP_002211866.1 SEQ ID No: 39		
I F* (variant)	<i>Shewanella putrefaciens</i> CN-32 NC_009438.1	Sputcn32_1819 Sputcn32_1820 Sputcn32_1821 Sputcn32_1822 Sputcn32_1823	WP_011789486.1 SEQ ID No: 40 WP_000989810.1 SEQ ID No: 41 WP_011919225.1 SEQ ID No: 42 WP_011919226.1 SEQ ID No: 43 WP_011919227.1 SEQ ID No: 44	SEQ ID: 115	N/A
II A	<i>Streptococcus pyogenes</i> MGAS10270 NC_008022.1	RS04390	WP_020905136.1 SEQ ID No: 45	SEQ ID: 116	SEQ ID: 134
II A	<i>Streptococcus thermophilus</i> LMD-9 NC_008532.1/CP000419.1	STER_1477	ABJ66636.1 SEQ ID No: 168	SEQ ID: 169	SEQ ID No: 170
II B	<i>Legionella pneumophila</i> str. Paris NC_006368.1	Lpp0160	WP_011212792.1 SEQ ID No: 46	SEQ ID: 117	SEQ ID: 135
II C	<i>Neisseria meningitidis</i> 8013 NC_017501.1	NMV_RS09835	WP_014574210.1 SEQ ID No: 47	SEQ ID: 118	SEQ ID: 136
III A	<i>Staphylococcus epidermidis</i> RP62A NC_002976.3	SERP2455 SERP2456 SERP2457 SERP2458 SERP2459 SERP2460 SERP2461 SERP2462 SERP2463	WP_002486027.1 SEQ ID No: 48 WP_002486034.1 SEQ ID No: 49 WP_002486031.1 SEQ ID No: 50 WP_002486041.1 SEQ ID No: 51 WP_002486018.1 SEQ ID No: 52 WP_002486044.1 SEQ ID No: 53 WP_002486045.1 SEQ ID No: 54 WP_002469399.1 SEQ ID No: 55 WP_002486040.1 SEQ ID No: 56	SEQ ID: 119	N/A
III B	<i>Pyrococcus furiosus</i> DSM 3638 NC_003413.1	PF1124 PF1125 PF1126 PF1127 PF1128 PF1129 PF1130 PF1131	WP_011012264.1 SEQ ID No: 57 WP_011012265.1 SEQ ID No: 58 WP_011012266.1 SEQ ID No: 59 WP_011012267.1 SEQ ID No: 60 WP_011012268.1 SEQ ID No: 61 WP_011012269.1 SEQ ID No: 62 WP_011012270.1 SEQ ID No: 63 WP_011012271.1 SEQ ID No: 64	SEQ ID: 120	N/A

III B* (variant)	<i>Marinomonas mediterranea</i> MMB_1 NC_015276.1	Marme_0668 Marme_0669 Marme_0670 Marme_0671 Marme_0673 Marme_0674 Marme_0675 Marme_0676 Marme_0677	WP_013659857.1 SEQ ID No: 65 WP_013659858.1 SEQ ID No: 66 WP_013659859.1 SEQ ID No: 67 WP_013659861.1 SEQ ID No: 68 WP_083799838.1 SEQ ID No: 69 WP_013659863.1 SEQ ID No: 70 WP_013659864.1 SEQ ID No: 71 WP_013659865.1 SEQ ID No: 72 WP_013659866.1 SEQ ID No: 73	SEQ ID: 121	N/A
III C	<i>Methanothermobacter thermautotrophicus</i> str. Delta H NC_000916.1	MTH323 MTH324 MTH325 MTH326 MTH327 MTH328	WP_010875962.1 SEQ ID No: 74 WP_010875963.1 SEQ ID No: 75 WP_010875964.1 SEQ ID No: 76 WP_010875965.1 SEQ ID No: 77 WP_010875966.1 SEQ ID No: 78 WP_010875967.1 SEQ ID No: 79	SEQ ID: 122	N/A
III D	<i>Synechocystis</i> sp. 6803 NC_005230.1	sII7063 sII7064 sII7065 sII7066 sII7067 slr7068	WP_011153732.1 SEQ ID No: 80 WP_011153733.1 SEQ ID No: 81 WP_011153734.1 SEQ ID No: 82 WP_011153735.1 SEQ ID No: 83 WP_011153736.1 SEQ ID No: 84 WP_011153737.1 SEQ ID No: 85	SEQ ID: 123	N/A
IV	<i>Thioalkalivibrio</i> sp. K90mix NC_013930.1	TK90_2699 TK90_2700 TK90_2701 TK90_2702 TK90_2703	WP_013006550.1 SEQ ID No: 86 WP_013006551.1 SEQ ID No: 87 WP_018940624.1 SEQ ID No: 88 WP_013006553.1 SEQ ID No: 89 WP_013006554.1 SEQ ID No: 90	SEQ ID: 124	N/A
V A	<i>Francisella</i> cf. <i>novicida</i> Fx1 NC_017450.1	FNFX1_1431	WP_014550095.1 SEQ ID No: 91	SEQ ID: 125	N/A
V B	<i>Alicyclobacillus acidoterrestris</i> ATCC 49025 NZ_AURB01000127.1	N007_06525 N007_06530 N007_06535	WP_021296342.1 SEQ ID No: 92 WP_021296343.1 SEQ ID No: 93 WP_021296344.1 SEQ ID No: 94	SEQ ID: 126	SEQ ID: 137
V C	<i>Oleiphilus</i> sp. HI0009 HI0009_c113 LWEK01000147.1	A3715_16885 A3715_16890	KZX85786.1 SEQ ID No: 95 KZX85787.1 SEQ ID No: 96	SEQ ID: 127	N/A
V D	<i>Bacterium</i> CG09_39_24	BK003_02070	OJI08768.1 SEQ ID No: 97	SEQ ID: 128	N/A

	<i>isolate CG09_39_24</i> MOEH01000029.1	BK003_02075	OJI08769.1 SEQ ID No: 98		
V E	<i>Deltaproteobacteria</i> <i>bacterium</i> <i>GWA2_43_19</i> <i>gwa2_scaffold_7984</i> MGPG01000094.1	A2Z89_08250 A2Z89_08255 A2Z89_08260 A2Z89_08265	OGP07438.1 SEQ ID No: 99 OGP07439.1 SEQ ID No: 100 OGP07440.1 SEQ ID No: 101 OGP07441.1 SEQ ID No: 102	SEQ ID: 129	SEQ ID: 138
VI A	<i>Leptotrichia shahii DSM</i> <i>19757</i> <i>B031DRAFT_scaffold_9</i> <i>.10</i> NZ_KB890278.1	B031_RS011044 5	WP_018451595.1 SEQ ID No: 103	SEQ ID: 130	N/A
VI B1	<i>Prevotella buccae</i> ATCC 33574 SCAFFOLD1 NZ_GL586311.1	HMPREF6485_R S00335 HMPREF6485_R S00340	WP_004343973.1 SEQ ID No: 104 WP_004343974.1 SEQ ID No: 105	SEQ ID: 131	N/A
VI B2	<i>Bergeyella zoohelcum</i> ATCC 43767 JH932293.1	HMPREF9699_0 2005 HMPREF9699_0 2006	EKB54193.1 SEQ ID No: 106 EKB54194.1 SEQ ID No: 107	SEQ ID: 132	N/A
VI C	<i>Fusobacterium</i> <i>perfoetens</i> ATCC 29250 NZ_JHXW01000011.1	T364_RS010511 0	WP_027128616.1 SEQ ID No: 108	SEQ ID: 133	N/A

N/A = not applicable or required; * non-essential Cas component of Type IB.

5 The nucleic acid molecule in the recombinant biological cell may comprise a sequence encoding a mature crRNA (mat-crRNA) comprising at least one spacer sequence flanked at least at one end (preferably at both ends) by a palindromic repeat sequence. The spacer sequence is complementary to a nucleic acid sequence of at least 15 nucleotides of at least one strand of said reporter gene or its transcript. This complementary sequence guides the polypeptide or polypeptide complex having RNA-guided endonuclease activity to the site of cleavage in the reporter gene or its transcript. Accordingly, the complementary sequence of the spacer is selected such that cleavage of the reporter gene or its transcript prevents expression of the reporter gene or its transcript.

10

The nucleic acid molecule in the recombinant biological cell may further comprise a sequence encoding a trans-activating crRNA (tracrRNA). The polypeptide or polypeptide complex having RNA-guided endonuclease activity selectively associates with its cognate palindromic repeat sequence and, where applicable also its cognate tracrRNA. Thus, when the polypeptide or polypeptide complex is selected from Table 1, then the cognate palindromic repeat sequence and cognate tracrRNA are encoded by the nucleotide sequences listed in Table 1.

In one embodiment, the nucleic acid molecule in the recombinant biological cell comprises a sequence encoding a single RNA transcript comprising a fusion of the mat-crRNA and the tracrRNA.

In a preferred embodiment, the amino acid sequence of the polypeptide or polypeptides of a protein complex having RNA-guided endonuclease activity has at least 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% amino acid sequence identity to an amino acid sequence selected from Table 1. In this embodiment the at least one cognate palindromic repeat sequence is encoded by a nucleic acid sequence at the corresponding position in the list of Table 1; while the cognate trans-activating crRNA is encoded by a nucleotide sequence at the corresponding position in the list of Table 1.

The reporter gene in the recombinant biological cell encodes a reporter polypeptide, whose expression in the cell is detectable; for example either by virtue of the inherent properties of the polypeptide (e.g. a polypeptide with fluorescent properties, such as GFP), or by means of the catalytic activity of the polypeptide (e.g. a polypeptide that confers resistance to a growth inhibitor (e.g. a protein conferring resistance to an antibiotic such as tetracycline, kanamycin, neomycin, tobramycin, chloramphenicol, spectinomycin, gentamycin, ampicillin and carbenicillin, as exemplified in Example 1.1) or whose activity complements an auxotrophic phenotype (e.g. vitamin auxotrophy). The reporter gene or its transcript is the target for the polypeptide or polypeptide complex having RNA-guided endonuclease activity. In one embodiment the reporter gene comprises a protospacer adjacent motif (PAM) located adjacent to the sequence complementary to the mat-crRNA spacer sequence, in particular where the RNA-guided endonuclease polypeptide is a Type II A Cas (e.g. Cas9). A canonical PAM is the sequence 5'-NGG-3' where "N" is any nucleobase.

Expression of the RNA-guided endonuclease protein or protein complex and respective cognate pre-crRNA, mat-crRNA and/or tracrRNA molecules (e.g. sgRNA) in the recombinant biological cell leads to a rapid inactivation of the reporter gene or its

transcript. For this reason, their expression should be tightly regulated. Preferably the first inducible promoter, or a combination of the first promoter and the first cis element conferring inducible expression is both structurally and functionally different from the second inducible promoter, or a combination of the second promoter and the second cis element; such that expression of the protein or protein complex having RNA-guided endonuclease activity and its cognate pre-crRNA, mat-crRNA and/or tracrRNA molecules (e.g. sgRNA) can be regulated independently. A suitable inducible promoter includes an IPTG-inducible T5 or T7 promoter. In a preferred embodiment, inducible expression may be conferred by a promoter regulated by a cis element encoding a biosensor transcription factor, for example the araBAD promoter pBAD, whose expression is regulated by metabolites L-arabinose, and mevalonate (as exemplified in Example 1.1). In another embodiment, expression of a gene is regulated translationally by means of a constitutive promoter in combination with a cis element encoding a riboswitch, for example a theophyllin riboswitch (as described in Example 1.1). The selectivity of the screening method is improved by employing a strong inducible promoter; for example a constitutive promoter selected from the Anderson promoter collection having a measured strength (activity) of >0.65, or 0.7 or above (<http://parts.ioem.org/Promoters/Catalog/Anderson>) such as J23100 [SEQ ID No.:172], J23101 [SEQ ID No.:173], J23102 [SEQ ID No.:174], and J23104 [SEQ ID No.:175], having a measured strength of 1.00; 0.70; 0.86; and 0.72 respectively (see Examples 2.2 and 6). Expression of the reporter gene in a recombinant biological cell is preferably driven by a constitutive promoter, (e.g. Ptet promoter, without the repressor promoter or a sigma70 promoter from the Anderson promoter collection).

Each of said first gene, first gene cluster, reporter gene and nucleic acid molecule may independently be a component of an episome that is stably replicated in the recombinant biological cell or an integrated component of its chromosome. Preferably, the episome is a plasmid comprising an origin of replication (ori) conferring a low copy number, such as pSC101 and pI5A described in Example 1.1 (Figure 2). When a population of recombinant cells of the invention are used to screen for regulators of a CRISPR-Cas associated activity; the use of a low copy number plasmid is advantageous, since it reduces variability between the recombinant cells; which in turn enhances the stringency of the screening assay.

The functional operation of the method of the invention involves the inactivation of a reporter gene such that a reporter polypeptide is no longer expressed. Inactivation of the reporter gene by DNA cleavage also leads to a progressive elimination of the episome or chromosome comprising the reporter gene. Accordingly, in a one embodiment, the reporter gene is located on a plasmid, separate from the location of

the other first genes and nucleic acid molecule required for expression of the CRISPR associated activity. In this manner, the recombinant cells continue to express inducible CRISPR associated activity independently of the reporter gene, ensuring complete inactivation of the reporter gene, thereby limiting the frequency of false positives.

5

The recombinant biological cell is preferably bacterial, belonging to a genus selected from the group consisting of: *Escherichia* spp., (e.g. *E. coli*); *Bacillus* spp., (e.g. *B. subtilis*); *Pseudomonas* spp., (e.g. *P. fluorescens*, *P. putida*); *Corynebacterium* spp., (e.g. *C. glutamicum*); *Agrobacterium* spp., (e.g. *A. tumefaciens*); *Caulobacter* spp., (e.g. *C. vibrioides*); *Burkholderia* spp., (e.g. *B. graminis*); *Rhizobium* spp., (e.g. *R. leguminosarum*), and *Raistonia* spp., (e.g. *R. metaioidurans*). Preferably the host biological cell is species of the genus *Escherichia* spp., in particular *E. coli*.

10

II The method for screening a small molecule library for a regulator of CRISPR-Cas associated activity

15

When the method according to the first embodiment is employed to screen a small molecule library for a regulator of CRISPR-Cas associated activity, the method preferably comprises the following steps:

20

a. providing one or more aliquots of recombinant biological cells having inducible CRISPR-Cas associated activity wherein said cells are prepared on a defined medium devoid of one or more agents for inducing expression of said activity;

b. providing a multi-compartment container suitable for cell culture;

25

c. delivering said aliquots of the recombinant cells into each of one or more compartments of said multi-compartment container; whereby each aliquot of cells is brought into contact with a single small molecule of a small molecule library;

d. providing each compartment of step (c) with said defined culture medium and culturing for a period sufficient to induce uptake of a candidate regulator of CRISPR-Cas associated activity;

30

e. culturing the cells obtained from step (d) on said defined culture medium wherein said medium is supplemented with one or more agents to induce expression of said CRISPR-Cas associated activity and

f. detecting cells expressing the reporter gene in the one or more compartments of said multi-compartment container obtained in step (d).

35

wherein said recombinant biological cells are according to the first embodiment as defined in section I.

The small molecule library is typically composed of small molecules, including

peptidomimetics, natural or synthetic compounds, for example libraries of diverse macrocycle compounds.

5 The growth medium in step (d) is either solid or liquid; and its composition is chosen to be compatible with the properties of the reporter protein expressed the the reporter gene in the recombinant cells. Thus, when the reporter polypeptide is detectable by virtue of the inherent properties of the polypeptide (e.g. a polypeptide with fluorescent properties, such as GFP), it is sufficient to employ a growth medium that supports growth of the recombinant cells. However, when the reporter polypeptide is detectable
10 by virtue of its catalytic activity (e.g. a polypeptide that confers resistance to a growth inhibitor (antibiotic resistance protein, as exemplified in Example 1.1) or whose activity complements an auxotrophic phenotype, then the composition of the growth medium should facilitate the mode of detection. Thus, where the polypeptide confers resistance to a growth inhibitor, the growth medium should be supplemented with growth
15 restrictive concentrations of the same growth inhibitor. Equally, where the polypeptide complements an auxotrophic phenotype, the nutrient required for growth of the auxotroph should be excluded from the growth medium.

20 Potential regulators of CRISPR-Cas associated activity are identified by detecting recombinant cells expressing the reporter polypeptide in the one or more compartments of said multi-compartment container in step (e). The means for detection is compatible with the properties of the reporter polypeptide. For example, fluorescence detection tools are available for detection of expressed fluorescent polypeptides; while detection of cell growth suffices to detect recombinant cells
25 cultured on restrictive growth medium.

III The methods for screening a library of non-host DNA fragments for a regulator of CRISPR-Cas associated activity

30 When the method according to the first embodiment is employed to screen a library of non-host DNA fragments for genes encoding a regulator of CRISPR-Cas associated activity, the method preferably comprises the following steps:

- a. providing one or more aliquots of recombinant biological cells having inducible CRISPR-Cas associated activity, wherein said cells are prepared on a defined medium devoid of one or more agents for inducing expression of said activity;
- 35 b. introducing a self-replicating library of non-host DNA fragments into one or more aliquots of said recombinant biological cells;
- c. culturing the one or more aliquots of cells obtained from step (b) on said defined culture medium for a period sufficient to induce expression of a candidate regulator of CRISPR-Cas associated activity;

- d. culturing the cells obtained from step (c) on said defined culture medium wherein said medium is supplemented with one or more agents to induce expression of said CRISPR-Cas associated activity, and
- e. detecting expression of a reporter gene present in cells obtained from step iii.
- 5 f. Optionally, individual self-replicating library members in cells detected in step (e) are isolated and re-transformed into recombinant biological cells having inducible CRISPR-Cas associated activity, and tested by repeating steps (c) to (e).

10 The construction of the library of non-host DNA fragments and its introduction into the recombinant cells is further detailed in Section V and Example 4.

The composition of the growth medium is one that facilitates the mode of selection provided by the reporter polypeptide, as described in section II. Similarly, potential
15 regulators of CRISPR-Cas associated activity encoded by genes in the library of DNA fragments are identified by detecting recombinant cells expressing the reporter polypeptide, as described in section II.

20 Preferably, the recombinant cells produced in step (b), that are cultured in step (c) are cultured on a growth medium that is solid, at a plating density that ensures the growth of single colonies. A suitable plating density is about 0.5 to 1.5×10^6 cells/cm²; which generates an overgrown plate during growth under non-selective growth medium.

25 Single colonies detected in step (d) can be selected; and the non-host DNA fragment cloned in a self-replicating plasmid of the library (and present in cells of the selected colony) can be isolated, amplified and sequenced in order to identify the candidate polypeptides as regulators of CRISPR-Cas associated activity.

30 **IV A recombinant biological cell comprising a genetic circuit for detecting regulators of CRISPR-Cas associated activity**

The invention according to the second embodiment provides recombinant bacterial cells comprising a genetic system for detecting a regulator of CRISPR-Cas associated activity (Figures 1 and 2), wherein the genetic circuit preferably comprises:

- a) a 1st self-replicating genetic element comprising,
 - 35 i. a first gene encoding a polypeptide or a first gene cluster encoding polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said gene or gene cluster is operably linked a first inducible promoter having a measured strength of >0.65 on the Anderson Scale, or a combination of a first promoter having

- a measured strength of >0.65 on the Anderson Scale and a first cis element conferring inducible expression wherein said inducible expression is induced by at least one agent;
- ii. a first reporter gene encoding a selectable marker polypeptide and operably linked to a constitutive promoter and
- 5 b) a 2nd self-replicating genetic element comprising,
- i. a second reporter gene operably linked to a constitutive promoter;

wherein either said 1st self-replicating genetic element or said 2nd self-replicating genetic element further comprises :

- 10 a nucleic acid molecule encoding any one of:
- o a pre-crRNA molecule (encoded by a DNA molecule comprising a CRISPR array) wherein said pre-crRNA comprises a spacer sequence complementary to a nucleic acid sequence of at least 15 nucleotides of at least one strand of said reporter gene or its transcript;
 - 15 o a mat-crRNA molecule comprising at least one spacer sequence flanked by at least one palindromic repeat sequence, wherein the spacer sequence is complementary to a nucleic acid sequence of at least 15 nucleotides of at least one strand of said reporter gene or its transcript;
 - 20 o said pre-crRNA molecule and a trans-activating crRNA; or
 - o said mat-crRNA molecule and a trans-activating crRNA (or a fusion thereof),

25 wherein said polypeptide or polypeptide complex having RNA-guided endonuclease activity is capable of inactivating said heterologous reporter gene or its transcript, and

wherein said nucleic acid molecule is operably linked to a second inducible promoter measured strength of >0.65 , or a combination of a second promoter and second cis element conferring inducible expression, wherein said inducible expression is induced by at least one agent.

30

Preferably, said reporter gene is comprised on a plasmid devoid of genes capable of expressing CRISPR-Cas associated activity.

35 The first gene or first gene cluster encoding a polypeptide or polypeptide complex, respectively having RNA-guided endonuclease activity, include those belonging to the Type I, Type II, Type III; Type IV, Type V, and Type VI and their subtypes; as exemplified in Table 1. In most cases the first gene or first gene cluster and their respective encoded proteins are heterologous with respect to the host cell in which they are expressed. In one preferred example the CRISPR associated protein is a Type

5 II subtype A protein, in particular a Cas9 protein. Alternatively, the CRISPR associated protein is preferably a class 2 type V system comprising the signature gene Cpf1 encoding a Cpf1 protein (e.g. WP_014550095.1). Alternatively, the CRISPR associated protein is preferably a Class I type 1 system comprising the signature gene Cas3
5 encoding a Cas3 protein.

10 The nucleic acid molecule in the genetic circuit encodes a pre-crRNA or a mat-crRNA molecule comprising a palindromic repeat sequence that is cognate to the polypeptide or polypeptide complex having RNA-guided endonuclease activity. When the
10 polypeptide or polypeptide complex is selected from Table 1, then the repeat and tracrRNA preferably have a cognate sequence listed in Table 1. In one embodiment, the nucleic acid molecule in the genetic circuit comprises a sequence encoding a single RNA transcript comprising a fusion of the gRNA and tracrRNA, and is heterologous with respect to the host cell in which it is expressed.

15 The first reporter gene encoding the selectable marker polypeptide in the 1st self-replicating genetic element allows the selection of recombinant cells comprising the 1st self-replicating genetic element and its maintenance therein. The selectable marker polypeptide encoded by the first reporter gene and the reporter polypeptide encoded
20 by the second reporter gene must be different, to allow plasmid selection and reporter detection to be independent of each other. Suitable marker polypeptides include those conferring resistance to a growth inhibitor (for a protein conferring resistance to an antibiotic such as tetracycline, kanamycin, neomycin, tobramycin, chloramphenicol, spectinomycin, gentamycin, ampicillin and carbenicillin, as exemplified in Example
25 1.1). Expression of the heterologous gene encoding the selectable marker polypeptide is preferably driven by a constitutive promoter, (e.g. Ptet promoter, without the repressor promoter or a sigma70 constitutive promoter (<http://parts.igem.org/Promoters/Catalog/Anderson>)).

30 The reporter gene in the genetic circuit encodes a reporter polypeptide having the same properties as the reporter gene in recombinant cells described in Section I.

35 Induced expression of CRISPR-Cas associated activity in recombinant cell comprising the genetic circuit leads to a rapid cleavage of the reporter gene or its transcript. For this reason, expression of the RNA-guided endonuclease protein or protein complex and its cognate gRNA in a host cell is tightly regulated. Inducible expression of the component genes in the genetic circuit is regulated in the same manner as the corresponding genes in the recombinant cells described in Section I.

The 1st and 2nd self-replicating genetic element may be any episome that is stably replicated in a host cell, for example a plasmid. Suitable plasmids for use in the genetic circuit are the same as the corresponding plasmids for use in the recombinant cells described in Section I.

5

Self-replicating genetic elements (e.g. plasmids); methods for assembling the genetic circuit therein; and their subsequent cloning in a suitable host cell, are illustrated in Example 1, using commercially available methods and tools that are known to those skilled in the art (see, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, 1989).

10

Suitable recombinant biological cells comprising the genetic circuit according to the second embodiment of the invention may be obtained using the same host cells as those used to obtain the recombinant biological cells described in Section I.

15

The genetic circuit of the invention may be transformed into the chosen biological host cell employing standard transformation protocols, including electroporation as illustrated in Example 1.3.

20 **V A recombinant biological cell library comprising a genetic circuit for detecting regulators of CRISPR-Cas associated activity**

The invention, according to a third embodiment, provides a recombinant biological cell library of non-host DNA fragments wherein each cell of the library comprises the genetic circuit described in section IV.

25

Preferably the library is a metagenomics DNA library where the DNA fragments of metagenomic DNA are cloned into a self-replicating vector (e.g. plasmid or cosmid comprising an origin of replication). The average size of the metagenomic DNA insert in a plasmid is about 1.5 kb; but using a cosmid can be up to about 40 kb. While expression of open reading frames in the cloned DNA fragments may be driven by endogenous promoters within the DNA fragments; a suitable constitutive promoter may be provided upstream of the DNA insert cloning site in the self-replicating vector. Preferably the self-replicating vector into which the metagenomic DNA is cloned is provided with a gene encoding a selectable marker polypeptide, which is different from the marker polypeptides encoded by the reporter gene and the heterologous gene encoding a marker polypeptide in the genetic circuit.

30

35

The recombinant biological cell library is conveniently constructed in the recombinant bacterial cells comprising the genetic circuit according to the invention. The DNA fragments of metagenomic DNA can be derived from human, cow and pig fecal samples and soil samples, and the library can be constructed as described in Example 4.1 and 4.2.

VI Enhancing the efficiency of the method for *in vivo* detection of anti-CRISPR polypeptides or small molecules using a recombinant biological cell of the invention.

The efficiency of the method for functional screening for anti-CRISPR molecules is improved by employing biological cells comprising the genetic circuit according to the invention, wherein expression of the RNA guided CRISPR-Cas activity is controlled by a strong promoter and inducible expression, and a selectable plasmid having the reporter gene (see section I, IV and Example 2). The frequency of false positives is further reduced by using defined media to precisely regulate the expression of the genetic circuit. This is achieved by using the following method, which is suitable for screening either a metagenomics library encoding candidate anti-CRISPR proteins; or one or more isolated candidate anti-CRISPR proteins; or one or more isolated anti-CRISPR small molecules. The method comprises 6 steps:

1) Provision of recombinant biological cells capable of inducible CRISPR-Cas associated activity and a reporter gene according to the invention; and the cultivation of these cells under aerobic conditions in a defined growth medium (such as M9 minimal medium for prokaryotic cells), comprising only those nutrients essential for growth, to produce a multiplied cell culture. Preferably, the medium is selective for recombinant biological cells comprising the gene or gene cluster encoding an RNA-guided endonuclease that confers said inducible CRISPR-Cas associated activity and the nucleic acid molecule encoding its cognate RNA molecule that is capable of guiding said RNA-guided endonuclease. Thus, when said gene or gene cluster and said nucleic acid molecule is provided on one or more plasmid, then the medium comprises both nutrients essential for growth and antibiotics selective for cells maintaining the respective plasmids. Importantly, the defined medium is devoid of agents required for, or capable of, inducing expression of the genes or gene clusters encoding an RNA-guided endonuclease and its cognate RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said reporter gene or its transcript.

2) An aliquot of the cell culture produced in step 1) is inoculated into said defined growth medium comprising nutrients essential for growth to produce a cell suspension; said medium being both devoid of agents required for, or capable of, inducing expression of the genes or gene clusters encoding an RNA-guided endonuclease and

its cognate RNA molecule; and devoid of antibiotics selective for cells maintaining the respective plasmids;

3) cells in the cell suspension produced in step 2) are either:

5 i) transformed with a metagenomic library to produce a transformed cell suspension,
or

ii) contacted with a candidate anti-CRISPR small molecule or a candidate anti-CRISPR protein by their addition to said cell suspension;

10 4) a cell suspension resulting from step 3) i) is incubated (preferably between 0.25 -2 h) to allow expression of candidate anti-CRISPR protein encoded by DNA molecules in the metagenomic library; or 3) ii) is incubated (preferably between 0.25 -2 h) to allow intracellular uptake of the candidate anti-CRISPR small molecule or candidate anti-CRISPR protein;

15 5) Agents are added to the incubated cell suspension(s) resulting from step 4; and the suspensions are incubated under conditions suitable for said agents to induce expression of the genes or gene clusters encoding an RNA-guided endonuclease and its cognate RNA molecule present in the biological cell provided in step 1); whereby said expressed endonuclease activity is capable of inactivating the reporter gene or its transcript present in the biological cell during said incubation (preferably between 1-3 h);

20 6) Cells derived from the incubated cell suspension(s) resulting from step 5) are plated on agar plates composed of a defined media comprising nutrients essential for growth and supplemented with agents required for, or capable of, inducing expression of the genes or gene clusters encoding an RNA-guided endonuclease and nucleic acid molecule(s) encoding its cognate RNA molecule(s). When said gene or gene cluster and said nucleic acid molecule(s) are provided on one or more plasmids, then the medium further comprises antibiotics selective for cells maintaining the respective plasmids. In the case that the reporter gene or its transcript encodes an antibiotic resistance protein, then the defined media agar is further supplemented with the respective antibiotic. The plates are incubated under suitable conditions for a time
25
30 sufficient to detect viable biological cells.

EXAMPLES

Example 1 Construction and cloning of a genetic circuit for screening for anti-CRISPR molecules

35 1.1 Construction of pCasens3 and pDual3 plasmids

The pCasens3 and pDual3 plasmids, illustrated in Figure 2, were constructed as follows:

Plasmid pCasens3 containing the *Streptococcus pyogenes* Cas9 gene (spCas9) [SEQ ID No.: 150] was constructed in a single step by using USER cloning (Genee et al. 2014).

A DNA fragment containing spCas9 gene [SEQ ID No.:150] and its endogenous terminator [SEQ ID No.:152] was PCR amplified from a plasmid DS-SPcas (supplied by addgene : Plasmid #48645) (Esvelt et al. 2013); and the amplified DNA fragment was cloned into the backbone of the plasmid pSEVA47 (as described by Martinez-Garcia et al. 2014) . The plasmid pSEVA47 contains: the low copy number origin of replication pSCIOI [SEQ ID No.:155], and the antibiotic resistance gene aadA [SEQ ID No.:153] that confers resistance against Spectinomycin and a restriction site for insertion of a gene. A DNA molecule [SEQ ID No.:149] encoding a theophylline translational riboswitch was placed in front of Cas9 using a long forward primer:

5'AAGTCTAGCGAACCGCACTTAATACGACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCCTGCTAAGGTAACAACAAGATGATGGATAGAAATACTCAATAGGCTTAGATATCGGCAC-3' [SEQ ID No.:156] . Additionally, a sigma70 constitutive promoter (J23100 [SEQ ID No.:172]; as defined by <http://parts.igem.org/Promoters/Catalog/Anderson>) was also introduced using a reverse primer in substitution for the endogenous of spCas9 promoter: 5'-ctctagTagctagcactgtacctagtagctgagctagccgctcaagTTAGCTGTGCTCTAGAAGCTAGCAG-3' [SEQ ID No.:157] .

Plasmid pDual3 for arabinose inducible sgRNA expression was constructed by using USER cloning (Genee et al. 2014) to insert the respective genes into the backbone of the plasmid was from pSEVA3610, described by Martinez-Garcia et al., 2015 . The chimeric gRNA encoding sequence [SEQ ID No.:161] with a TrnB terminator [SEQ ID No.:162] and a chloramphenicol resistance gene (CmR) [SEQ ID No.:164] were cloned downstream of the inducible pBAD promoter [SEQ ID No.:158] and araC gene [SEQ ID No.:159] encoding an L-arabinose-inducible transcription factor. In the absence of L-arabinose, the AraC protein binds to operator sites within pBAD effectively repressing transcription . On addition of L-arabinose, AraC protein binds to L-arabinose to form a complex having a DNA-binding conformation that activates pBAD leading to induction of transcription of its cognate genes, namely sgRNA and CmR. The plasmid comprises a low copy number origin of replication, p15A [SEQ ID No.:163] . The first twenty nucleotide sequence of the encoded gRNA transcript is complementary to nucleotides 96:115 of the + strand in the CmR gene having [SEQ ID No.:164] .

Positive strand nucleotides 96:115: CTATAACCAGACCGTTCAGC[SEQ ID No.:166]

35

1.2 Construction of positive and negative control plasmids for testing the genetic circuit

Control plasmids for testing the genetic circuit (Figure 3) include a negative control plasmid comprising a gfp gene [SEQ ID No.:140] with rrnB T1 terminator [SEQ ID

No.:142] capable of expression of GFP (A); and a positive control plasmid comprising an *acrIIA2* gene [SEQ ID No.:159] capable of expression of the anti-CRISPR protein, *AcrIIA2*, each under the control of a constitutive *Ptet* promoter (lacking the TET repressor) and RBS [SEQ ID No.:139] (B). Both plasmids have a *ColEI* ori [SEQ ID No.:143] and kanamycin resistance gene [SEQ ID No.:144]. When the control plasmids are transformed into host cells, comprising the genetic circuit (i.e. *pCasens3* and *pDual3* plasmids as described in Example 1.1); the expression of the anti-CRISPR protein, *AcrIIA2* is capable of inhibiting Cas9 mediated inactivation of the *CmR* gene.

10 1.3 Host cells transformed with *pCasens3* and *pDual3* plasmids

E.coli TOPIO were made electro-competent and transformed with the genetic circuit comprising the plasmids: *pCasens3* and *pDual3* (obtained in example 1.1) as follows:

- 5 i. *E.coli* TOPIO cells (supplied by ThermoFischer) were cultured in 3 ml of 2xYT media for 16-18 hours at 37°C and 250 rpm;
- 15 ii. the culture from (i) was inoculated 1:100 3 ml of pre-warmed 2xYT media; and cultured at 37°C at 250 rpm until the cell density reached 0.5-0.7 OD600;
- iii. the cultures from (ii) were cooled on ice; and the cells harvested by centrifugation at 10 000 xg for 1 min;
- 20 iv. the cells were washed by re-suspension in cold water and then harvested again by centrifugation; then the cells were re-suspended in 1 ml glycerol and harvested by centrifugation; and finally the cells were re-suspended in the residual water and then placed on ice, ready for electroporation.
- 25 v. A 40 μ l sample of the electrocompetent cells were transferred to microcentrifuge tubes; to which a 1 μ l sample of *pCasens3* and *pDual3* DNA was added; and this DNA-cell mixture was then transferred to a cold cuvette and electroporated with a pulse having 1.8 kv, 200 ohms and 25 μ F.
- 30 vi. Immediately thereafter, a 975 μ l volume of 37°C SOC [Ausubel F.M. et al. (1987)] was added to the electroporated cells, which were mixed and transferred to a 15 ml- tube; and incubated on rollers at 37°C for 1 h;
- 35 vii. The recovered transformed cells were diluted 10^0 - 10^8 ; plated on selective solid media comprising Luria Broth agar supplemented with: 50 μ g/mL spectinomycin; 30 μ g/mL chloramphenicol and 50 μ g/mL kanamycin; and incubate overnight at 37°C.
- viii. Transformed cell colonies comprising the *pCasens3* and *pDual3* plasmids (*E. coli* genetic circuit [*E. coli*-GC]) were selected; amplified and stored.

1.4 *E. coli*-GC cells transformed with control plasmids

E.coli TOPIO containing the genetic circuit comprising the plasmids: *pCasens3* and *pDual3* (*E. coli*-GC) were made electro-competent and transformed with either the

negative control plasmid expressing GFP or positive control comprised a plasmid expressing AcrIIA2, as described in Example 1.2, employing the transformation protocol described in Example 1.3, but where the host cells were E. coli-GC.

5 **Example 2 Optimisation of the genetic circuit**

The use of host cells comprising the genetic circuit to screen for regulators of RNA-guided endonucleases relies on the stringency of the genetic circuit; such that the activity of a candidate regulator can be clearly detected. As a first step in optimizing the circuit, the importance of regulation of the expression of the RNA-guided
10 endonuclease was investigated.

The pCasens3 plasmid was modified to facilitate the comparison of 3 plasmids, differing only in the promoter region operably linked to the Cas9 gene. These modifications allowed an analysis of the impact of both the strength and on-off
15 regulation of the promoter to be assessed.

2.1 Methodology

Three samples of E.coli TOPIO cells were transformed (as described in Example 1.3) with the genetic circuit comprising the pDual3 plasmid together with one of 3 versions
20 of the pCasens3 plasmid; and the transformed cells were then duplicate plated on LB agar comprising the antibiotics (50 µg/ mL spectinomycin; 30 µg/ mL chloramphenicol and 50 µg/ mL kanamycin); where one set of plates was further supplemented with 1% arabinose to induce sgRNA expression (alone) or together with 2mM theophylline for inducible Cas9 expression.

25

2.2 Optimised genetic circuit requires regulated and strong expression of the Cas9 gene

E.coli TOPIO cells, transformed with genetic circuits comprising the pDual3 plasmid and each of the 3 versions of the pCasens3 plasmid, are dependent on expression of a
30 functional camR gene in order to grow on the LB plates comprising chloramphenicol. Effective inactivation of the camR gene, by the Cas9-sgRNA can be determined by detecting the loss of cell growth under these conditions.

E.coli TOPIO cells comprising the first version of the pCasens3 plasmid comprising a
35 Cas9 gene regulated by a medium strength constitutive promoter (ProC promoter [SEQ ID No.:167]; **Esvelt**, K. M., 2013) (see Figure 4) showed slower growth than cells comprising the second and third versions of the pCasens3 plasmid when grown on growth medium lacking both theophylline and arabinose required for inducing expression of the Cas9 and sgRNA genes respectively. The observed growth

depression may be attributable to a potential fitness cost associated with the expression of Cas9, since it is a large protein (>1300 amino acids in length). When expression of sgRNA was induced in these cells with arabinose, in combination with the constitutively expressed Cas9, the growth rate fell as expected; indicating the
5 inactivation of the camR gene. However, the difference in growth rate between non-induced and induced states of the cas9-sgRNA was small.

E.coli TOPIO cells comprising the second version of the pCasens3 plasmid comprising a Cas9 gene regulated by the medium strength promoter (ProC promoter [SEQ ID
10 No. : 167] in combination with the inducible theophyllin riboswitch (see Figure 4) showed a faster growth than the first version, which correspondingly may be attributed to the "off" state of cas9 expression, and absence of its fitness cost. However, when plated on media inducing expression of the Cas9-sgRNA, the log reduction in growth was limited, indicating that the medium strength promoter following induction, failed
15 to drive sufficient expression of the Cas9 gen in order to obtain optimal inactivation of the camR gene.

E.coli TOPIO cells comprising the third version of the pCasens3 plasmid comprising a Cas9 gene regulated by the strong promoter (BBa J23100) combined with the
20 inducible theophyllin riboswitch (see Figure 4) also showed a faster growth than the first version, which correspondingly may be attributed to the "off" state of cas9 expression, and absence of its fitness cost. Additionally, when plated on media inducing expression of the Cas9-sgRNA, the log reduction in growth was substantial, indicating that a tightly regulated and strong promoter controlling expression of the
25 Cas9 gene is important in order to obtain optimal inactivation of the camR gene; and an effective screening genetic circuit.

E.coli TOPIO cells comprising a fourth version of the pCasens3 plasmid comprising a Cas9 gene operably linked a non-regulated strong constitutive promoter (BBa J23100)
30 is additionally tested to confirm the importance of promoter regulation.

2.3 Optimised genetic circuit locates reporter gene for CRISPR-Cas on a separate plasmid

35 The effect of the genetic location of the reporter gene on the efficacy of CRISPR-Cas (when located on the same or on a separate genetic element) is tested as follows:
The pCasens3 plasmid is modified by insertion of the arabinose inducible sgRNA expression cassette, comprising the chimeric gRNA encoding sequence [SEQ ID No.:161] with a TrnB terminator [SEQ ID No.:162] cloned downstream of the

inducible pBAD promoter [SEQ ID No.:158] and araC gene [SEQ ID No.:159] encoding an L-arabinose-inducible transcription factor, to create pCasens4. A chloramphenicol resistance gene (CmR) [SEQ ID No.:164] is inserted into a second plasmid having the low copy number origin of replication, p15A [SEQ ID No.:163] to create pCmR.

5 E.coli TOPIO cells, transformed either pCasens4 and pCmR, or pCasens3 and pDual3 (see Example 1.1) are cultivated under aerobic conditions in M9 minimal medium (<http://cshprotocols.cshlp.org/content/2010/8/pdb.recl2295.short>), supplemented with the antibiotics (spectinomycin 50ug/mL and chloramphenicol 30ug/mL) and the two cultures are incubated at 37°C; 250RPM. An aliquot of each of the resulting cultures
10 is transferred to fresh M9 minimal medium (devoid of said antibiotics), and cells in the resulting two cell suspensions are transformed with the negative control plasmid expressing GFP (see Example 1.2 and Figure 3B) and then incubated for at 37°C for a period of ≈1h, in M9 minimal medium, sufficient to allow expression of plasmid encoded genes, e.g. GFP (this negative control plasmid substitutes for a plasmid
15 expressing a candidate anti-CRISPR protein, where a corresponding period for expression of plasmid genes is required). The two transformed cell suspensions are then supplemented with 2mM theophylline; 1% arabinose; and incubated at 37°C; 250RPM for a period of ≈2h, to allow induced expression of the RNA-guided CRISPR-Cas genes, and inactivation of the CmR reporter gene. Finally, serial dilutions of the
20 cells are plated in solid M9-agar plates supplemented with 50 ug/ml Spectomycin ; 50 ug/ml Kanamycin ; 30ug/ml chloramphenicol ; and the inducing agents, 2mM theophylline; 1% arabinose.

Further tests demonstrate that when the reporter gene (in this example the CmR reporter gene) is located on a plasmid, separate from the location of the other first
25 genes and nucleic acid molecule required for expression of the CRISPR associated activity this serves to enhance inactivation of the reporter gene, thereby limiting the frequency of false positives.

30 **Example 3 Use of host cells comprising the genetic circuit to screen for anti-CRISPR molecules**

Growth of host cells comprising the genetic circuit, E. coli-GC (prepared as described in Example 1.3) on a medium comprising theophylline and arabinose leads to induction of expression of the cas9 endonuclease and sgRNA, respectively. Since the cas9
35 endonuclease, when complexed with the sgRNA transcript, will then selectively base-pair and cleave the chloramphenicol resistance gene (CmR) in the pDual3 plasmid, this results in a loss of chloramphenicol resistance, and a progressive elimination of this plasmid. Thus when expression of the cas9 endonuclease and sgRNA are co-induced, growth of the cells on a medium comprising chloramphenicol is inhibited. Anti-CRISPR

molecules capable of being imported into the host cells and inhibiting cas9 mediated inactivation of the CmR gene are detected by the survival and grow of the host cells on a medium comprising chloramphenicol. In brief, *E. coli*-GC cells from an overnight culture are diluted and plated in microtiter wells, into which small molecules are introduced (pin-transferred) and cells are then incubated at 37°C. Cell growth can be quantitated 24 h later via OD600 (or 630) readings on in Envision. Wells showing cell growth, detected by an increase OD600, identify molecules having anti-CRISPR activity.

E. coli-GC cells transformed with additional plasmids were employed as controls: *E. coli*-GC cells acting as a negative control comprised a plasmid expressing GFP; and *E. coli*-GC cells acting as a positive control comprised a plasmid expressing AcrIIA2.

3.1 Procedure for screening for anti-CRISPR molecules

- i. Overnight cultures of *E. coli* comprising the genetic circuit, *E. coli*-GC (see Example 1.2) were prepared in filtered sterilized minimal media supplemented with 50 µg/mL spectinomycin and 30 µg/mL chloramphenicol to maintain the genetic circuit, and comprising glucose as the sole carbon source.
- ii. Screening was performed on multiwell assay plates (Corning 3710), in a minimal media further containing 2mM theophylline to induce Cas9 endonuclease expression; and 1% arabinose to induce gRNA transcription as follows:
 - a. columns 1-23 of the multiwell assay plates were prefilled with 30 µL of minimal media comprising *E. coli*-GC cells of the strain carrying the negative control plasmid expressing GFP (the negative control), and column 24 of the multiwell assay plates were prefilled with 30 µL of minimal media comprising *E. coli*-GC cells of the strain carrying the positive plasmid expressing the inhibitor protein, AcrIIA2.
 - b. 300 nL of each candidate anti-CRISPR molecule were pin-transferred to each well of each plate. All molecules were tested in duplicate on duplicate multiwell assay plates under the same conditions. The optical density (OD 600nm) of each well was measure at t=0 in order to correct for compounds having an inherent absorption at 600nm. The multiwell assay plates were sealed and incubated at 37°C for 18 h.
- iii. Cell growth in each well of the multi-well assay plates was determined by measuring cell density using a PerkinElmer Envision (600 nm filter) and comparing OD600 of each well and subtracting the value of t=0h from t=18h

in order to correct for compounds having an inherent absorption at 600nm. The measured OD600 values were normalized to the positive and negative controls in each plate to determine a normalized percent of inhibitory activity of each compound. Z scores per plate per replicate were calculated, based on the normalized OD600 values. A positive hit was given for a Z score above 2.5. All hits were annotated based on the hit scores as Weak = W (<3.5), medium = M (3.5-4.5) and Strong=S (>4.5). The compounds were further ranked based on the number of hits per library and those hits having the strongest anti-CRISPR activity in each library.

3.2 E. coli-GC host cells detect proteins having anti-CRISPR activity

E. coli-GC host cells transformed with the negative control expressing GFP (Figure 3A); and the positive control plasmid expressing AcrIIA2 (Figure 3B and 6B), were screened for anti-CRISPR activity as follows :

- i. Overnight cultures of said *E.coli* host cells were prepared by culture in 3ml 2xYT media supplemented with the following antibiotics (50 µg/mL spectinomycin ; 30 µg/mL chloramphenicol and 50 µg/mL kanamycin) at 37°C and 250RPM;
- ii. Said cultures were used to prepared dilutions from 10^0 - 10^7 and 5µl volumes of each dilution was spotted onto a 1st and 2nd set of agar plates, in order to determine both cell death and cell viability, having the following agar composition :
1st set of plates used to determine cell death comprised LB agar supplemented with : 2mM theophylline; 1% arabinose and antibiotics (50 µg/mL spectinomycin ; 30 µg/mL chloramphenicol and 50 µg/mL kanamycin);
2nd set of plates used to determine cell viability comprised LB agar supplemented with : 2mM theophylline; 1% arabinose; and antibiotics (50 µg/mL spectinomycin; and 50 µg/mL kanamycin).

The survival of E. coli-GC host cells comprising the negative control plasmid, and grown under conditions that induce expression of the genetic circuit, is significantly reduced (see Figure 5B). However, expression of the, AcrIIA2 is seen to enable cell growth and survival, consistent with the known anti-CRISPR properties of this protein (Figure 5A and B).

3.3 Small molecule libraries screening using the genetic circuit

Host cells comprising the genetic circuit (e.g. E. coli-GC) have been screened using a wide range of small molecules, including peptidomimetics, natural or synthetic compounds, for example libraries of diverse macrocycle compounds. Circa 20 positive hits were obtained from screening a library of diverse macrocycles (Asinex Macrocycle library) providing evidence that the screening method is suitable for detecting molecules having anti-CRISPR activity.

10 Example 4 Construction and screening of Metagenomic Libraries for anti-CRISPR proteins

4.1 Construction of expression metagenome libraries from diverse sources

Functional metagenomic libraries from human, cow and pig fecal samples and soil samples were constructed as described previously [Sommer et al. 2009; Genee et al., 15 2016]. The steps of the procedure were as follows :

- i) total DNA was isolated from each of 5 g of fecal samples and a soil sample using the PowerMax Soil DNA Isolation Kit (Mebio Laboratories Inc.); where after total DNA isolated from each sample was treated as follows :
- ii) the extracted DNA, in a total sample volume of 200µl in minutes, was 20 fragmented by sonication into fragments of an average size of 4-5 kb using a Covaris E210 (Massachusetts, USA);
- iii) the sheared DNA (200 µl) was size-selected by separation on a 1 % agarose gel with DNA dye and DNA migrating in the 3-6 kb region was extracted using a Qiagen gel extraction kit;
- 25 iv) the gel-extracted DNA fragments were end-repaired using an End-It end repair kit (Epicentre) and selected using PCR clean up column; and
- v) the end-repaired fragments were blunt-end cloned into the HindI site, downstream of a constitutive Ptet promoter (lacking the tet repressor) and RBS [SEQ ID No.: 139] in the expression 100-200 ng pZE21 plasmid in a 1:5 ratio [Lutz and Bujard, 1997] using the Fast Link ligation kit (Epicentre) over-night and the reaction was cleaned with a spin column and eluted in a 15 µl twice;
- 30 vi) 2 µl cleaned ligation mix was transformation into 50 µl E. coli top10 cells by electroporation ; and cells in the transformation mix were recovered in 1 ml SOC medium for 2 h; using multiple electroporations to increase library size;
- 35 vii) the transformation mix were plated in 1:100 and 1:1000 dilutions on kanamycin 50 µg/ml plates to create a library;
- viii) the library was amplified by inoculation into 10 ml LB+Kan 50 and grown overnight; 5 ml 50 % glycerol is added and the library is aliquoted into 1 ml portions and frozen at -80 C.

In brief, plasmid DNA, comprising the amplified metagenomic library was extracted and recovered using Machery-Negel plasmid DNA purification kit as follows :

- 5 1. An overnight culture of bacteria comprising the metagenomic library was grown in 3 ml LB medium supplemented with 50 µg/mL kanamycin ;
2. the culture was centrifuged, and the pelleted bacteria were resuspended in the bacteria in buffer, to which a denaturing solution was added;
- 10 3. the proteins and genomic DNA were pelleted by centrifugation, and the plasmid-containing supernatant was recovered and purified through the Nucleospin® plasmid column on the kit; and bound plasmid DNA was eluted from the column using water or a neutral buffer such as Tris:EDTA, to provide a plasmid DNA comprising the metagenomic DNA expression library.

15 **4.2 Transformation and screening of Metagenomic DNA expression library into Host cells, comprising genetic circuit**

E.coli TOP10 containing the genetic circuit comprising the plasmids : pCasens3 and pDual3 (E. coli-GC) were made electrocompetent and transformed with plasmids for expression of each metagenomic library (obtained in example 4.1) as follows :

- 20 i. E.coli TOP10 cells (supplied by ThermoFischer) containing the plasmids : pCasens3 and pDual3 were cultured in 3 ml of 2xYT media with 50 µg/mL spectinomycin and 30 µg/mL chloramphenicol for 16-18 hours at 37°C and 250 rpm;
- 25 ii. the culture from (i) was inoculated 1:100 in 3 ml of pre-warmed 2xYT media with 50 µg/mL spectinomycin and 30 µg/mL chloramphenicol; and cultured at 37°C at 250 rpm until the cell density reached 0.5-0.7 OD₆₀₀;
- 30 iii. the cultures from (ii) were cooled on ice; and the cells harvested by centrifugation at 10 000 xg for 1 min;
- iv. the cells from (iii) were washed by re-suspension in cold water and then harvested again by centrifugation ; then the cells were re-suspended in 1 ml glycerol and harvested by centrifugation ; and finally the cells were re-suspended in the residual water and then placed on ice, ready for electroporation .
- 35 v. A 40 µl sample of the electrocompetent cells from (iv) were transferred to microcentrifuge tubes; to which a 1 µl sample of the metagenomic library was added; and this DNA-cell mixture was then transferred to a cold cuvette and electroporated with a pulse having 1.8 kv, 200 ohms and 25 µF.

- vi. Immediately thereafter, a 975 μ l volume of 37°C SOC [Ausubel F.M. et al. (1987)] was added to the electroporated cells from (v), which were mixed and transferred to a 15 ml- tube; and incubated on rollers at 37°C for 1 h;
- 5 vii. serial dilutions from 10^0 - 10^8 from the recovered cells from (vi) were made; and 100 μ l of each dilution were plated on selective solid media comprising Luria Broth agar supplemented with: 2mM theophylline; 1% arabinose; 50 μ g/ mL spectinomycin; 30 μ g/ mL chloramphenicol and 50 μ g/ mL kanamycin; and incubated overnight at 37°C.
- 10 viii. Colonies growing on the plates were collected by adding 1 ml H₂O, after which the colonies were scraped off the plate with a sterile loop. The bacterial cells were then pelleted and washed in water by centrifugation and the plasmid DNA in collected bacterial cells was extracted using a Machery-Negel plasmid DNA purification kit.

15

4.3 Detection of metagenomic DNA encoding proteins having anti-CRISPR activity

E.coli host cells containing the genetic circuit, transformed with the library of metagenomic DNA expression plasmids (Figure 6A) prepared as described in Example 20 2.2, were screened for expression of proteins having anti-CRISPR activity (as describe in Example 4.2 and figure 7). Colonies that appeared on the selective growth medium were selected, and their respective metagenomic DNA insert were each re-cloned in the metagenomics DNA expression plasmid (Figure 6A, plasmid to the left). 11 inserts showed anti-CRISPR activity above background in the selection assay, where in several 25 cases the detected activity was similar to that of AcrIIA2 (Figure 8); confirming that the genetic circuit provides an efficient genetic tool for selecting new candidate anti-CRISPR polypeptides from diverse genetic sources.

Example 5 A genetic circuit for screening for anti-CRISPR molecules of Class 30 1 (Type 1-B) and Class 2 (Type II-A and V-A) CRISPR-Cas systems

The genetic circuit for use in screening for anti-CRISPR molecules, comprising gene(s) encoding a Class 1 or 2 CRISPR-Cas system, in combination with a nucleic acid molecule encoding a cognate guide RNA (cognate palindromic repeat sequence and as 35 required a tracrRNA) were validated, as demonstrated below. In each case, the cognate RNA was expressed from the pDual plasmid which comprises a gene encoding cognate guide RNA targeting the chloramphenicol resistance gene (CmR), under the control of an arabinose-inducible promoter, a chloramphenicol resistance gene (cat), and a p15a origin of replication (see right-hand pDual3 plasmid in Figure 2).

5.1 Construction of a Type I-B CRISPR-Cas screening system

Class 1, type I CRISPR-Cas systems are composed of a multi-subunit protein complex, each subunit encoded by its respective gene. The type I-B CRISPR-Cas operon from *Clostridium difficile*, encoding the effector protein cas3 and cascade genes cas6, cas7, cas8 and cas5, was cloned into a medium copy number plasmid having a cloDF13 origin of replication; and spectinomycin resistance (aadA gene) to create pTypeI-B-Cas. The endogenous promoter of the type I-B CRISPR-Cas operon was replaced with the IPTG inducible strong Ptac promoter.

5.2 Construction of a Type IIA CRISPR-Cas system screening system

Class 2, type II CRISPR-Cas systems comprise a single multidomain protein encoded by a single gene. The type II-A CRISPR-Cas gene from *Streptococcus thermophilus*, encoding the *S. thermophilus* Cas9 protein (SEQ ID No: 168), was cloned in a plasmid (Esvelt et al., 2013), such that expression of Cas9 was placed under control of an IPTG-inducible Ptac promoter, creating plasmid pTypeII-A-Cas.

5.3 Construction of a Type V-A CRISPR-Cas system screening system

Class 2, type V CRISPR-Cas systems also comprise a single multidomain protein encoded by a single gene. The type V-A CRISPR-Cas from *Franciseiia novicida* strain ul12 comprises a CasI2 gene (formerly designated CpfI), encoding the *Franciseiia novicida* CasI2 effector protein (WP_014550095.1; SEQ ID No: 91), was cloned in a medium copy number plasmid having a cloDF13 origin of replication, and spectinomycin resistance (aadA gene) (as described by Zetsche et al., 2015). The plasmid was modified to place expression of CasI2 gene under control of the strong constitutive promoter J23100 [SEQ ID No.:172] (as defined by <http://parts.igem.org/Promoters/Catalog/Anderson>), regulated by a DNA sequence [SEQ ID No.:149] encoding a theophylline riboswitch; creating plasmid pTypeV-A-Cas.

5.4 Testing host cells transformed with genetic circuits expressing Class 1 (Type I-B) and Class 2 (Type II-A and V-A) CRISPR-Cas systems

E. coli TOPIO were made electro-competent and transformed with plasmids comprising:

- pTypeI-B-Cas (comprising coding sequences for cas3 [SEQ ID No.:12], cas6 [SEQ ID No.:13], cas7[SEQ ID No.:14], cas8 [SEQ ID No.:15] and cas5[SEQ ID No.:16]) and pDual4 (comprising Cognate palindromic repeat SEQ ID No.:110); or
- pTypeII-A-Cas (comprising coding sequence for Cas9 2 having SEQ ID No: 168) and pDual5 (comprising Cognate palindromic repeat SEQ ID No.:169 and tracrRNA gene having SEQ ID No: 170);

• pTypeV-A-Cas (comprising coding sequence for Cas2; SEQ ID No: 91) and pDual6 (comprising Cognate palindromic repeat SEQ ID No.:125) as described in example 1.3. Transformants were cultured overnight at 37°C with shaking (250RPM) in a growth medium supplemented with antibiotics (spectinomycin 50ug/mL, kanamycin 50ug/mL, and chloramphenicol 30ug/mL); but devoid of inducers of CRISPR-Cas expression. A subculture, derived from the over-night culture, was incubated under aerobic condition at 37C for ~6hrs in a growth medium devoid of antibiotics, but, supplemented with 1% arabinose and either 0, 10, 50 or 100uM IPTG (or 2mM theophylline as inducer for pTypeV-A-Cas) to activate expression of the sgRNA and CRISPR-Cas system respectively. The resulting cultures were serially diluted from 10^{-1} to 10^{-7} and then plated on LB-agar with and without chloramphenicol.

As demonstrated in Figures 9, 10 and 11, host *E. coli* cells transformed with each of the 3 tested CRISPR-Cas systems were able to grow under both non-selective and selective conditions (chloramphenicol media) provided that the genetic circuit encoding the CRISPR-Cas was not induced. Addition of as little as 10uM IPTG to induce expression of the genetic system, was for all 3 CRISPR-Cas systems sufficient to lead to a detectable loss of cell viability when grown under selective conditions.

20 **Example 6.0 Enhancing the efficiency of the method for *in vivo* detection of anti-CRISPR small molecules or polypeptides using recombinant biological cell of the invention.**

The frequency of false positives detected using the method for functional screening for anti-CRISPR molecules is reduced by employing biological cells comprising the genetic circuit according to the invention, wherein expression of the RNA guided CRISPR-Cas activity is controlled by a inducible strong promoter, and a selectable plasmid having the reporter gene, in combination with the use of a defined growth medium, such that induction of RNA guided CRISPR-Cas activity and reporter gene selection is restricted to defined steps of the method. The method (with reference to Figure 12) was tested by:

30 1) Providing *E. coli* cells, comprising the genetic circuit on pCasens3 and pDual3; by cultivating the cells under aerobic conditions in M9 minimal medium (<http://cshprotocols.cshlp.org/content/2010/8/pdb.rec12295.short>), supplemented with the antibiotics (spectinomycin 50ug/mL and chloramphenicol 30ug/mL) and incubating the culture at 37°C; 250RPM without inducers (figure 12, step 1). Since the medium is devoid of the inducing agents, theophylline and arabinose, *E. coli* cells comprising the genetic circuit are selected without inducing expression of the RNA-guided CRISPR-Cas genes.

2) An aliquot of the cell culture produced in step 1) is inoculated into M9 minimal medium; said medium being both devoid of the inducing agents (theophylline and arabinose) and antibiotics (spectinomycin and chloramphenicol) to produce a cell culture.

5 3) Cells, derived from the cell culture produced in step 2, are transformed with the control plasmid comprising a gene [SEQ ID No.:140] encoding Green Fluorescent Protein (GFP) under the control of a constitutive Ptet promoter [SEQ ID No.:139].

10 4) Transformed cells, obtained in step 3, are incubated at 37°C for a period of ≈1 h, in M9 minimal medium, sufficient to allow expression of the GFP protein, and providing a transformed cell culture.

15 5) The transformed cell culture obtained in step 4 is supplemented with 2mM theophylline; 1% arabinose; and the suspensions are incubated at 37°C; 250RPM for a period of ≈2h, to allow induced expression of the RNA-guided CRISPR-Cas genes, and inactivation of the CmR reporter gene in cells. Since the growth medium is devoid of chloramphenicol, growth conditions are non-selective for functional expression of this reporter gene.

20 6) Cells derived from the incubated cell suspension resulting from step 5) are plated on a defined media agar supplemented with both 2mM theophylline; 1% arabinose; and the antibiotics: spectinomycin 50ug/mL, kanamycin 50 pg/mL, and chloramphenicol 30ug/mL) and incubated at 37°C for a time sufficient to detect viable biological cells.

The use of a defined growth medium for using the CRISPR-Cas screening method of the invention is shown to reduce the number of false positives for each of 5 different types of CRISPR-Cas (Figure 13).

25

The effect of the strength of the promoter used to control inducible expression of RNA guided CRISPR-Cas activity on the efficiency of inactivation of the reporter gene (camR) gene cleavage, measured as cell survivors, and corresponding reduction in false-positives using the screening assay is demonstrated in Figure 14. The three types of cells tested comprised the CRISPR-Cas9 genetic circuit, differing only in that their respective Cas9 gene was operably linked to a promoter having promoter strengths of 0.16; 0.58 or 1.00 on the Anderson scale, each combined with a theophylline riboswitch 5' of the cas9 gene. The cells were tested according to steps 1, 2, 5 and 6, and omitting transformation steps 3 and 4 (Figure 12).

35

References

- Ausubel F.M. et al. (1987)** *Current Protocols in Molecular Biology* ;
- Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Young, S. J., & Church, G. M. (2013)** . Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods*, 10(11), 1116-21. <https://doi.org/10.1038/nmeth.2681>
- 5 **Genee, H. J., Bonde, M. T., Bagger, F. O., Jespersen, J. B., Sommer, M. O. a., Wernersson, R., & Olsen, L. R. (2014)** . Software-supported USER cloning strategies for site-directed mutagenesis and DNA assembly. *ACS Synthetic Biology*, 4(3), 342-349. <https://doi.org/10.1021/sb500194z>
- 10 **Genee H J, Bali A P, Petersen SD, Siedler, S, Bonde MT, Gronenberg LS, Kristensen, M and Harrison SJ and MOA Sommer (2016)** Functional mining of transporters using synthetic selections. *Nature Chemical Biology* 12, 1015-1022;
- Koonin, E. V., Makarova, K. S., & Zhang, F. (2017)** . Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, 37, 67-78.
- 15 <https://doi.org/10.1016/j.mib.2017.05.008>
- Lutz, R., Bujard, H., 1997.** Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Pharmacia* 25: 1203-1210.
- Martinez-Garcia, E., Aparicio, T., Gohi-Moreno, A., Fraile, S., & De Lorenzo, V. (2015).** SEVA 2.0: An update of the Standard European Vector Architecture for de/re-construction of bacterial functionalities. *Nucleic Acids Research*, 43(D1), D1183-D1189. <https://doi.org/10.1093/nar/gku114>
- Sommer, M.O., Dantas, G. & Church, G.M. (2009)** Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* 325, 1128-1131.
- 25 **Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., ... Zhang, F. (2015)** . Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. <https://doi.org/10.1016/j.cell.2015.09.038>

Claims

1. A method for *in vivo* detection of regulators of CRISPR-Cas associated activity comprising :
- 5 a. providing biological cells capable of CRISPR-Cas associated activity wherein said cells are prepared on a defined medium devoid of one or more agents for inducing expression of said activity, and either
- I. contacting said cells with a molecule or library of molecules; or
- 10 II. introducing a self-replicating library of non-host DNA fragments into said cells;
- b. culturing the cells obtained from step (a) (I) or (II) on said defined culture medium for a period sufficient for uptake or expression of a candidate regulator of CRISPR-Cas associated activity;
- 15 c. culturing the cells obtained from step (b) on said defined culture medium, wherein said medium is supplemented with one or more agents to induce expression of said CRISPR-Cas associated activity and
- d. detecting expression of a reporter gene present in cultured cells obtained from step (c),
- 20 wherein said cells comprise :
- i. a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said gene or gene cluster is operably linked to a promoter having a measured strength of >0.65 on the Anderson scale, and whose expression is inducible by at least one
- 25 of said one or more agents;
- ii. a reporter gene operably linked to a promoter;
- iii. a nucleic acid molecule operably linked to a promoter and whose expression is inducible by at least one of said one or more agents,
- 30 wherein said nucleic acid molecule encodes one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said reporter gene or its transcript, and
- wherein the polypeptide or polypeptide complex having RNA-guided endonuclease activity is capable of inactivating said reporter gene or its transcript.
- 35
2. A method for *in vivo* detection of regulators of CRISPR-Cas associated activity of claim 1, wherein said polypeptide or polypeptide complex having RNA-

guided endonuclease activity is selected from a group of CRISPR-Cas systems consisting of Type 1, Type II, Type III; Type IV, Type V, and Type VI.

- 5 3. A method for *in vivo* detection of regulators of CRISPR-Cas associated activity of claims 1 or 2, wherein the amino acid sequence(s) of said polypeptide or each member of said polypeptide complex having RNA-guided endonuclease activity has at least 80% amino acid sequence identity to an amino acid sequence or a corresponding member of a set of amino acid sequences selected from among Group A; and
- 10 wherein one of said RNA molecules comprises at least one cognate palindromic repeat sequence and one of said RNA molecules comprises a cognate trans-activating crRNA or one of said RNA molecules comprises a fusion at least one cognate palindromic repeat sequence and one cognate trans-activating crRNA, wherein said repeat sequence is encoded by a nucleotide sequence at the
- 15 corresponding position in the list of Group B; and wherein said cognate trans-activating crRNA is encoded by a nucleotide sequence at the corresponding position in the list of Group C

	Group A	Group B	Group C
Type II A	SEQ ID No.: 45	SEQ ID No.: 116	SEQ ID No.:134
Type II A	SEQ ID No.: 168	SEQ ID No.: 169	SEQ ID No.:170
Type II B	SEQ ID No.: 46	SEQ ID No.: 117	SEQ ID No.:135
Type II C	SEQ ID No.: 47	SEQ ID No.: 118	SEQ ID No.:136
Type V B	SEQ ID No.: 92-94	SEQ ID No.: 126	SEQ ID No.:137
Type V E	SEQ ID No.: 99-102	SEQ ID No.: 129	SEQ ID No.:138.

- 20 4. A method for *in vivo* detection of regulators of CRISPR-Cas associated activity of claims 1 or 2, wherein the amino acid sequence of each member of said polypeptide complex having RNA-guided endonuclease activity has at least 80% amino acid sequence identity to a corresponding member of a set of amino acid sequences selected from among Group A;
- 25 wherein said one or more RNA molecules comprises at least one cognate palindromic repeat sequence, wherein said repeat sequence is encoded by a nucleotide sequence at the corresponding position in the list of Group B

	Group A	Group B
Type I A	SEQ ID No.: 1-10	SEQ ID No.: 109
Type I B	SEQ ID No.: 12-16	SEQ ID No.: 110
Type I C	SEQ ID No.: 17-23	SEQ ID No.: 111

Type I D	SEQ ID No.: 24-31	SEQ ID No.: 112
Type I E	SEQ ID No.: 32-33	SEQ ID No.: 113
Type I F	SEQ ID No.: 34-39	SEQ ID No.: 114
Type I F*	SEQ ID No.: 40-44	SEQ ID No.: 115
Type III A	SEQ ID No.: 48-56	SEQ ID No.: 119
Type III B	SEQ ID No.: 57-64	SEQ ID No.: 120
Type III B*	SEQ ID No.: 65-73	SEQ ID No.: 121
Type III C	SEQ ID No.: 74-79	SEQ ID No.: 122
Type III D	SEQ ID No.: 80-85	SEQ ID No.: 123
Type IV	SEQ ID No.: 86-90	SEQ ID No.: 124
Type V A	SEQ ID No.: 91	SEQ ID No.: 125
Type V C	SEQ ID No.: 95-96	SEQ ID No.: 127
Type V D	SEQ ID No.: 97-98	SEQ ID No.: 128
Type VI A	SEQ ID No.: 103	SEQ ID No.: 130
Type VI B1	SEQ ID No.: 104-105	SEQ ID No.: 131
Type VI B2	SEQ ID No.: 106-107	SEQ ID No.: 132
Type VI C	SEQ ID No.: 108	SEQ ID No.: 133.

5. A method for *in vivo* detection of regulators of CRISPR-Cas associated activity according to any one of claims 1-3, wherein said first gene encodes a Cas9 polypeptide having RNA-guided endonuclease activity, and wherein one of said RNA molecules is a mat-crRNA and one of said RNA molecules is a cognate trans-activating crRNA, or said one RNA molecule is a fusion of a mat-crRNA and a cognate trans-activating crRNA.
6. A method for *in vivo* detection of regulators of CRISPR-Cas associated activity according to any one of claims 1-5, wherein the non-host DNA fragments of said self-replicating library are synthetic DNA or metagenomic DNA.
7. A method for *in vivo* detection of regulators of CRISPR-Cas associated activity according to any one of claims 1-6, wherein said biological cell is a bacterial cell.
8. A biological cell comprising:
- i. a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said first gene or first gene cluster is

- operably linked to a promoter having a measured strength of >0.65 on the Anderson scale, and whose expression is inducible by at least one agent;
- ii. a reporter gene operably linked to a promoter,
- iii. a nucleic acid molecule operably linked to a promoter and whose
- 5 expression is inducible by at least one agent,
- wherein said nucleic acid molecule encodes one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said reporter gene or its transcript and
- 10 wherein the polypeptide or polypeptide complex having RNA-guided endonuclease activity is capable of inactivating said reporter gene or its transcript.
9. A biological cell according to claim 8, where said first gene or said nucleic acid
- 15 molecule is operably linked to a constitutive promoter and a riboswitch capable of regulating translation of said gene or gene cluster in response to said at least one agent.
10. A biological cell according to claims 8 or 9, wherein said first gene or said
- 20 nucleic acid molecule is operably linked to an inducible promoter and a cis element encoding a transcription factor capable of inducing expression of said a nucleic acid molecule in response to said at least one agent.
11. A biological cell according to any one of claims 8 to 10, wherein said first gene
- 25 encodes a Cas9 polypeptide having RNA-guided endonuclease activity, and wherein one of said RNA molecules is a mat-crRNA and one of said RNA molecules is a cognate trans-activating crRNA, or said one RNA molecule is a fusion of a mat-crRNA and a cognate trans-activating crRNA.
12. A biological cell according to any one of claims 8 to 11, wherein said cell is a
- 30 bacterial cell.
13. Use of a biological cell according to any one of claims 8 to 12, for detecting a regulator of CRISPR-Cas associated activity.
14. A biological cell library of non-host DNA fragments, wherein each cell of the
- 35 library is a cell according to any one of claims 8 to 12, and wherein each cell further comprises a self-replicating genetic element comprising
- a. a non-host DNA fragment, and

- b. a gene encoding a selectable marker polypeptide and operably linked to a promoter.

5 **15.** Use of a biological cell library of non-host DNA fragments according to claim 14, for detecting a regulator of CRISPR-Cas associated activity.

Figure 1

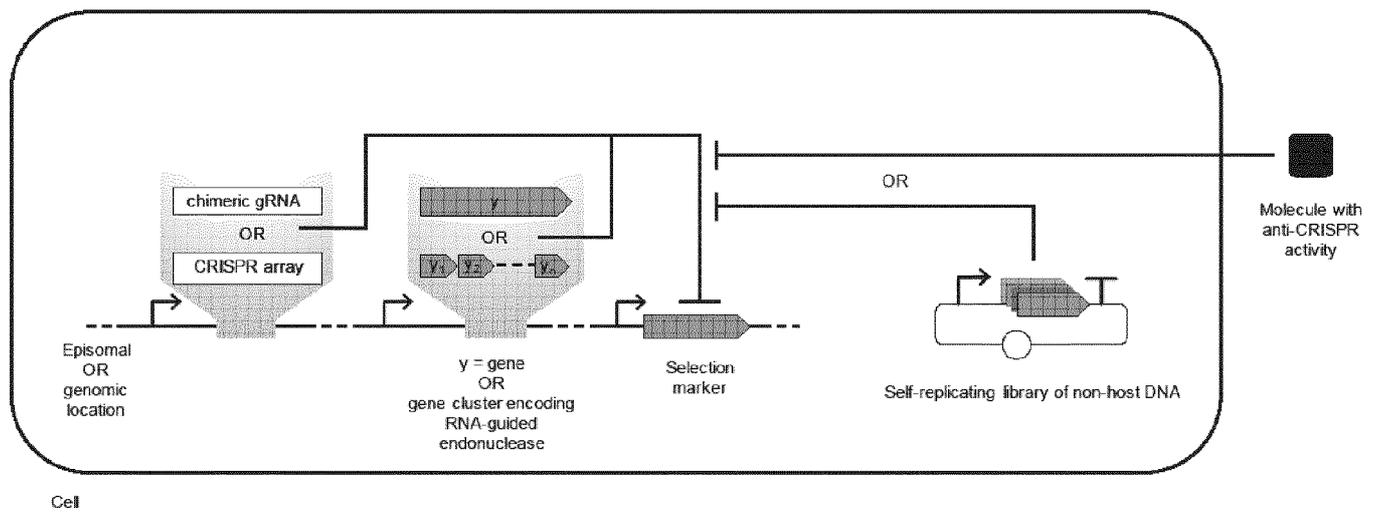


Figure 2

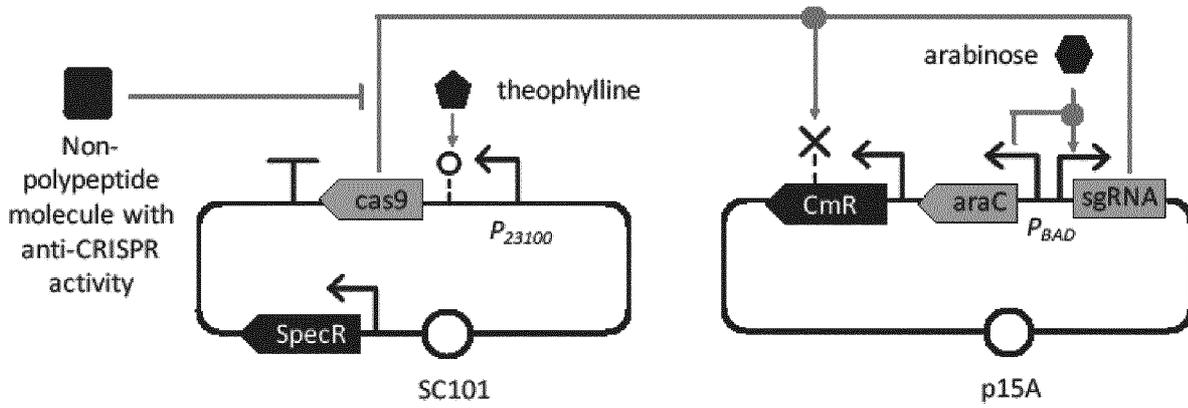
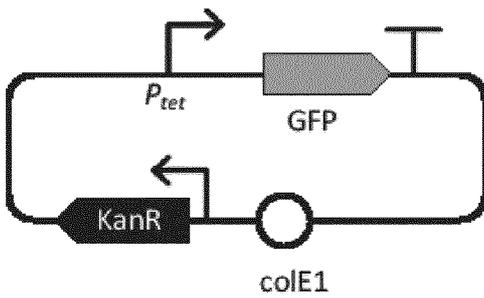


Figure 3

A



B

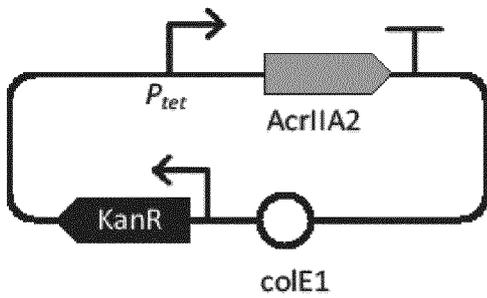


Figure 4

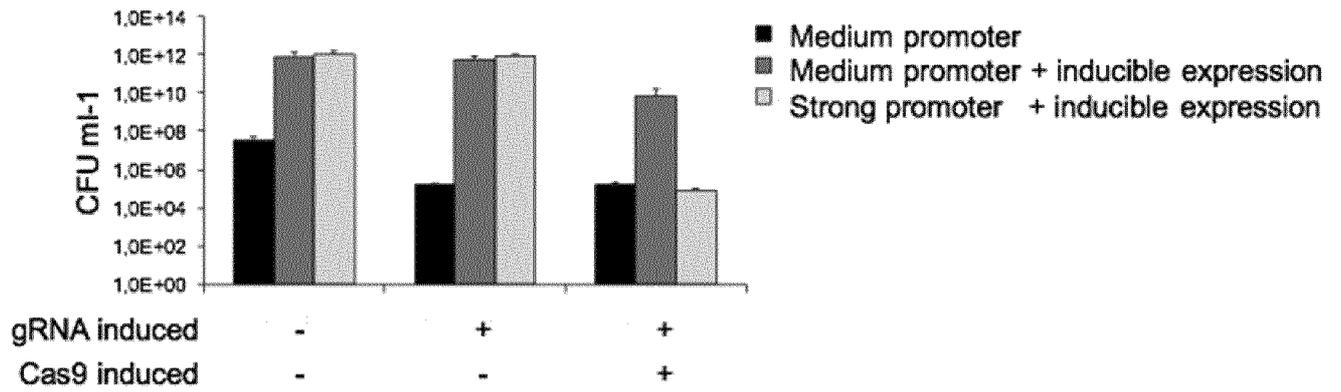
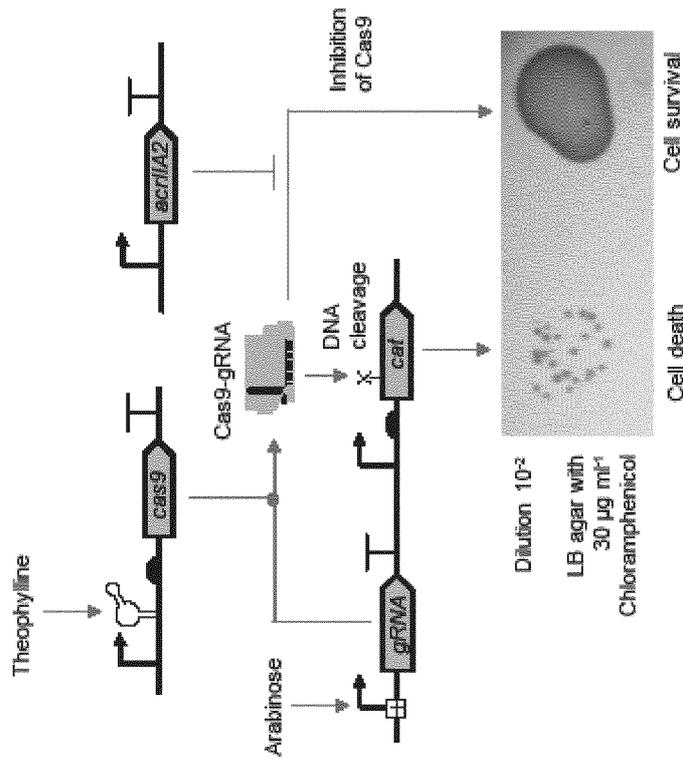


Figure 5

A



B

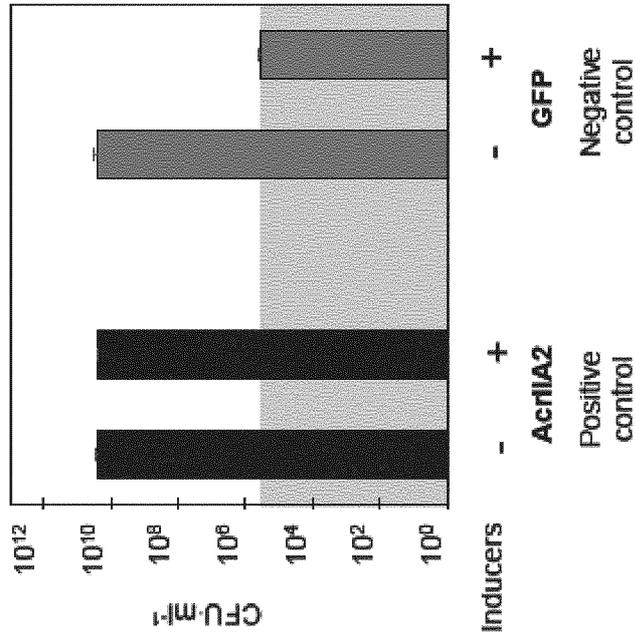
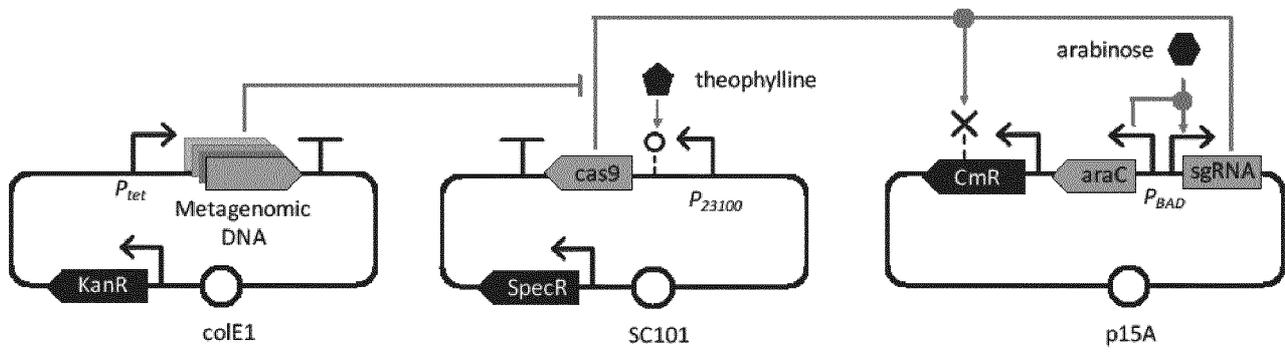


Figure 6

A



B

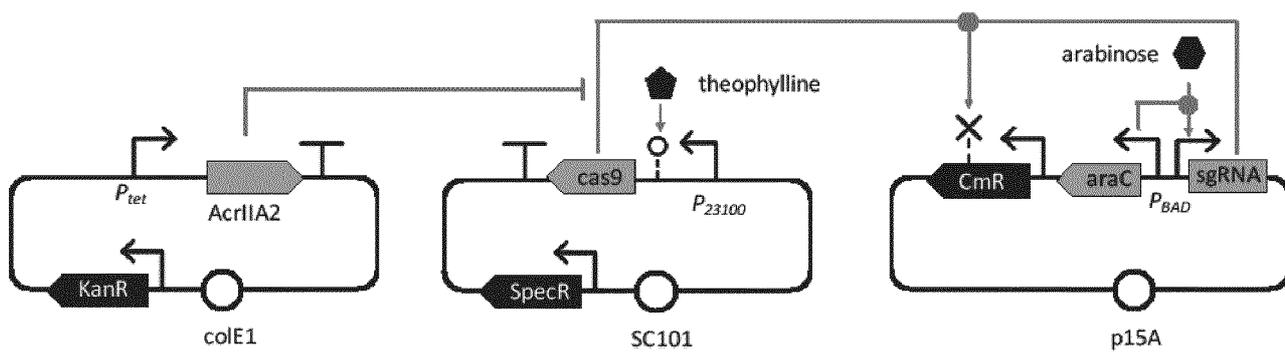


Figure 7

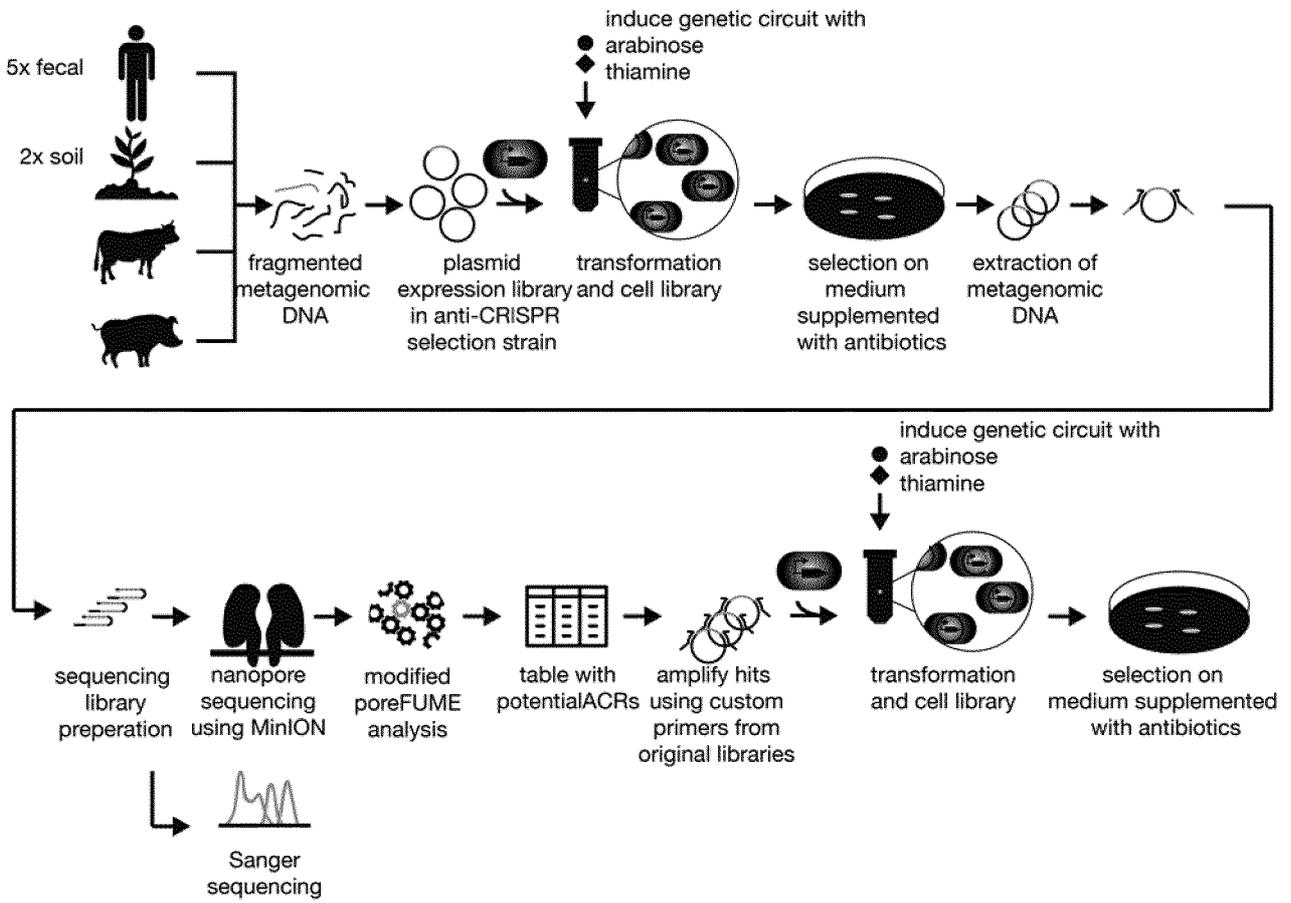


Figure 8

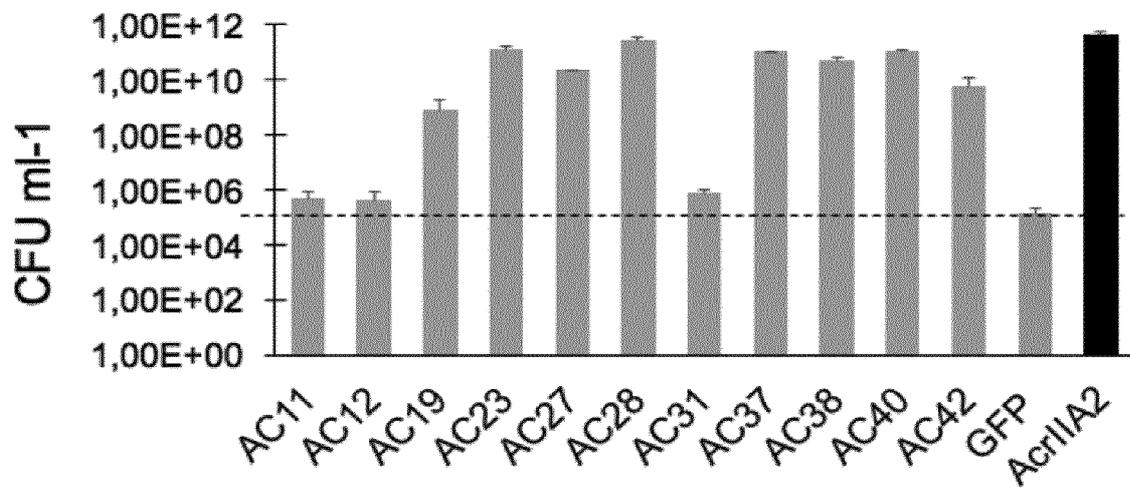


Figure 9

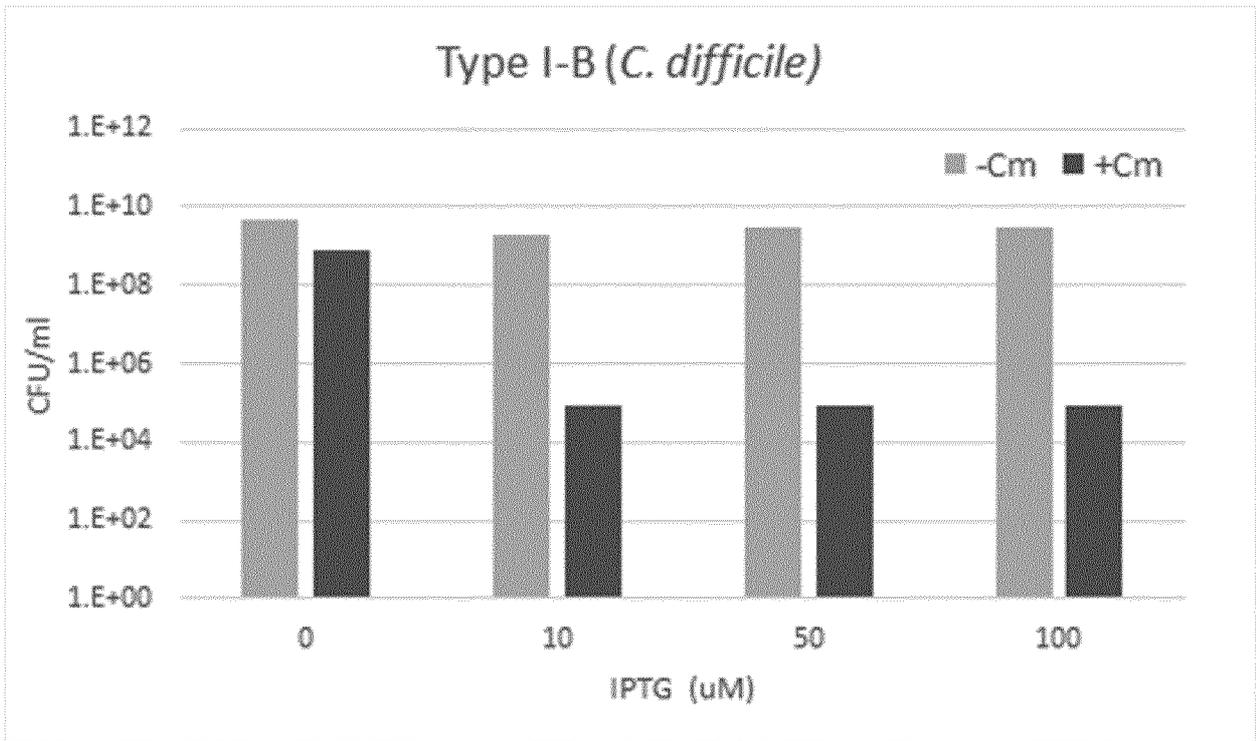


Figure 10

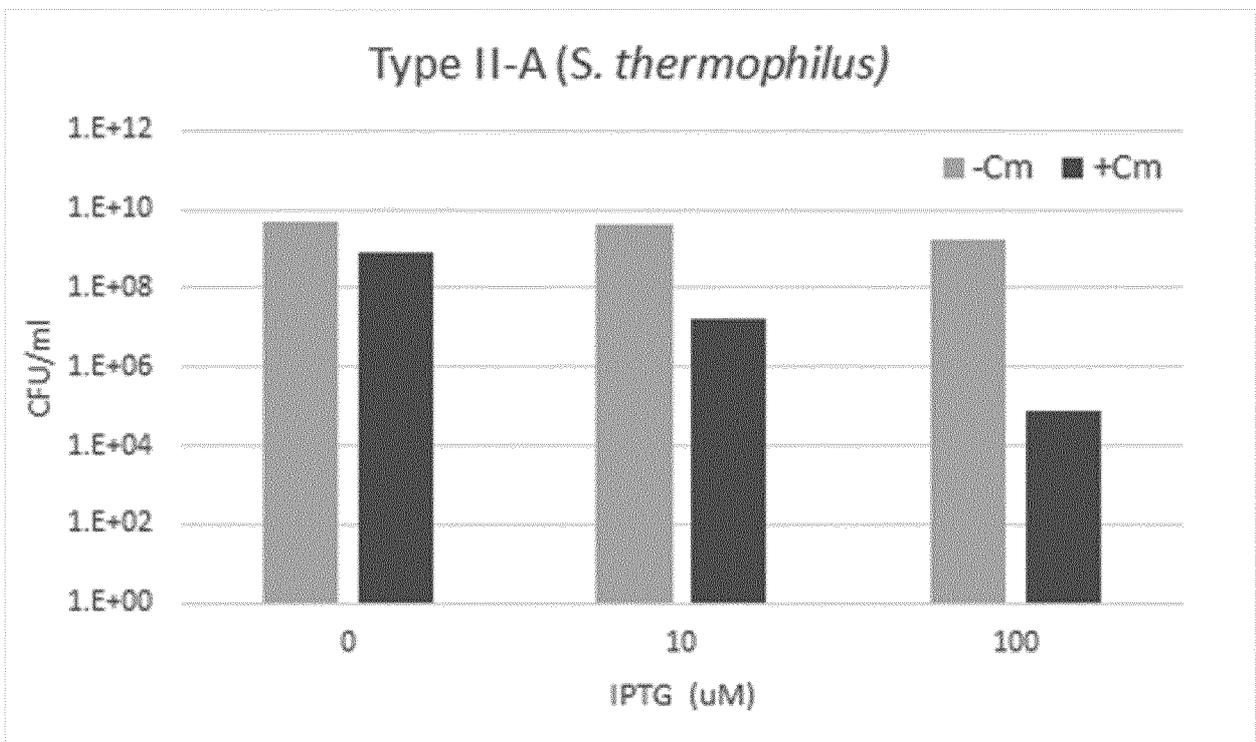


Figure 11

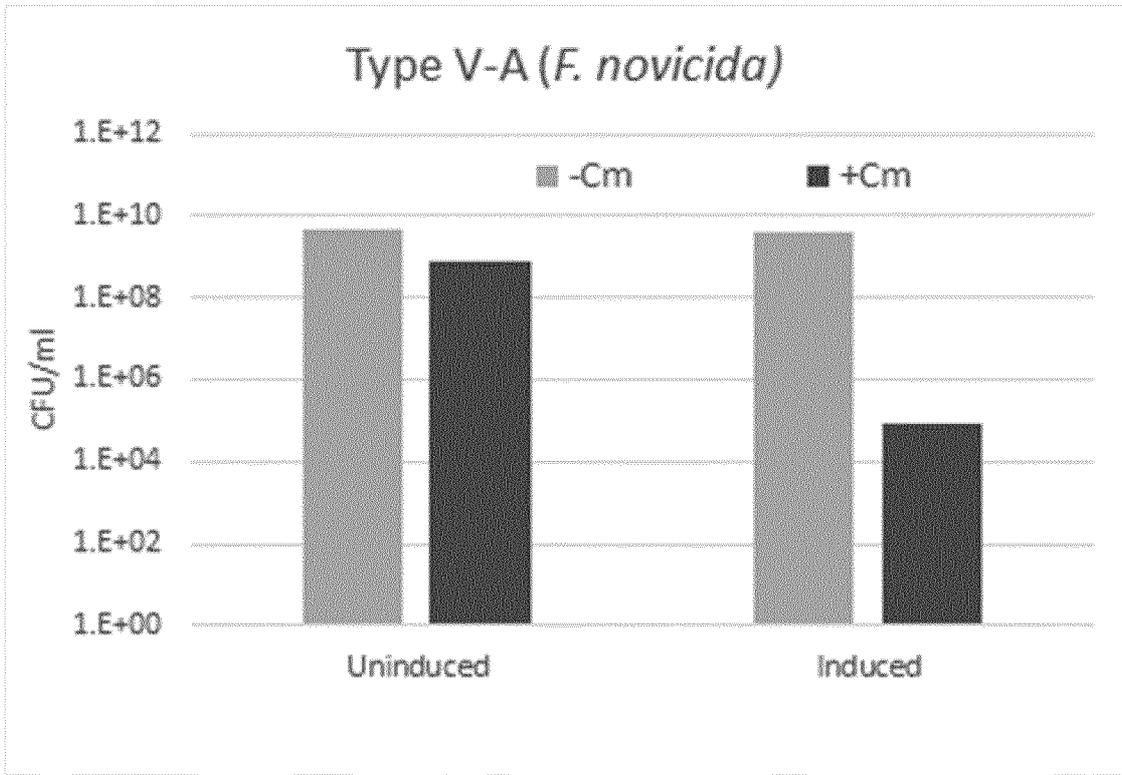


Figure 12

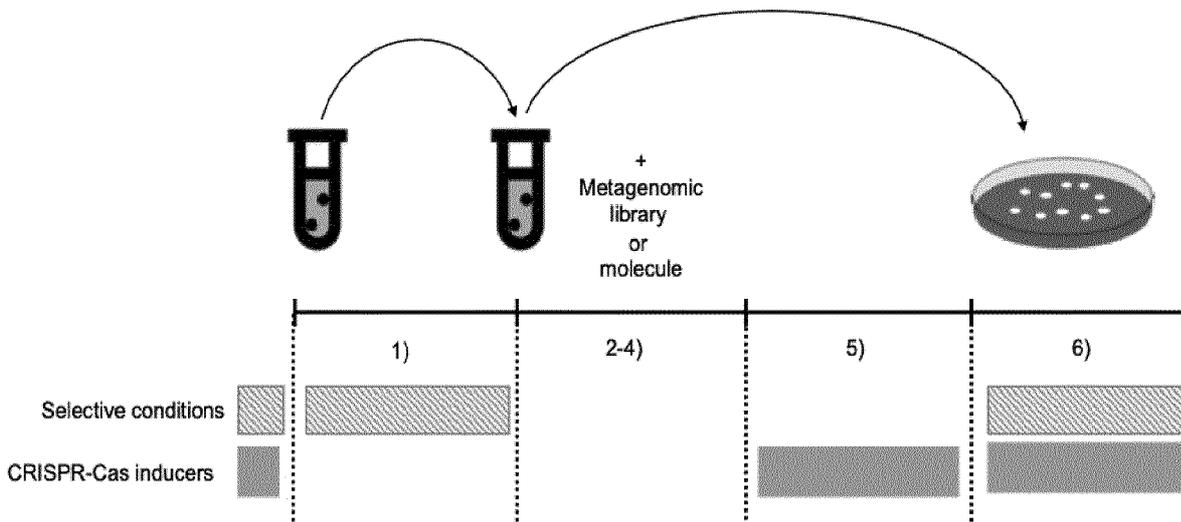


Figure 13

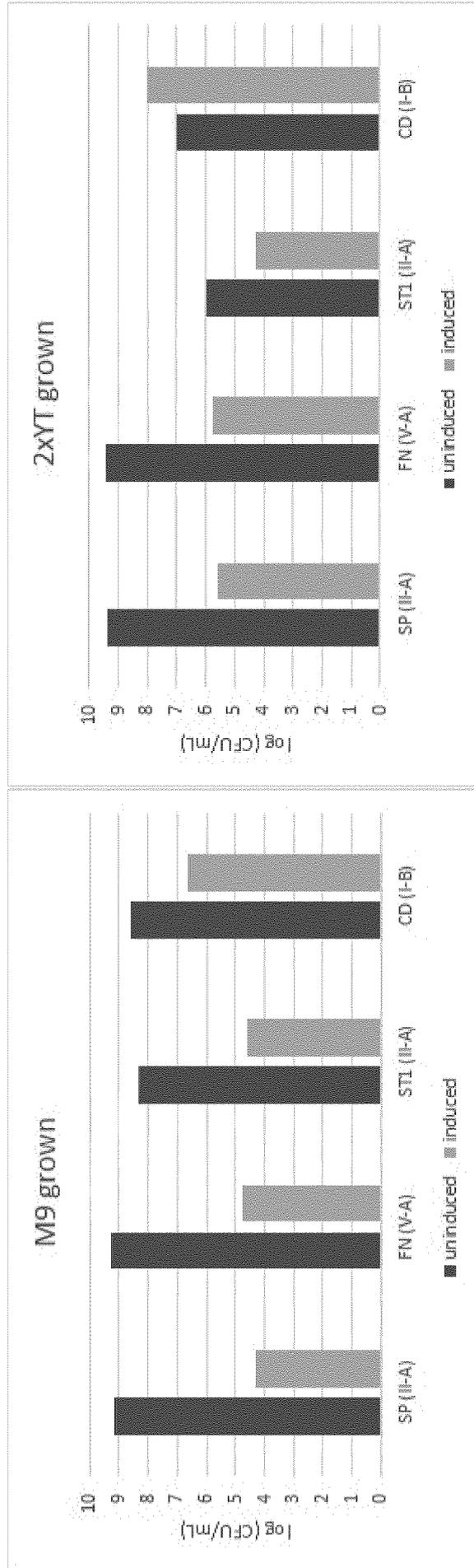
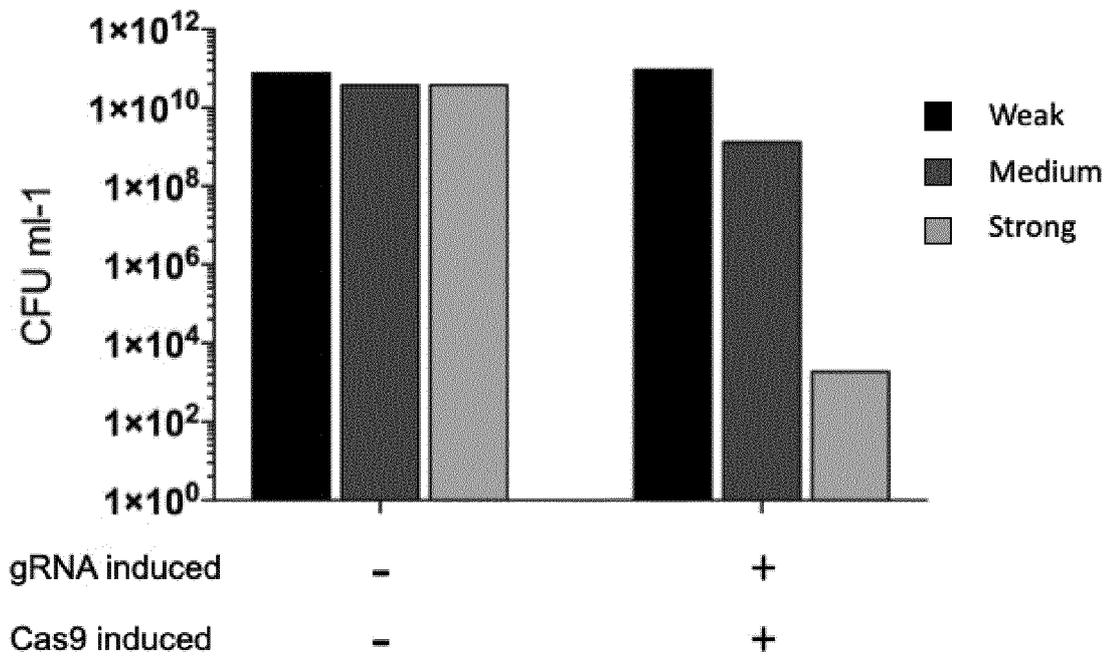


Figure 14



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057785

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/10
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)
C12N C12Q

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>JOSEPH BONDY-DENOMY ET AL: "Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins", NATURE, vol. 526, no. 7571, 23 September 2015 (2015-09-23), pages 136-139, XP055474577, GB ISSN: 0028-0836, DOI: 10.1038/nature15254 see abstract and pages 136-137 and Fig. 1-2,4</p> <p style="text-align: center;">----- -/--</p>	1-15

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"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</p> <p>"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</p> <p>"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</p> <p>"&" document member of the same patent family</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Date of the actual completion of the international search 17 April 2019	Date of mailing of the international search report 07/05/2019
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 Vix, Olivier

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057785

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>RAUCH BENJAMIN J ET AL: "Inhibition of CRISPR-Cas9 with Bacteriophage Proteins", CELL, CELL PRESS, AMSTERDAM, NL, vol. 168, no. 1, 29 December 2016 (2016-12-29), page 150, XP029882136, ISSN: 0092-8674, DOI: 10.1016/J.CELL.2016.12.009 see abstract and pages 151-155 and Fig. 2, 6</p> <p style="text-align: center;">-----</p>	1-15
Y	<p>A. PAWLUK ET AL: "A New Group of Phage Anti-CRISPR Genes Inhibits the Type I-E CRISPR-Cas System of Pseudomonas aeruginosa", MBIO, vol. 5, no. 2, 15 April 2014 (2014-04-15), pages e00896-14, XP055474866, DOI: 10.1128/mBio.00896-14 the whole document</p> <p style="text-align: center;">-----</p>	1-15
Y	<p>DE DONG ET AL: "Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein", NATURE, vol. 546, no. 7658, 27 April 2017 (2017-04-27), pages 436-439, XP055452726, GB ISSN: 0028-0836, DOI: 10.1038/nature22377 the whole document</p> <p style="text-align: center;">-----</p>	1-15
Y	<p>YUWEI ZHU ET AL: "Structural insights into the inactivation of CRISPR-Cas systems by diverse anti-CRISPR proteins", BMC BIOLOGY, vol. 16, no. 1, 19 March 2018 (2018-03-19), XP055474878, DOI: 10.1186/s12915-018-0504-9 the whole document</p> <p style="text-align: center;">-----</p>	1-15
Y	<p>YANG HUI ET AL: "Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4 Targeting SpyCas9", MOLECULAR CELL, ELSEVIER, AMSTERDAM, NL, vol. 67, no. 1, 9 June 2017 (2017-06-09), page 117, XP085122090, ISSN: 1097-2765, DOI: 10.1016/J.MOLCEL.2017.05.024 the whole document</p> <p style="text-align: center;">-----</p> <p style="text-align: center;">-/--</p>	1-15

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057785

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	APRIL PAWLUK ET AL: "Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species", NATURE MICROBIOLOGY, vol. 1, no. 8, 13 June 2016 (2016-06-13), XP055474868, DOI: 10.1038/nmicrobiol.2016.85 the whole document -----	1-15
Y	WO 2017/160689 A1 (UNIV MASSACHUSETTS [US]) 21 September 2017 (2017-09-21) see claims 1-6 and pages 2-8 -----	1-15
X	WO 2016/108926 A1 (BROAD INST INC [US]; MASSACHUSETTS INST TECHNOLOGY [US]) 7 July 2016 (2016-07-07) -----	8-10,12, 14
Y	see claims 1-3, 10-1, pages 2-9 and also [00256] -----	1-15
X	WO 2016/011080 A2 (UNIV CALIFORNIA [US]) 21 January 2016 (2016-01-21) -----	8-10,12, 14
Y	see claims 1 and 56 and pages 2-6 -----	1-15
Y	WO 2017/186854 A1 (DANMARKS TEKNISKE UNIV [DK]) 2 November 2017 (2017-11-02) see claims and pages 2-4 -----	1-7
Y	APRIL PAWLUK ET AL: "Anti-CRISPR: discovery, mechanism and function", NATURE REVIEWS. MICROBIOLOGY, vol. 16, no. 1, 24 October 2017 (2017-10-24), pages 12-17, XP055475543, GB ISSN: 1740-1526, DOI: 10.1038/nrmicro.2017.120 in particular see page 16 conclusion; the whole document -----	1-15
Y	CHUAN DONG ET AL: "Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins", NUCLEIC ACIDS RESEARCH, vol. 46, no. D1, 25 September 2017 (2017-09-25), pages D393-D398, XP055475568, ISSN: 0305-1048, DOI: 10.1093/nar/gkx835 the whole document -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2019/057785

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2017160689 A1	21-09-2017	CA 3029735 A1	21-09-2017
		CN 109152848 A	04-01-2019
		EP 3429635 A1	23-01-2019
		WO 2017160689 A1	21-09-2017

WO 2016108926 A1	07-07-2016	US 2018112255 A1	26-04-2018
		WO 2016108926 A1	07-07-2016

WO 2016011080 A2	21-01-2016	CA 2954791 A1	21-01-2016
		EP 3169776 A2	24-05-2017
		US 2017204407 A1	20-07-2017
		WO 2016011080 A2	21-01-2016

WO 2017186854 A1	02-11-2017	NONE	
