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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 November 2016 (17.11.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/183448 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/85* (2006.01)
C12N 15/10 (2006.01)

(21) International Application Number:
PCT/US2016/032367

(22) International Filing Date:
13 May 2016 (13.05.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/161,487 14 May 2015 (14.05.2015) US

(71) Applicant: **UNIVERSITY OF SOUTHERN CALIFORNIA** [US/US]; 1150 South Olive Street, Suite 2300, Los Angeles, California 90015 (US).

(72) Inventors: **JUNGE, Jason**; 1441 Indiana Ave., South Pasadena, California 91030 (US). **HUNT, Timothy**; 1338 13th Ave S., Seattle, Washington 98144 (US). **FRASER, Scott**; 1845 Niodrara Dr., Glendale, California 91208 (US).

(74) Agents: **CHEN, Stephen, W.** et al.; Nixon Peabody LLP, 300 South Grand Avenue, Suite 4100, Los Angeles, California 90071-3151 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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).

Published:

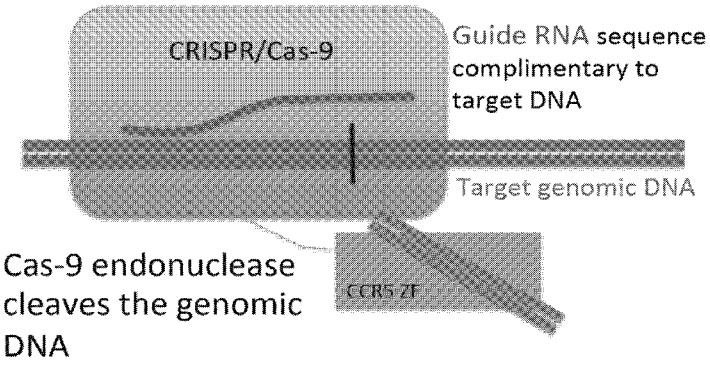
— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

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

(54) Title: OPTIMIZED GENE EDITING UTILIZING A RECOMBINANT ENDONUCLEASE SYSTEM

Figure 1



(57) Abstract: Described herein are methods and compositions for genomic editing. Endonucleases for genomic editing involve inducing breaks in double stranded DNA, for which knock-ins are notoriously inefficient for relying on random integration of homologous DNA sequences into the break site by repair proteins. To address these issues, described herein are novel recombinant fusion proteins that actively recruit linear DNA inserts in closer proximity to the genomic cleavage site, increasing integration efficiency of large DNA fragments into the genome. Such improvements to genomic editing technology allow one to use lower linear DNA concentrations without sacrificing efficiency and can be further combined with other features, such as fluorescent protein reporting systems.

WO 2016/183448 A1

OPTIMIZED GENE EDITING UTILIZING A RECOMBINANT ENDONUCLEASE SYSTEM

FIELD OF THE INVENTION

Described herein are methods and compositions for use in *in vitro* and *in vivo* manipulation of genetic sequences for research and therapeutic activities, including generation of knockout and knock-in sequences for genomic editing.

BACKGROUND

CRISPR-Cas genome editing has become a more mature technique within the last 10 years. Like all other genome editing techniques, it is adequate at generating knockout phenotypes, but not very efficient at introducing knock-in of donor DNA sequences. The development of more useful enzymes for biotechnology and primary research has included improving efficacy of enzymes, creating new enzymes to fill needs, or adapting them to expand their functionality. For example, Klenow exonuclease minus enzyme is a truncation of the Klenow large-fragment. This modified DNA polymerase retains its polymerase activity while losing the large-fragment's ability to chew back 3' DNA overhangs. Another example includes Phusion® DNA polymerase, a proprietary fusion protein consisting of a proofreading DNA polymerase plus the domain of another DNA polymerase that confers higher copying rate. Extending these principles to the realm of genome editing, it should be possible to create a new Cas-9 protein that preserves the existing DNA endonuclease activity while increasing the integration efficiency of donor DNA sequence into the genome. Several strategies are proposed to address this need.

There are several obstacles in particular that concern insertion of donor DNA into the genome. Non-homologous end joining (NHEJ), while an appropriate method for knocking out genes, is wholly insufficient for integration of donor DNA without the presence of any other mutations. Homology directed repair (HDR) encompasses at least two forms: a long sequence homology arm requiring Holliday structures to resolve integrations, or short homology arms which rely on strand invasion and the emergency DNA repair system to resolve the integration events. Any combination of HDR and NHEJ will predominantly yield indel (insertion and deletion) mutations within the portion of the damaged DNA resolved by NHEJ. In essence, stochastic processes determine the possible integration of homologous DNA and plague efficient genomic editing for knock-ins, including the inversely

proportional relationship between the length of the donor DNA fragment and the efficiency of integration, integration efficiency of donor DNA is directly proportional to the donor DNA concentration in the cell and cytotoxic overabundance of linear DNA fragments, and spatial restriction of the two free ends of a double stranded break therefore causing a most likely outcome of re-annealing of the two ends.

Described herein is the development of recombinant fusion protein that actively recruit linear DNA inserts in closer proximity to the genomic cleavage site, thereby allowing increasing integration efficiency, particularly of large DNA fragments, into the genome. Such improvements to genomic editing technology allow one to use lower linear DNA concentrations without sacrificing efficiency and can be further combined with other features, such as fluorescent protein reporting systems.

SUMMARY OF THE INVENTION

Described herein is a composition including a vector encoding a fusion protein including at least one endonuclease and a DNA binding moiety. In various embodiments, the fusion protein includes at least one endonuclease selected from the group consisting of: cas regularly interspaced short palindromic (CRISPR) protein, a zinc finger nuclease (ZFNs) and transcription activator-like effector nucleases (TALENs). In various embodiments, the CRISPR protein includes cas9. In various embodiments, the DNA binding moiety includes a zinc finger protein. In various embodiments, the zinc finger includes a left handed CCR5 binding protein. In various embodiments, the at least one endonuclease and DNA binding moiety are joined by a linker including two, three, four, five, six, seven, eight, nine, ten or more amino acids. In various embodiments, the fusion protein includes a fluorescent labeled protein. In various embodiments, the fluorescent labeled protein includes one or more proteins selected from the group consisting of: green fluorescent protein (GFP), enhanced (eGFP), red fluorescent protein (RFP) and mCherry. In various embodiments, the fusion protein includes a nuclear localization signal (NLS). In various embodiments, the NLS is SV40 NLS.

Further described herein is a method of genomic editing including (a) providing a quantity of one or more vectors encoding a fusion protein including at least one endonuclease and a DNA binding moiety and (b) contacting a population of cells with the quantity of the one or more vectors, wherein the fusion protein is capable of inducing double stranded break (DSB) and homologous recombination (HR) of the DSB results in editing of the

genome of the population of cells. In various embodiments, the method includes contacting the population of cells with one or more guide RNAs (gRNAs) in step (b). In various embodiments, the method includes contacting the population of cells with template DNA in step (b). In various embodiments, the template DNA includes at least one expression cassette, two flanking sequences, and a DNA binding moiety sequence. In various embodiments, the two flanking sequences are each at least 10 bp, and homologous to sequences in the genome of the population of cells. In various embodiments, the DNA binding moiety sequence includes CCR5. In various embodiments, contacting the population of cells includes a technique selected from the group consisting of: transfection, electroporation, and transformation. In various embodiments, the population of cells includes stem cells or progenitor cells.

Described herein is a kit for genomic editing including one or more vectors encoding a fusion protein including at least one endonuclease and a DNA binding moiety and template DNA including at least one expression cassette, two flanking sequences, and a DNA binding moiety sequence. In various embodiments, the kit includes one or more guide RNAs (gRNAs). In various embodiments, the kit fusion protein includes at least one endonuclease CRISPR protein, a DNA binding moiety that is a zinc finger protein. In various embodiments, the kit includes a fluorescent labeled protein. In various embodiments, the kit includes a nuclear localization signal (NLS).

Also described herein is a method of multi-locus genomic editing including inducing sticky end formation at one or more loci by adding a CRISPR protein, providing a quantity of one or more guide strand RNAs, ligating one or more single-stranded donor DNA, hybridizing one or more double-stranded DNA with a terminating oligonucleotide, synthesis of one or more double stranded DNA from the one or more single-stranded donor DNA to the one or more double-stranded DNA completing the donor DNA strand to form a sticky end, and joining compatible sticky ends at one or more loci.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1. Recombinant Cas-9 design. Conventionally, guide RNA (gRNA) is complementary to target DNA, and Cas-9 endonuclease cleaves the genomic, with stochastic processes determining the possible integration of homologous DNA. To shunt these processes towards favorable recombination events, a recombinant fusion protein including CCR5 zinc-finger as a DNA binding moiety to bind the homologous linear DNA fragment to

be inserted, and bringing the fragment closer to the double stranded break (DSB). A recombinant Cas-9 that aids in increasing proximity of DSB and linear DNA will lead to more efficient generation of genome editing.

Figure 2. Comparison between wildtype Cas-9 and recombinant Cas-9 using GFP as reporter and beta-actin as the target gene. Wild-type condition included 5 out of 18 CFP positive colonies that were also GFP positive, 2 of which were only GFP positive (i.e., stable integrants). Under Cas-9 zinc finger conditions included 9 out of 12 CFP positive colonies that were also GFP positive, 4 of which were only GFP positive. Scale bars = 10 μM.

Figure 3: **(Fig. 3A)** Data showing that the addition of the Zinc-finger to Cas9 does not confer additional zinc-finger dependent endonuclease activity. Donor DNA (900 bp) contains the CCR5 DNA binding sequence (base pairs 80-92 on the 5' end of the linear DNA. Combinations of protein and guide strand are labeled for each condition. **(Fig. 3B)** Three examples of GFP signal due to integration of donor DNA. The GFP positive clones are being validated. mCherry is the internal viability control.

Figure 4: Integration of GFP in beta-actin validated by PCR. PCR of GFP is a 300 bp fragment of the total GFP sequence, beta-actin reverse primer anneals 300 bp downstream of the sfGFP integration site.

Figure 5: Cas9-zinc finger fusion protein linearizes plasmid DNA efficiently. It appears there off-target cleavage that occurs perhaps due to the zinc finger, although off-target cleavage is ameliorated when the donor DNA is included. Lanes 5 and 6 are one sample loaded into two wells because well 5 was partially occluded.

Figure 6: Parallel multi-locus directed by guide-strand RNA mixture. Begins with T7 RNA synthesis of guide-strand RNA **(Fig. 6A)**, followed by splint **(Fig. 6B)** mediated ligation **(Fig. 6C)** of single-stranded donor DNA sequence **(Fig. 6D)** onto the 5' end of the guide-strand RNA. The 5' nucleotide of the synthesized donor DNA **(Fig. 6E)** has an exonuclease resistant phosphorothioate bond with its neighboring nucleotide. Thereafter, hybridization of double-stranded DNA terminating oligonucleotide **(Fig. 6F)** is followed by an isothermal DNA polymerase (Klenow exo-) reaction to fill-in the double strandedness **(Fig. 6G)** from the splinting primer (step 2) to the terminating oligo (step 3) and T4-DNA ligase reaction **(Fig. 6H)** to complete the donor fragment with the appropriate 5' sticky end **(Fig. 6I)** for donor DNA sticky end ligation to the sticky end of the genomic DNA **(Fig. 6J)** digested by cpf1. Complete integration happens upon ligation of the sticky end **(Fig. 6K)** of

the donor DNA and homologous recombination between donor and genomic DNA **(Fig. 6L)** due to their proximity and sequence homology.

DETAILED DESCRIPTION OF THE INVENTION

All references cited herein are incorporated by reference in their entirety as though fully set forth. Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Allen *et al.*, *Remington: The Science and Practice of Pharmacy 22nd ed.*, Pharmaceutical Press (September 15, 2012); Hornyak *et al.*, *Introduction to Nanoscience and Nanotechnology*, CRC Press (2008); Singleton and Sainsbury, *Dictionary of Microbiology and Molecular Biology 3rd ed., revised ed.*, J. Wiley & Sons (New York, NY 2006); Smith, *March's Advanced Organic Chemistry Reactions, Mechanisms and Structure 7th ed.*, J. Wiley & Sons (New York, NY 2013); Singleton, *Dictionary of DNA and Genome Technology 3rd ed.*, Wiley-Blackwell (November 28, 2012); and Green and Sambrook, *Molecular Cloning: A Laboratory Manual 4th ed.*, Cold Spring Harbor Laboratory Press (Cold Spring Harbor, NY 2012), provide one skilled in the art with a general guide to many of the terms used in the present application. For references on how to prepare antibodies, see Greenfield, *Antibodies A Laboratory Manual 2nd ed.*, Cold Spring Harbor Press (Cold Spring Harbor NY, 2013); Köhler and Milstein, *Derivation of specific antibody-producing tissue culture and tumor lines by cell fusion*, Eur. J. Immunol. 1976 Jul, 6(7):511-9; Queen and Selick, *Humanized immunoglobulins*, U. S. Patent No. 5,585,089 (1996 Dec); and Riechmann *et al.*, *Reshaping human antibodies for therapy*, Nature 1988 Mar 24, 332(6162):323-7.

One skilled in the art will recognize many methods and materials similar or equivalent to those described herein, which could be used in the practice of the present invention. Indeed, the present invention is in no way limited to the methods described herein. For purposes of the present invention, the following terms are defined below.

"Administering" and/or "administer" as used herein refer to any route for delivering a pharmaceutical composition to a patient. Routes of delivery may include non-invasive peroral (through the mouth), topical (skin), transmucosal (nasal, buccal/sublingual, vaginal, ocular and rectal) and inhalation routes, as well as parenteral routes, and other methods known in the art. Parenteral refers to a route of delivery that is generally associated with injection, including intraorbital, infusion, intraarterial, intracarotid, intracapsular, intracardiac, intradermal, intramuscular, intraperitoneal, intrapulmonary, intraspinal,

intrasternal, intrathecal, intrauterine, intravenous, subarachnoid, subcapsular, subcutaneous, transmucosal, or transtracheal. Via the parenteral route, the compositions may be in the form of solutions or suspensions for infusion or for injection, or as lyophilized powders.

"Modulation" or "modulates" or "modulating" as used herein refers to upregulation (i.e., activation or stimulation), down regulation (i.e., inhibition or suppression) of a response or the two in combination or apart.

"Pharmaceutically acceptable carriers" as used herein refer to conventional pharmaceutically acceptable carriers useful in this invention.

"Promote" and/or "promoting" as used herein refer to an augmentation in a particular behavior of a cell or organism.

"Subject" as used herein includes all animals, including mammals and other animals, including, but not limited to, companion animals, farm animals and zoo animals. The term "animal" can include any living multi-cellular vertebrate organisms, a category that includes, for example, a mammal, a bird, a simian, a dog, a cat, a horse, a cow, a rodent, and the like. Likewise, the term "mammal" includes both human and non-human mammals.

"Therapeutically effective amount" as used herein refers to the quantity of a specified composition, or active agent in the composition, sufficient to achieve a desired effect in a subject being treated. A therapeutically effective amount may vary depending upon a variety of factors, including but not limited to the physiological condition of the subject (including age, sex, disease type and stage, general physical condition, responsiveness to a given dosage, desired clinical effect) and the route of administration. One skilled in the clinical and pharmacological arts will be able to determine a therapeutically effective amount through routine experimentation.

"Treat," "treating" and "treatment" as used herein refer to both therapeutic treatment and prophylactic or preventative measures, wherein the object is to prevent or slow down (lessen) the targeted condition, disease or disorder (collectively "ailment") even if the treatment is ultimately unsuccessful. Those in need of treatment may include those already with the ailment as well as those prone to have the ailment or those in whom the ailment is to be prevented.

Genome editing has been a hot area of research for the past decade with an emphasis on generating genetic knockouts for genes of interest within model animals. Most methods for knocking in donated genetic sequences have typically relied on either brute force methods

with random sites of integration or on homologous recombination events. Some of these techniques only work in defined model organisms, while others have only limited application. Therefore, a more wide-ranging and efficient toolbox for the integration of donor DNA with sequences ranging from 300-3,500 bp in length would present solutions for answering certain questions within biology, more rapid generation of useful transgenic animals, and making theranostics cheaper, faster, and more accessible to clinicians and their patients. This toolbox could make even a small lab more efficient at making "footprint-free" transgenic cell lines/animals using industry standard cell culture techniques, quickly, and with fewer attempts at a reasonable overhead cost.

A fundamental concern regarding donor DNA integration, on top of the problems addressed previously about mutations peri-integration, is the inversely proportional relationship between the length of the donor DNA fragment and the efficiency of integration. It has been reported that 27 and 54 bases of inserted donor DNA sequence with 41 and 49 bases of sequence homology flanking those 27 bases (117 total length) or 33 base pairs on either side of 54 base pairs (120bp total) result in the following efficiencies: 64% mutation rate, 3.2% integration rate, with half of those lacking other mutations for the 117 base donor DNA; and 86% mutation rate, 15.6% integration, with 3.5% precise integration. Other have used TALEN endonucleases combined with very creative mathematics to report a mutation rate of 0% and a 50% integration rate (but using worst case scenario could be as low as 8.9%) and 1.5% germline transmission of 700 bases encoding GFP flanked by 827 bases and 904 bases of homology to the integration site.

Second, integration efficiency of donor DNA into the genome is directly proportional to the donor DNA concentration in the cell. The more abundant a fragment is in the cell, the more likely it is to participate in the DNA repair mechanism. A problem arises from overabundance of linear DNA fragments causing cytotoxicity due in large part to the innate ability of cells to defend themselves against DNA viruses, and also to saturating the cell's ability to recover from the endonuclease damage. The inability of a cell to discriminate between damaged DNA ends and the donor DNA ends could leave the genomic DNA unrepaired.

Third, the nucleus of a cell is a structure densely packed with genomic DNA. The two free ends of a double stranded break in the genomic DNA are spatially restricted, cannot diffuse away from each other, and therefore the most likely outcome is re-annealing of the two ends with some sequence added and/or removed at the locus of endonuclease activity.

Genomic Editing. What has been shown up to this point is that genome engineering is versatile and powerful tool to correct genetic mutations. Site-specific chromosomal integration can target desired nucleotide changes, including introducing therapeutic gene cassettes in safe landing sites within chromosomes, disrupting the coding or non-coding regions of specific alleles and correcting the genetic mutations to reverse the disease phenotype. Conventional technologies such as Zinc Finger Nuclease (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs) have provided a significant groundwork of proof-of-concept studies for genome editing and therapy. Yet, the most recent advances in Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated endonuclease protein (cas) system extend this versatility and convenience by reducing the number of steps required for designing targeting of a particular mutation.

Briefly, genome editing using tools such as ZFNs can be based on the introduction of a site-specific DNA double stranded break (DSB) into the locus of interest. Key to this process is the cellular repair mechanism for efficiently repairing DSBs via the homology-directed repair (HDR), or non-homologous end joining (NHEJ) pathways. The mechanisms of these DNA repair pathways can generate defined genetic outcomes. More specifically, genome editing using ZFNs can be based on the introduction of a site-specific DNA DSB into the locus of interest. Thereafter, NHEJ repair, can rapidly and efficiently ligate two broken ends, providing opportunity for the gain or loss of genetic information. This feature can be exploited to introduce small insertions and/or deletions at the site of the break, thereby allowing disruption of a target gene. If, for example, a disease results from toxic protein buildup, instruction of a nonsense or missense sequence effectively eliminates aberrant protein to correct human disorders caused by inherited gene defects. Alternatively, if a specifically-designed homologous donor DNA is provided in combination with the ZFNs, this template can result in gene correction or insertion, as repair of the DSB can include a few nucleotides changed at the endogenous site or the addition of a new gene at the break site.

While pioneering much of what is known about genomic editing process, significant challenges exist with conventional technologies such as ZFNs, and TALENs. These early generation nucleases, ZFNs and TALENs are artificial fusion proteins composed of an engineered DNA binding domain fused to a non-specific nuclease cleavage domain from the FokI restriction enzyme. Zinc finger and transcription activator-like effector repeat domains with customized specificities can be joined to bind to extended DNA sequences. While

adaptation of ZFNs and TALENs by modifying the DNA-binding specificities provide a significant level of targeting control, individual zinc finger domains provide some heterogeneity requiring some context-dependence for DNA binding. TALE repeat domains appear less susceptible to these context-dependent effects and can be modularly assembled to recognize virtually any DNA sequence via a simple one-to-one code between individual repeats and the four possible DNA nucleotides, but assembly of DNAs encoding large numbers of highly conserved TALE repeats can require the use of non-standard molecular biology cloning methods.

Whereas both ZFNs and TALENs involving use protein–DNA interactions for targeting, bacterial CRISPR-Cas system is unique and flexible due to utilization of RNA as the moiety that targets the nuclease to a desired DNA sequence. In contrast to ZFN and TALEN platforms, CRISPR-CAS uses simple Watson-Crick base pairing rules between an engineered RNA and the target DNA site. Generally, two components form the core of a CRISPR nuclease system, a Cas nuclease (e.g., cas9) and a guide RNA (grNA), the gRNA derived from a fusion of CRISPR-derived RNAs ("crRNA") and trans-acting antisense RNA ("tracRNA"). In the most well-studied example, the single gRNA complexes with a cas protein (e.g., cas9) to mediate cleavage of target DNA sites that are complementary to the first (5′) 20 nts of the gRNA and that lie next to a protospacer adjacent motif ("PAM") sequence (canonical form of 5'-NGG for *Streptococcus pyogenes* cas9, but also alternate 5'-NAG exist). Thus, with this system, Cas9 nuclease activity can be directed to any DNA sequence of the form N20-NGG simply by altering the first 20 nts of the gRNA to correspond to the target DNA sequence. It is notes that Type II CRISPR systems from other species of bacteria recognize alternative PAM sequences and that utilize different crRNA and tracrRNA sequences could also beused to perform targeted genome editing.

The Cas9-induced DSBs have been used to introduce NHEJ-mediated indel mutations as well as to stimulate HDR with both double-stranded plasmid DNA and single-stranded oligonucleotide donor templates. The capability to introduce DSBs at multiple sites in parallel using the Cas9 system is a unique advantage of this platform relative to ZFNs, or TALENs. For example, expression of Cas9 and multiple gRNAs has been used to induce small and large deletions or inversions between the DSBs, to simultaneously introduce parallel genetic editing mutations altering different genes in rats, mouse ES cell clones, and zebrafish. Together, these advances in CRISPR/cas-mediated gene editing technology can accelerated

the pace of gene-function relationship discovery, and a focused approach for developing personalized therapeutics.

An alternative is the use of Cpf1 a putative class 2 CRISPR effector with features distinct from Cas9. Cpf1 is a single RNA-guided endonuclease that does not require use of tracrRNA. Instead, it utilizes a T-rich protospacer-adjacent motif, TTTN, and on the 5' side of the guide. As a result, the cut CPF1 makes is staggered, occurring 19 bp after the PAM on the targeted + strand, and 23 bp on the opposite strand. Cpf1 generates a staggered cut with a 5' overhang, in contrast to the blunt ends generated by Cas9. This allows for editing through non-homologous end joining (NHEJ). Being able to program the exact sequence of a sticky end would allow researchers to design the DNA insert so that it integrates in the proper orientation. AsCpf1 (from *Acidaminococcus*) and LbCpf1 (from *Lachnospiraceae*) have been demonstrated as possessing genomic editing capacity in human cells.

The elegance of the CRISPR/Cas or CRISPR/Cpf1 system is allowing for tailoring to target the patient's particular mutation, combined with a delivery system via adeno-associated virus ("AAV"), or also via adenovirus, vectors as optimal vehicles for genome editing machinery can deliver components directly to the organ or cells of interest. There have been limited reports that, for example, systemic injection of an AAV vector carrying a zinc-finger nuclease and donor template construct was able to correct mutant transgenic clotting Factor IX in mice and reconstitute low but clinically detectable levels of circulating protein. In this regard, an AAV-CRISPR system could be delivered to treat dominant mutations via the same gene correction mechanism used for recessive mutations, compactly deliver targeted genomic editing machinery with a limited footprint capable of being delivered via viral vectors, as constant and agnostic to the size of the target gene and maintain the endogenous gene expression stoichiometry. These and other advantages of CRISPR/Cas, CRISPR/Cpf1 editing give it a wide range of possible clinical applications.

Thus, for any genome editing system with engineered nucleases, the use of an endonuclease joined to a DNA binding moiety such as zinc finger protein that binds donor DNA, or other methods to associate a donor DNA. The proximity of donor DNA and the double stand break leads to increased integration efficiency of the donor DNA.

Despite these advances, a paramount technical problem remains the low integration efficiency of donor DNA using genome editing tools. Herein, the Inventors describe a solution based on the realization that it is possible to bring the donor DNA in close proximity

to the site of the double stranded break in the genomic DNA. This can be accomplished through creation of a Cas9:DNA-binding protein motif fusion. By recruiting the donor DNA to the cleavage site, it is possible to use a donor DNA concentration that is non-toxic to cells and yet providing a local donor DNA concentration that should improve integration efficiency. Because the donor DNA is held in such close proximity to the damaged DNA, one can engineer the sequence homology arms on both ends of the donor DNA to be relatively short (<100bp) such that the cell's SOS DNA repair system favors the: 5'-3' exonuclease V trimming, RPA coating of the single strand DNA, 3' strand invasion, 3' end trimming by exonucleases Rad1/Rad10, and DNA nick ligation to repair the DNA versus pure Holliday junction dependent homologous recombination. Exploiting this type of DNA repair, if a favored result, will have a robust genome editing system which can be further optimized for integration of very large donor DNA fragments. By shunting stochastic processes towards favorable recombination events, recombinant fusion proteins aid in increasing proximity of DSB and linear DNA for more efficient genome editing.

In order for this increased integration efficiency to take place, the donor DNA must be in a position that is as close, or closer, to the site of DNA damage than the two opposing ends of the damaged genomic DNA are to each other. Recruitment of the donor DNA to these sites is performed either by direct or indirect association of the donor DNA with the enzymatically active Cas-9 RNA-dependent endonuclease. These DNA binding elements entail specific domains or full-length proteins in their entirety, including but not limited to these naturally occurring or engineered examples: transcription factors, endonucleases, zinc fingers, TALENs, endonuclease-minus Cas-9+guide strand RNA, or other such ribonucleoprotein that can bind directly or indirectly to specific DNA sequences. Some, but certainly not all, of the possible configurations of Cas9-endonuclease and donor DNA binding elements are: direct fusion, association (multimerization domains like leucine zippers, fkbp/FRB, etc.), engineered association via antibody mimetics, or any synthetic macromolecule (carbohydrate-, protein-, or lipid-based) which bind to Cas9 endonuclease and also bind to the donor DNA in a sequence specific manner. Preliminary results suggest this modified Cas-9 works much better than wild-type Cas-9 in generating knock-in gene modifications in cell lines.

Described herein is a composition including a vector encoding a fusion protein including at least one endonuclease and a DNA binding moiety. In various embodiments, the fusion protein endonuclease includes at least one endonuclease selected from the group consisting of: cas regularly interspaced short palindromic (CRISPR) protein, a zinc finger nuclease (ZFNs) or transcription activator-like effector nucleases (TALENs). In other embodiments, the DNA binding moiety includes a zinc finger protein. In other embodiments, the fusion protein includes at least one endonuclease CRISPR protein and a DNA binding moiety zinc finger protein. In other embodiments, the zinc finger protein includes a left handed, right handed, or both zinc fingers. In other embodiments, the zinc finger includes a left handed CCR5 sequence. In other embodiments, DNA binding moieties can include specific domains or full-length proteins in their entirety, including transcription factors, endonucleases, zinc fingers, TALENs, endonuclease-minus Cas-9+guide strand RNA, or other such ribonucleoprotein that can bind directly or indirectly to specific DNA sequences. In other embodiments, the at least one endonuclease and DNA binding moiety are joined by a linker including two, three, four, five, six, seven, eight, nine, ten or more amino acids. In other embodiments, some configurations of Cas9-endonuclease and donor DNA binding moieties are: direct fusion, association (multimerization domains like leucine zippers, fkbp/FRB, etc.), engineered association via antibody mimetics, or any synthetic macromolecule (carbohydrate-, protein-, or lipid-based) which bind to Cas9 endonuclease and also bind to the donor DNA in a sequence specific manner. In other embodiments, the fusion protein further includes a nuclear localization signal (NLS), such as SV40 NLS. In other embodiments, the CRISPR protein is a *Streptococcus pyogenes*-derived cas protein. In other embodiments, the CRISPR protein is not a *Streptococcus pyogenes*-derived cas protein. In various embodiments, CRISPR protein is cpf1, such as AsCpf1 from *Acidaminococcus* and LbCpf1 is from *Lachnospiraceae*. In other embodiments, the CRISPR protein is cas9. In other embodiments, the CRISPR protein is cpf1. In other embodiments, the fusion protein includes a reporter protein. In various embodiments, the report protein includes a fluorescent labeled protein including green or red fluorescent protein (GFP or RFP, including enhanced eGFP), mCherry, or similar proteins. In other embodiments, the vector is a DNA vector, plasmid, artificial chromosome. In other embodiments, the vector is a virus, such as adenovirus, adeno associated virus, or lentivirus.

In other embodiments, the vector encodes one or more guide RNAs (gRNAs), wherein the one or more gRNAs include a sequence capable of binding to a protospacer

adjacent motif (PAM). In other embodiments, the one or more gRNAs include a sequence capable of binding to a PAM. In other embodiments, the PAM includes the sequence NGG. In other embodiments, the PAM includes the sequence NAG. In other embodiments, the gRNA comprise a CRISPR-derived RNAs (crRNA) and trans-acting antisense RNA (tracRNA). In various embodiments, the gRNA is 10, 20, 30, or 40 or more nucleotides in length. In various embodiments, about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more nucleotides are cognate to a gene of interest. In various embodiments, about 20 nucleotides are cognate to a genetic loci of interest. For example, this includes gRNA designs that hybridize to a target sequence with N_{20}NGG. In some embodiments, the CRISPR protein is cas9. In other embodiments, the CRISPR protein is cpf1. In various embodiments, the composition is used in a method for altering a target polynucleotide sequence in a cell including contacting the polynucleotide sequence with a CRISPR protein (e.g., cas9) with at least one gRNA directing CRISPR to hybridize to a cognate sequence on a target polynucleotide sequence, wherein the target polynucleotide sequence is cleaved, and wherein the efficiency of alteration of cells that express CRISPR protein is from about 10-20%, 30-40%, 40-50%, or 50-80% or more. In various embodiments, the efficiency of alteration is improved 1x, 2, 3x, 4x, 5x, 6x, 7x, 8x, 9x, 10x, 20x, 30x, 40x, 50x, 100x when compared to a method using a native wild-type endonuclease. Further described herein is a quantity of cells produced using the described method.

Further described is a method of genomic editing including providing a quantity of one or more vectors each encoding a fusion protein including at least one endonuclease and a DNA binding moiety and contacting a population of cells with the quantity of the one or more vectors. In various embodiments, the fusion protein endonuclease includes at least one endonuclease selected from the group consisting of: cas regularly interspaced short palindromic (CRISPR) protein, a zinc finger nuclease (ZFNs) or transcription activator-like effector nucleases (TALENs). In other embodiments, the DNA binding moiety includes a zinc finger protein. In other embodiments, the fusion protein includes at least one endonuclease CRISPR protein and a DNA binding moiety zinc finger protein. In other embodiments, the zinc finger protein includes a left handed, right handed, or both zinc fingers. In other embodiments, the zinc finger includes a left handed CCR5 sequence. In other embodiments, DNA binding moieties can include specific domains or full-length proteins in their entirety, including transcription factors, endonucleases, zinc fingers,

TALENs, endonuclease-minus Cas-9+guide strand RNA, or other such ribonucleoprotein that can bind directly or indirectly to specific DNA sequences. In other embodiments, the at least one endonuclease and DNA binding moiety are joined by a linker including two, three, four, five, six, seven, eight, nine, ten or more amino acids. In other embodiments, Some, configurations of Cas9-endonuclease and donor DNA binding moieties are: direct fusion, association (multimerization domains like leucine zippers, fkbp/FRB, etc.), engineered association via antibody mimetics, or any synthetic macromolecule (carbohydrate-, protein-, or lipid-based) which bind to Cas9 endonuclease and also bind to the donor DNA in a sequence specific manner. In other embodiments, the fusion protein further includes a nuclear localization signal (NLS), such as SV40 NLS. In other embodiments, the CRISPR protein is a *Streptococcus pyogenes*-derived cas protein. In other embodiments, the CRISPR protein is not a *Streptococcus pyogenes*-derived cas protein. In various embodiments, CRISPR protein cpf1, such as AsCpf1 from *Acidaminococcus* and LbCpf1 is from *Lachnospiraceae*. In other embodiments, the CRISPR protein is cas9. In other embodiments, the CRISPR protein is cpf1. In other embodiments, the fusion protein includes a reporter protein. In various embodiments, the report protein includes a fluorescent labeled protein including green or red fluorescent protein (GFP or RFP, including enhanced eGFP), mCherry, or similar proteins. In other embodiments, the method is an *in vivo* method. In other embodiments, the method is an *in vitro* method. In certain embodiments, the population of cells include embryonic stem cells, including human or mouse embryonic stem cells. In various embodiments, the method includes generation of a double stranded break (DSB) in the quantity of cells, wherein homologous recombination (HR) of the DSB results in editing of the genome of the cells. In other embodiments, HR includes non-homologous end joining (NHEJ) introducing missense or nonsense of a protein expressed at the locus. In other embodiments, the missense or nonsense results in a knockout of a target sequence in the genome. In other embodiments, HR includes homology directed repair (HDR) introduces template DNA. In other embodiments, the HDR results in a knock-in of a target sequence in the genome. In other embodiments, the template DNA is cognate to a target sequence. In other embodiments, the template DNA is cognate to a wild-type genetic sequence. In other embodiments, the template DNA contains an expression cassette, for example, including a sequence transcribed and translated into a protein of interest. In other embodiments, the template DNA includes at least 80 bases of exact sequence homology both upstream and downstream of about 20 bases cognate to a target sequence, or cognate to a wild-type genetic

sequence. In other embodiments, the upstream and downstream sequences are at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 200, 250, 300, 350, 400, 450, 500 or more base pairs. In other embodiments, the template DNA includes a sequence for binding of a DNA binding moiety. In other embodiments, the template DNA includes about 12 base pairs constituting the left-handed CCR5-zinc finger-binding site. In other embodiments, contacting a population of cells with the quantity of the one or more vectors include transfection, electroporation, and/or lipofection. In other embodiments, the vector is a DNA vector, plasmid, artificial chromosome. In other embodiments, the vector is a virus, such as adenovirus, adeno associated virus, or lentivirus.

In other embodiments, the vector encodes one or more guide RNAs (gRNAs), wherein the one or more gRNAs include a sequence capable of binding to a protospacer adjacent motif (PAM). In other embodiments, one or more exogenous gRNAs are introduced to the quantity of cells. In other embodiments, the one or more gRNAs include a sequence capable of binding to a PAM. In other embodiments, the PAM includes the sequence NGG. In other embodiments, the PAM includes the sequence NAG. In other embodiments, the gRNA comprise a CRISPR-derived RNAs (crRNA) and trans-acting antisense RNA (tracRNA). In various embodiments, the gRNA is 10, 20, 30, or 40 or more nucleotides in length. In various embodiments, about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more nucleotides are cognate to a gene of interest. In various embodiments, about 20 nucleotides are cognate to a genetic loci of interest. For example, this includes gRNA designs that hybridize to a target sequence with N_{20}NGG. In some embodiments, the CRISPR protein is cas9. In various embodiments, the composition is used in a method for altering a target polynucleotide sequence in a cell including contacting the polynucleotide sequence with a CRISPR protein (e.g., cas9) with at least one gRNA directing CRISPR to hybridize to a cognate sequence on a target polynucleotide sequence, wherein the target polynucleotide sequence is cleaved, and wherein the efficiency of alteration of cells that express CRISPR protein is from about 10-20%, 30-40%, 40-50%, or 50-80% or more. In various embodiments, the efficiency of alteration is improved 1x, 2, 3x, 4x, 5x, 6x, 7x, 8x, 9x, 10x, 20x, 30x, 40x, 50x, 100x when compared to a method using a native wild-type endonuclease. Further described herein is a quantity of cells produced using the described method.

For example, the method of genomic editing including providing a quantity of one or more vectors each encoding a fusion protein including at least one endonuclease and a DNA binding moiety and contacting a population of cells with the quantity of the one or more

vectors, wherein the at least one endonuclease includes a CRISPR protein and a DNA binding moiety includes zinc finger protein, contacting a population of cells with the quantity of the one or more vectors includes transfection, electroporation, and/or lipofection, further including gRNA and/or template DNA, and after contacting a population of cells, results in generation of a double stranded break (DSB) and homology directed repair (HDR) introduces the template DNA in the genome of the population of cells. In other embodiments, the template DNA is cognate to a target sequence. In other embodiments, the template DNA is cognate to a wild-type genetic sequence. In other embodiments, the template DNA contains an expression cassette, for example, including a sequence transcribed and translated into a protein of interest. In other embodiments, the template DNA includes at least 80 bases of exact sequence homology both upstream and downstream of about 20 bases cognate to a target sequence, or cognate to a wild-type genetic sequence. In other embodiments, the template DNA includes a sequence for binding of a DNA binding moiety. In other embodiments, the template DNA includes about 12 base pairs constituting the left-handed CCR5-zinc finger-binding site.

Also described is a kit for genomic editing including a quantity of one or more vectors each encoding a fusion protein including at least one endonuclease and a DNA binding moiety and contacting a population of cells with the quantity of the one or more vectors, and further including a template DNA cognate to a target sequence or a wild-type genetic sequence, the template DNA including at least 10 bases of exact sequence homology both upstream and downstream of about 20 bases cognate to a target sequence, or cognate to a wild-type genetic sequence. In other embodiments, the DNA binding moiety includes a zinc finger protein. In other embodiments, the fusion protein includes at least one endonuclease CRISPR protein and a DNA binding moiety zinc finger protein. In other embodiments, the kit includes one or more guide RNAs (gRNAs), wherein the one or more gRNAs include a sequence capable of binding to a protospacer adjacent motif (PAM).

In other embodiments, the upstream and downstream sequences are at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 200, 250, 300, 350, 400, 450, 500 or more base pairs. In various embodiments, the fusion protein endonuclease includes at least one endonuclease selected from the group consisting of: cas regularly interspaced short palindromic (CRISPR) protein, a zinc finger nuclease (ZFNs) or transcription activator-like effector nucleases (TALENs). In other embodiments, the DNA binding moiety includes a

zinc finger protein. In other embodiments, the fusion protein includes at least one endonuclease CRISPR protein and a DNA binding moiety zinc finger protein. In other embodiments, the zinc finger protein includes a left handed, right handed, or both zinc fingers. In other embodiments, the zinc finger includes a left handed CCR5 sequence. In other embodiments, DNA binding moieties can include specific domains or full-length proteins in their entirety, including transcription factors, endonucleases, zinc fingers, TALENs, endonuclease-minus Cas-9+guide strand RNA, or other such ribonucleoprotein that can bind directly or indirectly to specific DNA sequences. In other embodiments, the at least one endonuclease and DNA binding moiety are joined by a linker including two, three, four, five, six, seven, eight, nine, ten or more amino acids. In other embodiments, some, configurations of Cas9-endonuclease and donor DNA binding moieties are: direct fusion, association (multimerization domains like leucine zippers, fkbp/FRB, etc.), engineered association via antibody mimetics, or any synthetic macromolecule (carbohydrate-, protein-, or lipid-based) which bind to Cas9 endonuclease and also bind to the donor DNA in a sequence specific manner. In other embodiments, the fusion protein further includes a nuclear localization signal (NLS), such as SV40 NLS. In other embodiments, the CRISPR protein is a *Streptococcus pyogenes*-derived cas protein. In other embodiments, the CRISPR protein is not a *Streptococcus pyogenes*-derived cas protein. In various embodiments, CRISPR protein is cpf1, such as AsCpf1 from *Acidaminococcus* and LbCpf1 is from *Lachnospiraceae*. In other embodiments, the CRISPR protein is cas9. In other embodiments, the fusion protein includes a reporter protein. In various embodiments, the report protein includes a fluorescent labeled protein including green or red fluorescent protein (GFP or RFP, including enhanced eGFP), mCherry, or similar proteins. In various embodiments, the kit is capable of generating a double stranded break (DSB) in the quantity of cells, wherein homologous recombination (HR) of the DSB results in editing of the genome of the cells. In other embodiments, HR includes non-homologous end joining (NHEJ) introducing missense or nonsense of a protein expressed at the locus. In other embodiments, the missense or nonsense results in a knockout of a target sequence in the genome. In other embodiments, HR includes homology directed repair (HDR) introduces template DNA. In other embodiments, the HDR results in a knock-in of a target sequence in the genome. In other embodiments, the template DNA is cognate to a target sequence. In other embodiments, the template DNA is cognate to a wild-type genetic sequence. In other embodiments, the template DNA contains an expression cassette, for example, including a

sequence transcribed and translated into a protein of interest. In other embodiments, the template DNA includes at least 80 bases of exact sequence homology both upstream and downstream of about 20 bases cognate to a target sequence, or cognate to a wild-type genetic sequence. In other embodiments, the upstream and downstream sequences are at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 200, 250, 300, 350, 400, 450, 500 or more base pairs. In other embodiments, the template DNA includes a sequence for binding of a DNA binding moiety. In other embodiments, the template DNA includes about 12 base pairs constituting the left-handed CCR5-zinc finger-binding site. In other embodiments, contacting a population of cells with the quantity of the one or more vectors include transfection, electroporation, and/or lipofection. In other embodiments, the vector is a DNA vector, plasmid, artificial chromosome. In other embodiments, the vector is a virus, such as adenovirus, adeno associated virus, or lentivirus.

In other embodiments, the vector encodes one or more gRNAs, wherein the one or more gRNAs include a sequence capable of binding to a PAM. In other embodiments, one or more exogenous gRNAs are introduced to the quantity of cells. In other embodiments, the one or more gRNAs include a sequence capable of binding to a PAM. In other embodiments, the PAM includes the sequence NGG. In other embodiments, the PAM includes the sequence NAG. In other embodiments, the gRNA comprise a CRISPR-derived RNAs (crRNA) and trans-acting antisense RNA (tracRNA). In various embodiments, the gRNA is 10, 20, 30, or 40 or more nucleotides in length. In various embodiments, about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more nucleotides are cognate to a gene of interest. In various embodiments, about 20 nucleotides are cognate to a genetic loci of interest. For example, this includes gRNA designs that hybridize to a target sequence with N_{20}NGG. In some embodiments, the CRISPR protein is cas9. In other embodiments, the CRISPR protein is cpf1.In various embodiments, the kit is used in a method for altering a target polynucleotide sequence in a cell including contacting the polynucleotide sequence with a CRISPR protein (e.g., cas9) with at least one gRNA directing CRISPR to hybridize to a cognate sequence on a target polynucleotide sequence, wherein the target polynucleotide sequence is cleaved, and wherein the efficiency of alteration of cells that express CRISPR protein is from about 10-20%, 30-40%, 40-50%, or 50-80% or more. In various embodiments, the efficiency of alteration is improved 1x, 2, 3x, 4x, 5x, 6x, 7x, 8x, 9x, 10x, 20x, 30x, 40x, 50x, 100x when compared to a method using a native wild-type endonuclease. Further described herein is a quantity of cells produced using the described method.

Also described herein is a method of multi-locus genomic editing including inducing sticky end formation at one or more loci by adding a CRISPR protein, providing a quantity of one or more guide strand RNAs, ligating one or more single-stranded donor DNA, hybridizing one or more double-stranded DNA with a terminating oligonucleotide, synthesis of one or more double stranded DNA from the one or more single-stranded donor DNA to the one or more double-stranded DNA completing the donor DNA strand to form a sticky end, and joining compatible sticky ends at one or more loci. In various embodiments, providing a quantity of one or more guide strand RNAs includes T7 RNA synthesis of guide-strand RNA. In other embodiments, ligating one or more single-stranded donor DNA includes splint mediated ligation of single-stranded donor DNA sequence onto the 5' end of the guide-strand RNA. In various embodiments, the 5' nucleotide of the synthesized donor DNA has an exonuclease resistant phosphorothioate bond with its neighboring nucleotide. In other embodiments, hybridization of double-stranded DNA terminating oligonucleotide occurs. In other embodiments, synthesis of one or more double stranded DNA from the one or more single-stranded donor DNA to the one or more double-stranded DNA completing the donor DNA strand includes isothermal DNA polymerase (Klenow exo-) reaction. In other embodiments, synthesis of one or more double stranded DNA from the one or more single-stranded donor DNA to the one or more double-stranded DNA completing the donor DNA strand to form a sticky end is by T4-DNA ligase reaction. In various embodiments, complete integration happens upon ligation of the sticky end of the donor DNA and homologous recombination between donor and genomic DNA due to their proximity and sequence homology. In various embodiments, the PAM is TTTN. In various embodiments, the cpf1 sticky end is 19 bp after the PAM on the targeted + strand, and 23 bp on the opposite strand with a 5' overhang. In various embodiments, cpf1 such as AsCpf1 from *Acidaminococcus* and LbCpf1 is from *Lachnospiraceae*.

Example 1

Experimental design

As described, gene manipulation by Cas-9 includes genetic knockouts through Cas-9 DNA cleavage and emergency DNA repair systems are relatively easy to produce, yet gene knock-ins/fusions are much more challenging because: of larger DNA inserts (GFP=717 base pairs) integrate with low efficiency, dependency of intregration rate rate is dependent on insert

DNA concentration, higher concentrations of linear DNA is toxic to cells

Through actively recruiting the linear DNA insert in closer proximity to the genomic cleavage site, one can increase integration efficiency of large DNA fragments into the genome, be able to use lower linear DNA concentrations without sacrificing efficiency, and quickly screen through various protein configurations due to cell culture system. A nice biological (fluorescent) readout allows for quick optimization.

To aid development of a quick and easy analysis of CRISPR/Cas9 optimization, a GFP:B-actin reporter system is developed including Step 1) Building Cas9:CCR5:NLSfusion protein; Step 2) Synthesize sgRNA against human genomic B-actin ; Step 3) Design suitable linear GFP DNA sequence optimized for efficient DNA repair, Step 4) Screen for green cells and optimize if needed.

Example 2

Study model

GFP-βactin is our first genomic GFP-tagged goal protein for the following reasons: it expresses constantly and one can screen cells quickly, any frame shift and/or cleavage leading to knocking out actin results in cell death. The two fluorescent labels, mCherry and GFP, are the means of screening integration efficiency and cell viability. GFP can only express if it is integrated into the genome and is in-frame with surrounding sequence. The mCherry should express in all living transfected cells.

The experimental conditions included: 1) mCherry only (tranfection efficiency/cell viability control); 2) Cas-9-nls, gsRNA, donor DNA, mCherry (NHEJ efficiency); 3)Donor DNA, gsRNA, mCherry (cell viability due to linear DNA); 4) Cas-9-ZF, donor DNA, mCherry (off-target CCR5 effects); 5) Cas-9-ZF, gsRNA, donor DNA, mCherry (test).

Fluorescently labeled GFP signal allows rapid determination of cell viability due to the constituent reagents in the transfection in order to optimize future electroporations. GFP signal will only come from cells positive for integration and colonies can be isolated, collected, expanded, genotyped to determine GFP insertion sites, and imaged to determine proper protein localization.

Example 3

Methods

A vector was purchased from Addgene encoding Cas-9-NLS (Plasmid #42251). The C-terminal NLS was removed from the 3' end of the DNA sequence and replaced with the coding sequence for a short 3x glycine-serine linker (6 amino acids total), the left-handed CCR5, and the NLS previously removed was added back. The vector was linearized and used as template for a T7 mMessage® capped mRNA synthesis kit. The RNA is cleaned up using an RNeasy kit followed by phenol/chloroform extraction and ethanol precipitation.

Guide-strand RNA is synthesized using the following protocol. A GFP-gene fusion (GFP-βactin) was published as a functionally viable protein and therefore became the desired genomic GFP-tagged endogenous protein. PCR primers were designed to amplify sequence surrounding this region for targeted insertion within the HEK293 cell line and the PCR product was sequenced to confirm accuracy of the published chromosomal sequence data. 20 bases of sequence within the genome are chosen as the hybridization target for the guide-strand RNA and run through several algorithms on-line to both maximize guide-strand hybridization to the genome and to minimize off-target hybridizations. After a sequence is selected, a 120bp DNA template is assembled by Klenow fill-in and PCR amplification from 4 ssDNA oligo nucleotides ordered from IDT, only one of which is unique to each guide-strand (i.e. only one oligo, $18, must be purchased for a unique gsRNA). The template consists of a T7 promoter, 20 bases of unique homology to the genomic target, and 80 bases that will encode the Cas9-binding hairpins. The recipe is as follows: 5pmoles DNA template, 125uM NTP mix, transcription buffer (HEPES-$MgCl_2$, DTT, spermidine, pH 7.5), 150ug T7 RNA pol (20ul, made in house) in a total volume of 200ul. The reaction is incubated at 37°C for 2-4hours until a white precipitate of Mg^{2+}(pyrophosphate) collects at the bottom of the tube. 25ul of 0.5M EDTA is added to clear the precipitate and halt further polymerization. The newly synthesized RNA is run out on an 8% acrylamide Urea-PAGE gel to separate unincorporated NTPs, truncated RNAs, and DNA template from the full-length 100base gsRNA. The RNA is visualized via shadow imaging over a TLC plate, the bands are cut out from the gel and eluted from the gel by electrophoresis on a Whatman elutrap. The RNA is then ethanol precipitated and dissolved in 10mM Tris pH 8.0 to the desired concentration.

After some test integrations, it was determined that the donor DNA sequence should consist of at least 80 bases of exact sequence homology both upstream and downstream of the

20 bases required by Cas-9 to cause the double stranded DNA break. These 80bp up- and 80bp downstream are engineered on the 5' and 3' ends of the donor DNA sequence that are to be added into the genome. It is significant to note that the sequencing data used to determine the optimal gsRNA hybridization sequence is also useful to generate the precise homology arms. The 12 base pairs constituting the left-handed CCR5-zinc finger-binding site are added strategically as to not interfere with either the final coding sequence or splicing of the mRNA transcribed from the target locus. Super-folder GFP (sfGFP), mCerulean, and tagRFP were codon optimized, through silent mutations, such that any consensus mRNA splice donor and splice acceptor sequences were removed from both sense and antisense strands. The βactin homology domains and CCR5-binding site were extended in both directions from sfGFP in PCR reactions. The linear DNA was purified from an LMP agarose gel, extracted using a Qiagen Gel Extraction Kit, phenol/chloroform extracted, ethanol precipitated, and dissolved in TE buffer (10mM Tris, 1mM EDTA, pH 8.0) to the desired concentration.

Because of the mixed nature of the various components (100base gsRNA, 4500base Cas-9-ZF mRNA, 6500bp Ubi6::mCherry plasmid as transfection control, 890bp linear donor DNA) going into the transfection, it was determined that electroporation was the delivery method of choice as opposed to lipophilic or receptor mediated transfection reagents. $1x10^6$ cells are suspended in 1xPBS containing: 2ug gsRNA, 1.5ug Cas-9-ZF mRNA, 50ng donor DNA, and 1ug plasmid DNA in 40ul total volume. The cells are electroporated in a square wave, 42V, 50ms, pulse supplied by Harvard Apparatus BTX-840. The cells are then plated on 60mm glass bottom cell culture dishes for ease of imaging, analysis, and clonal harvesting. HEK293 cells are grown in DMEM with 5%FBS and gentamicin. The CRISPR/Cas mRNA or protein, gsRNA, and donor DNA mixture is injected into zebrafish embryos at 10-30 minutes post fertilization to make our zebrafish transgenic lines.

Example 4

Preliminary results

As shown in Figure 2, a wild-type condition resulted in only 5 out of 18 CFP positive colonies that were also GFP positive, 2 of which were only GFP positive (i.e., stable integrants). By contrast, using the describe Cas-9 zinc finger suste, 9 out of 12 CFP positive colonies that were also GFP positive, 4 of which were only GFP positive. In Figure 3, results demonstrate that addition of the Zinc-finger to Cas9 does not confer additional zinc-finger

dependent endonuclease activity (Figure 3A) and three examples of GFP signal due to integration of donor DNA are shown (Figure 3B). The GFP positive clones are being validated (Figure 4). mCherry is the internal viability control. This work was done in HEK-293t cells to generate GFP fused to endogenous beta actin.

This proof-of concept of labeling endogenous proteins in stem cells can readily be translated to generate transgenic animals quickly. The ability to recruit donor DNA, via the Cas-9:CCR5 zinc finger, to the break site of Cas-9 increases the integration of foreign DNA (723 base pairs, GFP) 2.7 fold over wildtype Cas-9 in this first iteration. By adding more genomic sequence on either side of the donor DNA one should be able to get higher integration efficiencies.

Example 5

Subsequent projects and applications

Once proper conditions for integration are worked out, one can genetically tag Oct4 with tagRFP, Sox2 with CFP, and Nanog with sfGFP in separate mouse ES cell lines, and combine them in all possible permutations. These cell lines will be useful in their own right but are useful to generate transgenic mouse lines where we are able to visualize absolute amounts and localizations of these developmentally important proteins expressed as endogenous levels of fusion proteins.

Additionally, one can skip *in vivo* translation of Cas-9 mutant by bacterial expression and subsequently purify Cas-9 fusion protein. This purified protein would be premixed and equilibrated with both guide strand RNA and donor DNA necessary for all unique genomic integrations. These preassembled endonuclease/donor DNA integration units could be injected or transfected into embryos or any cell line. This configuration allows not only genomic editing at the earliest stages but also would allow multiple integration events simultaneously with high efficiency. Such would be the first successful generation of a mouse ES stem cell line with all three of the previously described integrations, simultaneously, as the first multiple knock-in lines. In other examples, one could also target RNA probes such as Spinach (1 or 2) into the genome to get visual read outs of nascent transcription and/or mRNA splicing events.

Example 6

Construct formats

Additionally, one could potentially combine all of the elements of this CRISPR, Cas-9:fusion, donor DNA-targeting system into a single virus. The attenuated virus would express Cas-9:fusion, and guide strand RNA under an orthogonally induced promoter (i.e. ecdysone receptor-driven expression would work in all species except insect) or cell type-specific promoter, and the viral genome would contain the donor DNA sequence flanked on both sides by the 20 base pairs recognized by the guide strand RNA. This configuration would generate linear donor DNA with the requisite flanking homology sequence to the genomic target and would be recruited by Cas-9:fusion-gsRNA endonuclease to the target locus. This viral construct could be used to: correct genetic mutations, make drugable targets more sensitive to drugs, or make previously non-drugable targets drugable, and would self-destruct in the process.

Example 7

Exemplary sequences used in applications

A variety of constructs were prepared to test the variable designs described herein. In on example, a Cas-9 zinc-finger fusion protein was generated, including the following elements: the Cas-9 endonuclease, a short 9 amino acid linker (ssagagaga), left-handed CCR5 zinc-finger, 4 amino acids (wrlp), and a nuclear localization sequence, stop. Nucleotide sequence is described in [SEQ ID NO: 1] and amino acid sequence of the construct is described in [SEQ ID NO: 2]. For Ni-NTA purification of the Cas-9 fusion protein, the Inventors added a 6x histidine tag to the extreme carboxy-terminal end of the protein.

Other donor DNA sequence examples used or synthesized include a super-folder GFP: Beta-actin [SEQ ID NO: 3] utilized in the human cancer cell line HEK293. This sequence includes 80 base pairs of sequence homology upstream of the Cas-9 endonuclease cleavage site, left-handed CCR5 binding sequence, 12 bases encoding a short linker, super-folder GFP, and 80 base pairs of sequence homology downstream of the endonuclease cleavage site.

Another example includes a Sox-2:PS-mOrange2:Spinach2 [SEQ ID NO: 4] for use in zebrafish (D.rerio). 81 base pairs of sequence homology upstream of the Cas-9 endonuclease cleavage site, silent mutations to the Sox-2 carboxy-terminal end, 48 bases

coding for 16 glycine/serine residues, photoswitchable orange fluorescent protein (PS-mOrange2), 144 base pairs non-coding, left-handed CCR5 binding sequence, Spinach-2 (mRNA fluorescent reporter), and 82 base pairs of sequence homology downstream of the endonuclease cleavage site.

Additional example includes Oct-4:GFP [SEQ ID NO: 5] for use in mouse (M.musculus). This sequence includes 82 base pairs of sequence homology upstream of the Cas-9 endonuclease cleavage site, 27 bases encoding a glycine/alanine linker, super-folder GFP, left-handed CCR5 binding sequence, and 94 base pairs of sequence homology downstream of the endonuclease cleavage site.

Finally, a variety of guide-strand RNA sequences were utilized including human beta-Actin [SEQ ID NO: 6], zebrafish Sox-2 [SEQ ID NO: 7], and mouse Oct-4 [SEQ ID NO: 8].

Example 8

Experimental protocol

An exemplary experimental protocols for applying the above constructs is described further herein:

- All cas9 protein/sgRNA/donor DNA complexes are allowed to equilibrate at room temp for 30 minutes.
- Human B-actin sgRNA and donor DNA can used
- A "1:2:3" mixture strategy is utilized (e.g. 1 pmol plasmid: 2pmol sgRNA: 3pmol protein: 3pmol of sgRNA) with 90ng(0.013pmol) plasmid DNA. These conditions are optimized for minimal protein usage and for clear read-out on the gel.
- Cas9 reactions are carried out at 28 degrees ^{o}C for two hours.
- XmaI digest (single cut) is utilized, two hour reaction at 37 degrees C.
- Samples were cleaned up with qiagen PCR clean-up reagents/columns prior to gel electrophoresis.
- Samples were run on a 1% agarose gel in TAE.

The above protocol demonstrates that Cas9-zinc finger fusion protein linearizes plasmid DNA efficiently (Fig. 5). It appears that some off-target cleavage may occur due to the presence of the zinc finger. However, off-target cleavage is ameliorated when donor DNA is included. The above approach has allowed for optimized conditions for

injections into zebrafish, as well as *in vitro* confirmation that the B-actin guide strand and cas-9 fusion are functional together. Further confirmation is available based on running flow cytometry cell sorting (FACS) on frozen GFP:B-actin HEK cells. Once a GFP sorted culture is prepared, genomic analysis can confirm editing.

Example 9

All-in-one CRISPR editing

In an alternative embodiment, mixtures of specific guides strand RNA-donor DNA hybrids could allow for parallel, multi-locus mutations. More specifically, mixtures of specific guide strand RNA-donor DNA hybrids and a single CRISPR protein preparation are deployed. This approach would allow a researcher to make simultaneous additions/mutations to multiple loci in the genome of any organism or cell line.

By using Cpf1 instead of Cas9 for CRISPR genome editing, this allows one to take advantage of highly efficient sticky-end donor DNA-genomic DNA ligation repairs. A partial unnatural nucleotide backbone in the donor DNA assembly primers make the sticky-end of the donor DNA less vulnerable to degradation, which preservation theoretically will allow for much higher ligation efficiencies.

Parallel, multi-locus, genetic mutations for developing disease or research models for any organism can be made and allow the simultaneous editing of any number of target genes just by adding additional guide-strand RNA/donor DNA fusions.

While the above is described using the CRISPR/cpf1 system, as efficient for these purposes, the above approach appears compatiable with a variety of current or future RNA-dependent endonucleases such as cas9 among others.

The guide strand/donor fusion is assembled as shown in Figure 6. More specifically,

1) T7 RNA synthesis of guide-strand RNA (Fig. 6A)

2) Splint (Fig. 6B) mediated ligation (Fig. 6C) of single-stranded donor DNA sequence (Fig. 6D) onto the 5' end of the guide-strand RNA. The 5' nucleotide of the synthesized donor DNA (Fig. 6E) has an exonuclease resistant phosphorothioate bond with its neighboring nucleotide.

3) Hybridization of double-stranded DNA terminating oligonucleotide (Fig. 6F).

4) Isothermal DNA polymerase (Klenow exo-) reaction to fill-in the double strandedness (Fig. 6G) from the splinting primer (step 2) to the terminating oligo (step

3) and T4-DNA ligase reaction (Fig. 6H) to complete the donor fragment with the appropriate 5' sticky end (Fig. 6I) for donor DNA sticky end ligation to the sticky end of the genomic DNA (Fig. 6J) digested by cpf1.

5) Complete integration happens upon ligation of the sticky end (Fig. 6K) of the donor DNA and homologous recombination between donor and genomic DNA (Fig.6) due to their proximity and sequence homology.

Given the fact that each sgRNA/donor is unique and confers both digestion specificity and the specific donor DNA mutation for the targeted gene, one can combine these hybrid nucleotides with a single protein preparation and get multiple targeted mutations simultaneously.

The various methods and techniques described above provide a number of ways to carry out the invention. Of course, it is to be understood that not necessarily all objectives or advantages described may be achieved in accordance with any particular embodiment described herein. Thus, for example, those skilled in the art will recognize that the methods can be performed in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other objectives or advantages as may be taught or suggested herein. A variety of advantageous and disadvantageous alternatives are mentioned herein. It is to be understood that some preferred embodiments specifically include one, another, or several advantageous features, while others specifically exclude one, another, or several disadvantageous features, while still others specifically mitigate a present disadvantageous feature by inclusion of one, another, or several advantageous features.

Furthermore, the skilled artisan will recognize the applicability of various features from different embodiments. Similarly, the various elements, features and steps discussed above, as well as other known equivalents for each such element, feature or step, can be mixed and matched by one of ordinary skill in this art to perform methods in accordance with principles described herein. Among the various elements, features, and steps some will be specifically included and others specifically excluded in diverse embodiments.

Although the invention has been disclosed in the context of certain embodiments and examples, it will be understood by those skilled in the art that the embodiments of the

invention extend beyond the specifically disclosed embodiments to other alternative embodiments and/or uses and modifications and equivalents thereof.

Many variations and alternative elements have been disclosed in embodiments of the present invention. Still further variations and alternate elements will be apparent to one of skill in the art. Among these variations, without limitation, are the compositions for, and methods of, genetic editing, in vivo methods associated with genetic editing, compositions of cells generated by the aforementioned techniques, treatment of diseases and/or conditions that relate to the teachings of the invention, techniques and composition and use of solutions used therein, and the particular use of the products created through the teachings of the invention. Various embodiments of the invention can specifically include or exclude any of these variations or elements.

In some embodiments, the numbers expressing quantities of ingredients, properties such as concentration, reaction conditions, and so forth, used to describe and claim certain embodiments of the invention are to be understood as being modified in some instances by the term "about." Accordingly, in some embodiments, the numerical parameters set forth in the written description and attached claims are approximations that can vary depending upon the desired properties sought to be obtained by a particular embodiment. In some embodiments, the numerical parameters should be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. Notwithstanding that the numerical ranges and parameters setting forth the broad scope of some embodiments of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as practicable. The numerical values presented in some embodiments of the invention may contain certain errors necessarily resulting from the standard deviation found in their respective testing measurements.

In some embodiments, the terms "a" and "an" and "the" and similar references used in the context of describing a particular embodiment of the invention (especially in the context of certain of the following claims) can be construed to cover both the singular and the plural. The recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g. "such as") provided with

respect to certain embodiments herein is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention otherwise claimed. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the invention.

Groupings of alternative elements or embodiments of the invention disclosed herein are not to be construed as limitations. Each group member can be referred to and claimed individually or in any combination with other members of the group or other elements found herein. One or more members of a group can be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the specification is herein deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

Preferred embodiments of this invention are described herein, including the best mode known to the inventor for carrying out the invention. Variations on those preferred embodiments will become apparent to those of ordinary skill in the art upon reading the foregoing description. It is contemplated that skilled artisans can employ such variations as appropriate, and the invention can be practiced otherwise than specifically described herein. Accordingly, many embodiments of this invention include all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

Furthermore, numerous references have been made to patents and printed publications throughout this specification. Each of the above cited references and printed publications are herein individually incorporated by reference in their entirety.

In closing, it is to be understood that the embodiments of the invention disclosed herein are illustrative of the principles of the present invention. Other modifications that can be employed can be within the scope of the invention. Thus, by way of example, but not of limitation, alternative configurations of the present invention can be utilized in accordance with the teachings herein. Accordingly, embodiments of the present invention are not limited to that precisely as shown and described.

THE CLAIMS

1. A composition comprising a vector encoding a fusion protein comprising at least one endonuclease and a DNA binding moiety.

2. The composition of claim 1, wherein the fusion protein comprises at least one endonuclease selected from the group consisting of: cas regularly interspaced short palindromic (CRISPR) protein, a zinc finger nuclease (ZFNs) and transcription activator-like effector nucleases (TALENs).

3. The composition of claim 2, wherein the CRISPR protein comprises cas9.

4. The composition of claim 1, wherein the DNA binding moiety comprises a zinc finger protein.

5. The composition of claim 4, wherein the zinc finger comprises a left handed CCR5 binding protein.

6. The composition of claim 1, wherein the at least one endonuclease and DNA binding moiety are joined by a linker comprising two, three, four, five, six, seven, eight, nine, ten or more amino acids.

7. The composition of claim 1, wherein the fusion protein comprises a fluorescent labeled protein.

8. The composition of claim 7, wherein the fluorescent labeled protein comprises one or more proteins selected from the group consisting of: green fluorescent protein (GFP), enhanced (eGFP), red fluorescent protein (RFP) and mCherry.

9. The composition of claim 1, wherein the fusion protein comprises a nuclear localization signal (NLS).

10. The composition of claim 9, wherein the NLS is SV40 NLS.

11. A method of genomic editing comprising:

(a) providing a quantity of one or more vectors encoding a fusion protein comprising at least one endonuclease and a DNA binding moiety; and

(b) contacting a population of cells with the quantity of the one or more vectors, wherein the fusion protein is capable of inducing double stranded break (DSB) and homologous recombination (HR) of the DSB results in editing of the genome of the population of cells.

12. The method of claim 11, further comprising contacting the population of cells with one or more guide RNAs (gRNAs) in step (b).

13. The method of claim 11, further comprising contacting the population of cells with template DNA in step (b).

14. The method of claim 13, wherein the template DNA comprises at least one expression cassette, two flanking sequences, and a DNA binding moiety sequence.

15. The method of claim 14, wherein the two flanking sequences are each at least 10 bp, and homologous to sequences in the genome of the population of cells.

16. The method of claim 15, wherein the DNA binding moiety sequence comprises CCR5.

17. The method of claim 11, wherein contacting the population of cells comprises a technique selected from the group consisting of: transfection, electroporation, and transformation.

18. The method of claim 11, wherein the population of cells comprise stem cells or progenitor cells.

19. A kit for genomic editing comprising:

one or more vectors encoding a fusion protein comprising at least one endonuclease and a DNA binding moiety; and

template DNA comprising at least one expression cassette, two flanking sequences, and a DNA binding moiety sequence.

20. The kit of claim 19, further comprising one or more guide RNAs (gRNAs).

21. The kit of claim 19, wherein the fusion protein comprises at least one endonuclease CRISPR protein, a DNA binding moiety that is a zinc finger protein.

22. The kit of claim 21, further comprising a fluorescent labeled protein.

23. The kit of claim 21, further comprising a nuclear localization signal (NLS).

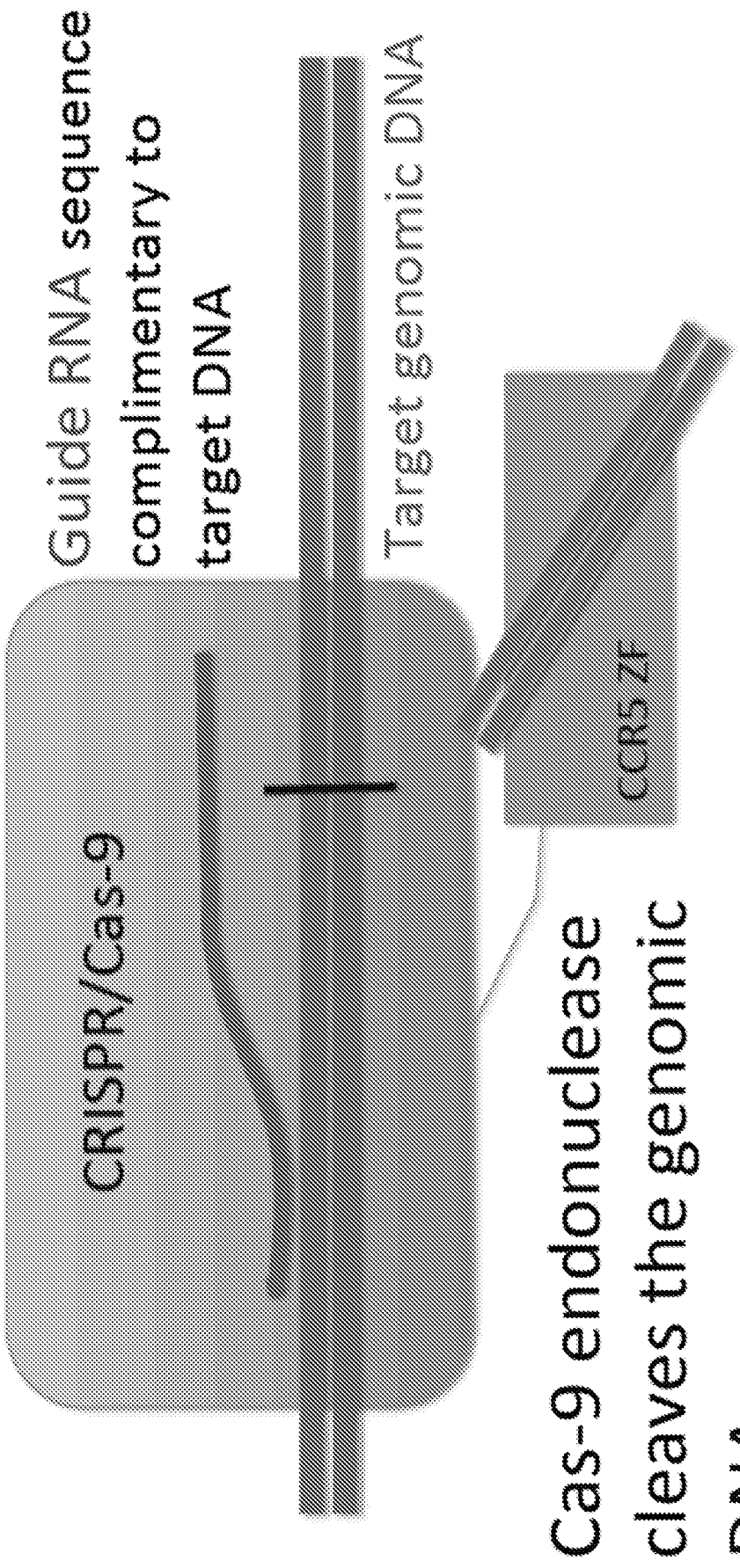

CRISPR/Cas-9
Guide RNA sequence complimentary to target DNA
Target genomic DNA
CCR5 ZF
Cas-9 endonuclease cleaves the genomic DNA

Figure 1

Figure 2

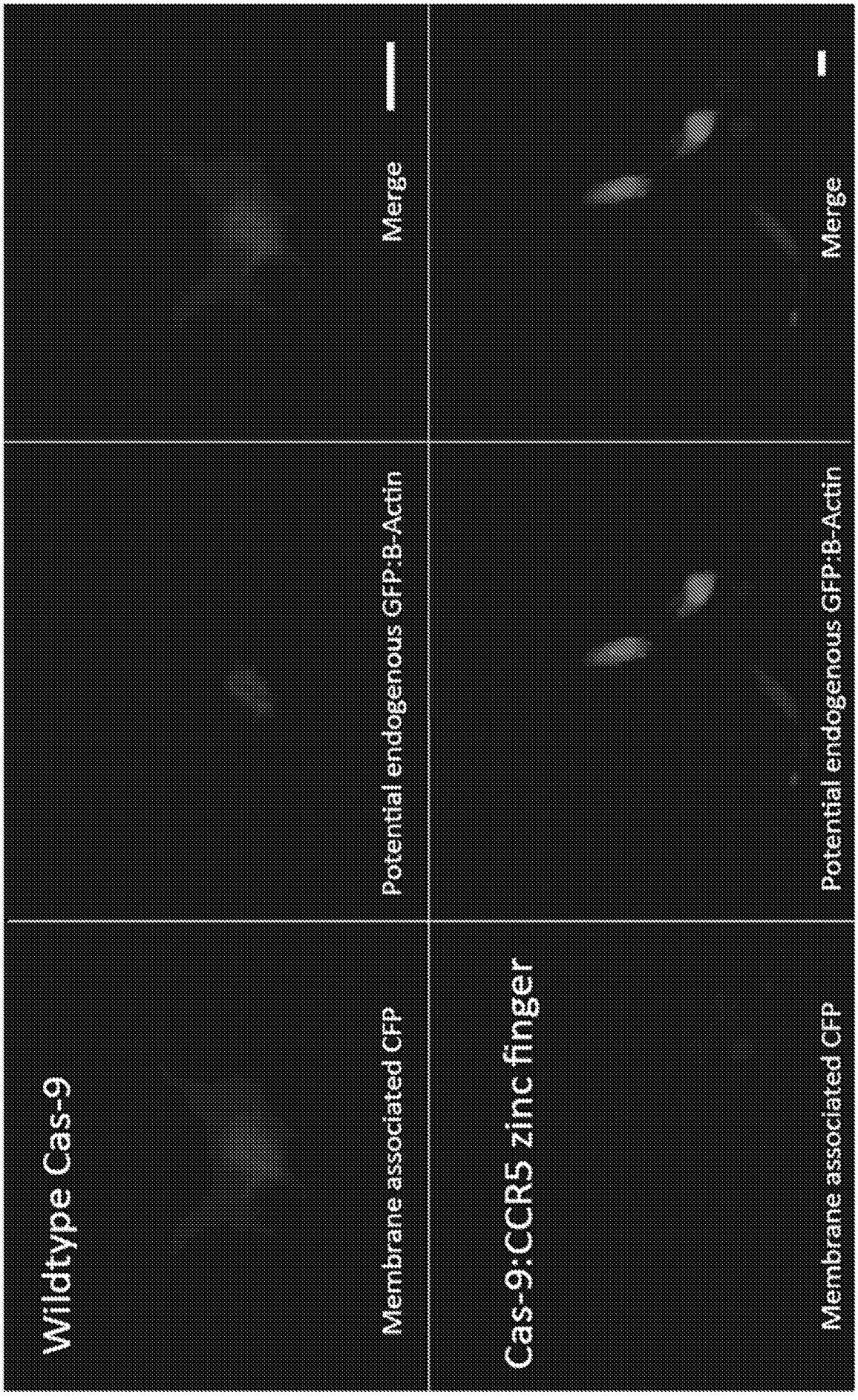

Wildtype Cas-9
Membrane associated CFP
Potential endogenous GFP:B-Actin
Merge
Cas-9:CCR5 zinc finger
Membrane associated CFP
Potential endogenous GFP:B-Actin
Merge

Figure 3

Fig. 3A

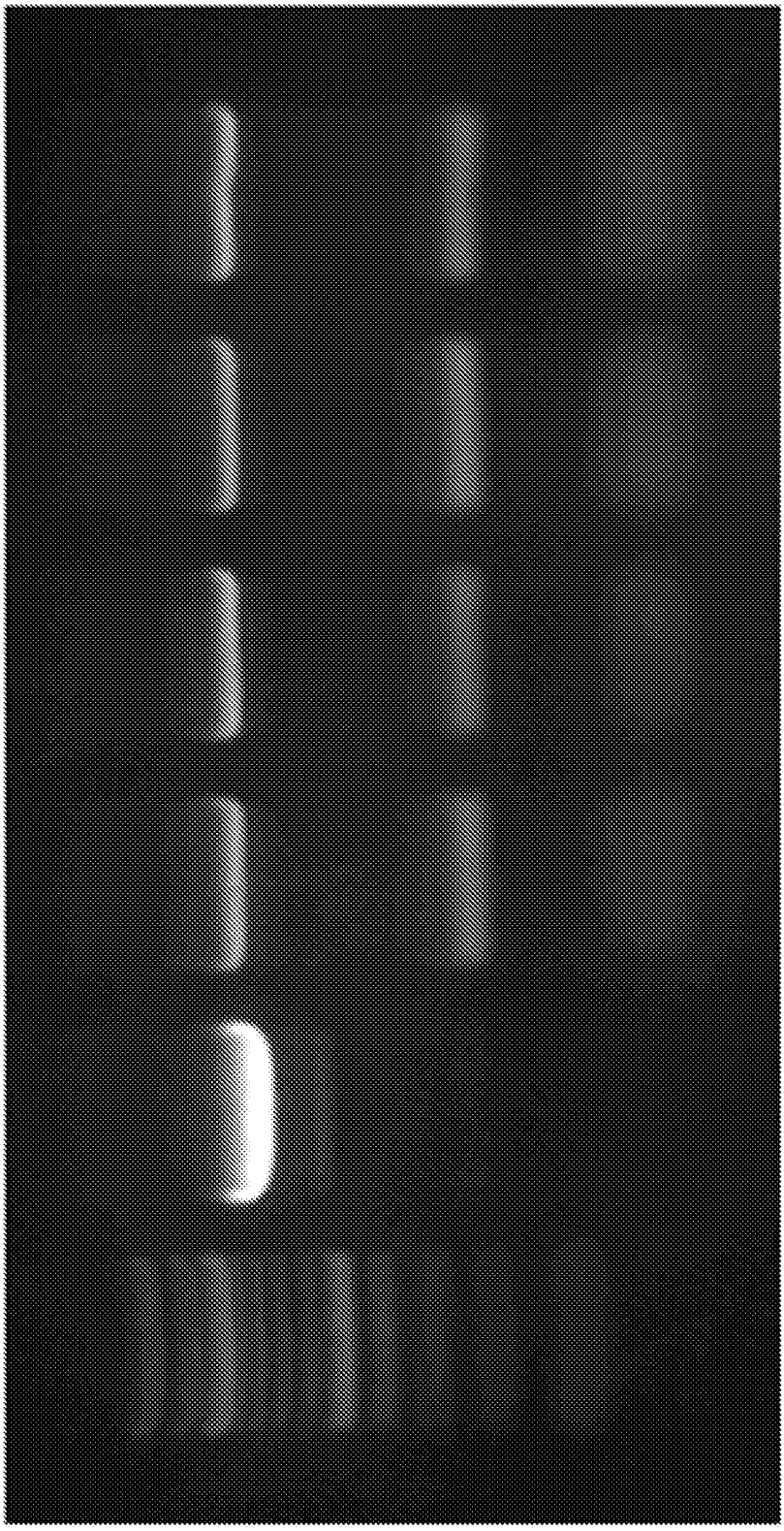


Figure 3

Fig. 3B

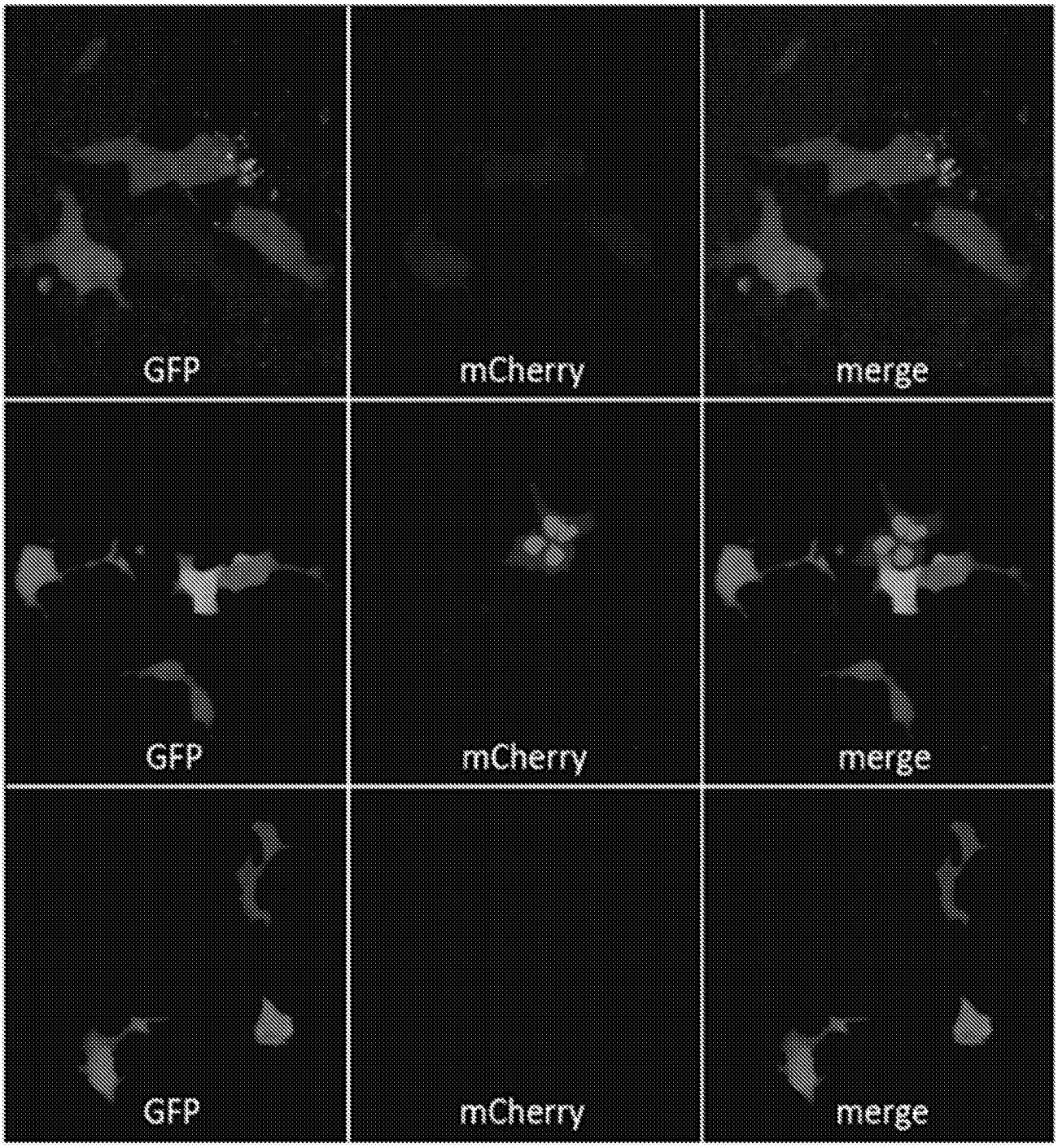


Figure 4

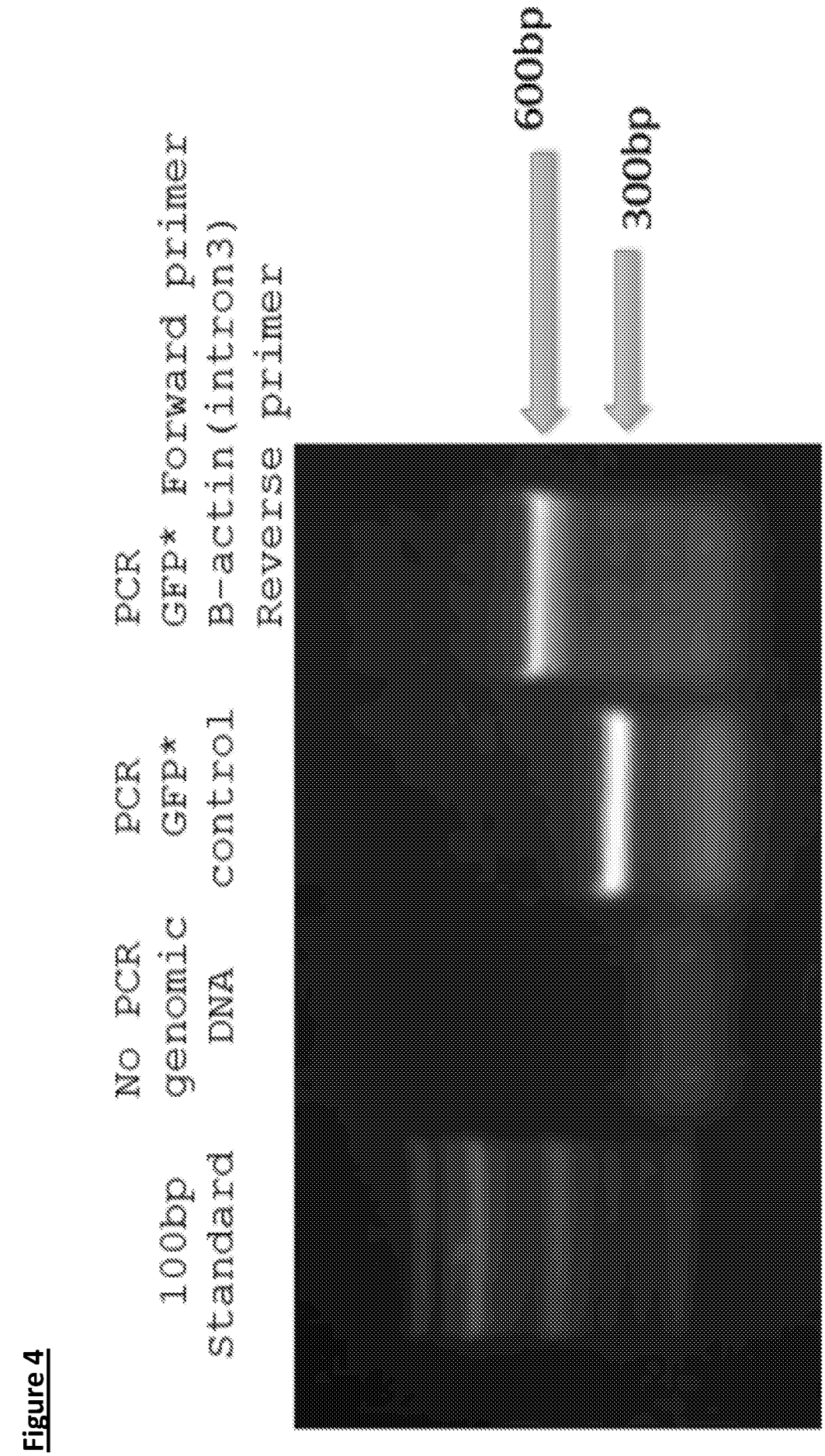

100bp Standard
No PCR genomic DNA
PCR GFP* control
PCR GFP* Forward primer B-actin(intron3) Reverse primer
600bp
300bp

Figure 5

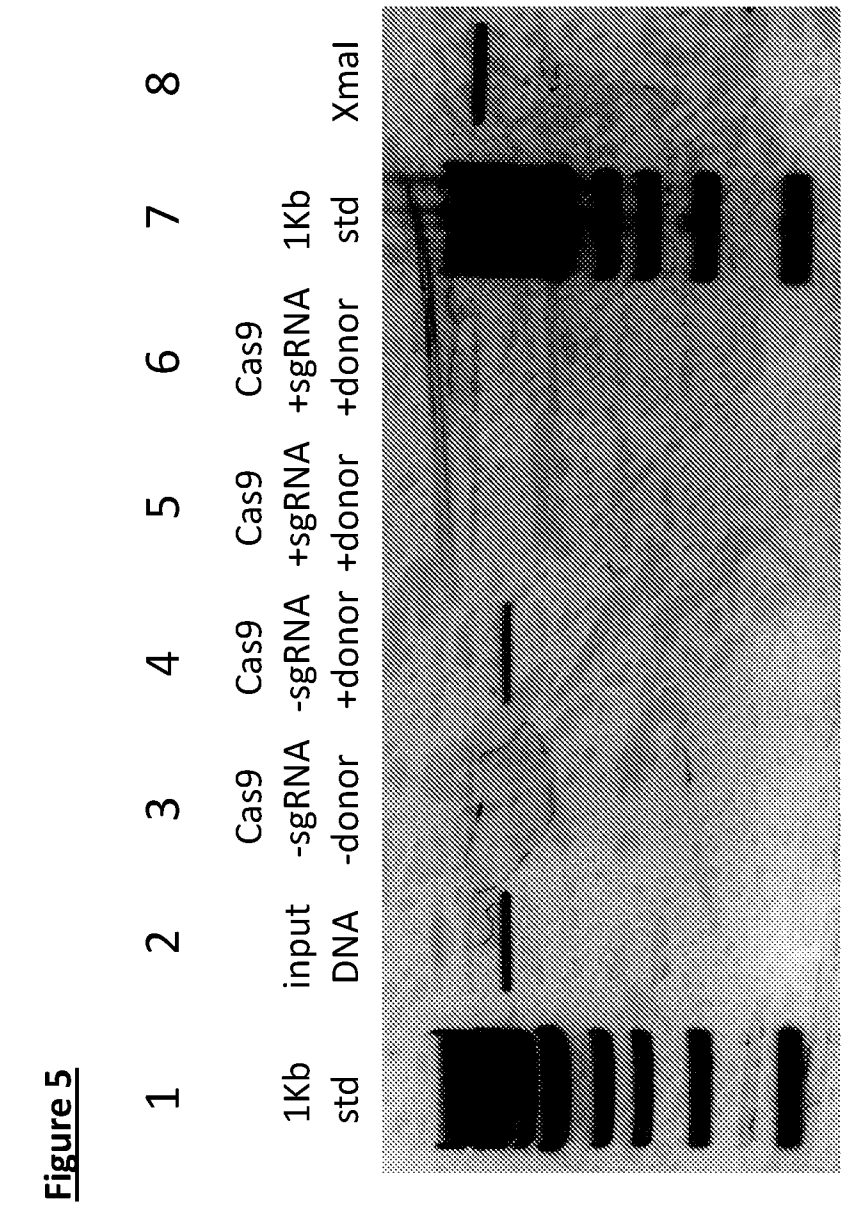

1
2
3
4
5
6
7
8
1Kb std
input DNA
Cas9 -sgRNA -donor
Cas9 -sgRNA +donor
Cas9 +sgRNA -donor
Cas9 +sgRNA +donor
1Kb std
XmaI

Figure 5

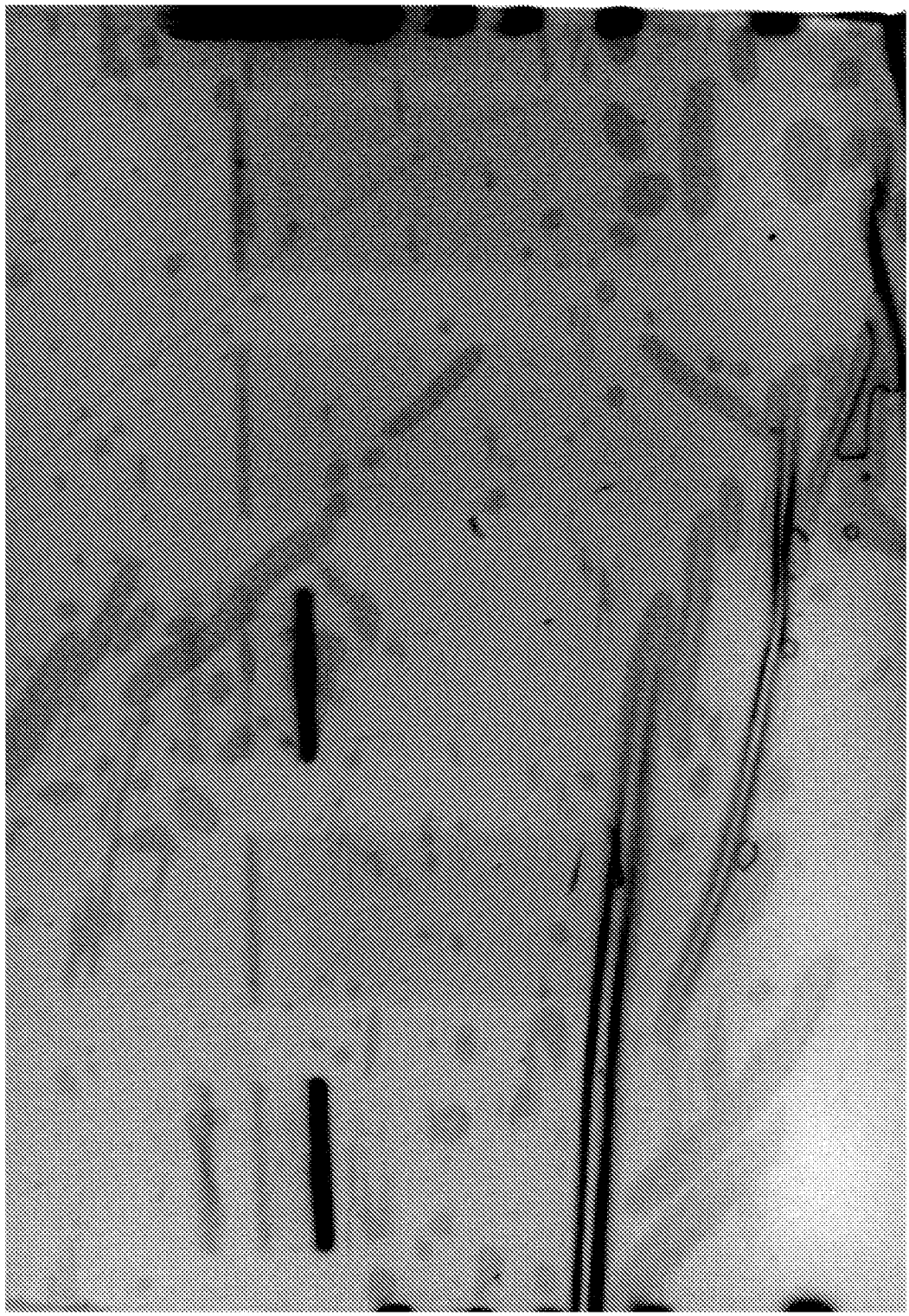

2
Cas9
-sgRNA
-donor
3
Cas9
-sgRNA
+donor
4
Cas9
+sgRNA
+donor
5
6
Cas9
+sgRNA
+donor

Figure 6

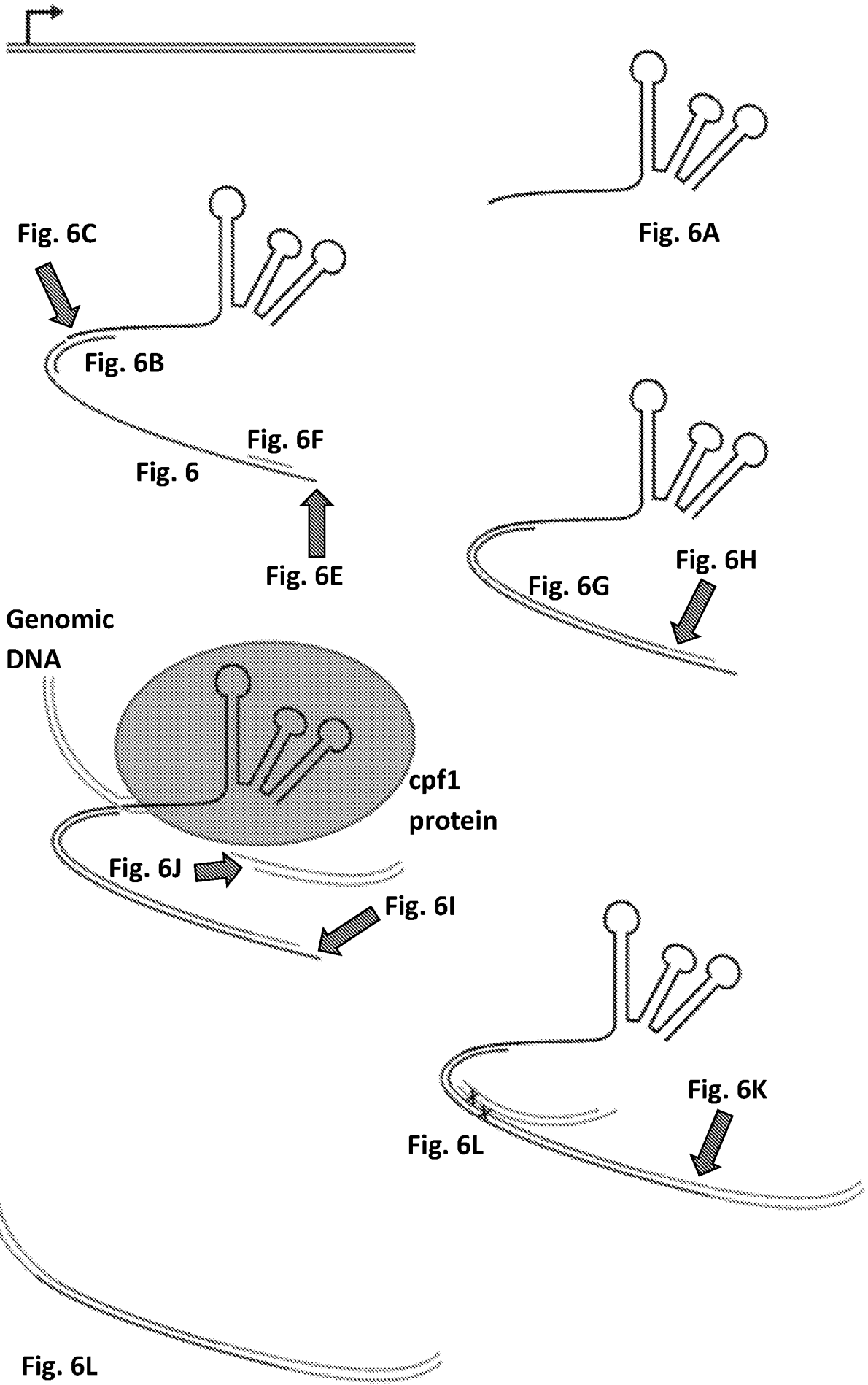


Fig. 6C
Fig. 6A
Fig. 6B
Fig. 6F
Fig. 6
Fig. 6E
Fig. 6G
Fig. 6H
Genomic
DNA
cpf1
protein
Fig. 6J
Fig. 6I
Fig. 6K
Fig. 6L
Fig. 6L

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/032367

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 9/22; C12N 15/10; C12N 15/85 (2016.01)
CPC - C12N 9/22; C12N 15/1082; C12N 15/63 (2016.08)
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)
IPC - C12N 9/22; C12N 15/10; C12N 15/85
CPC - C12N 9/22; C12N 15/1082; C12N 15/63

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 435/463; 435/320.1; 536.23.1 (keyword delimited)

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

Search terms used: CRISPR cas9 vector cassette editing targeting GFP flanking NLS SV40 ZFN fusion

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 2015/0044772 A1 (SAGE LABS, INC.) 12 February 2015 (12.02.2015) entire document	1-4, 6, 9, 11-13, 17 --- 5, 7, 8, 10, 14-16, 18-23
Y	US 2014/0287426 A1 (ARNOLD et al) 25 September 2014 (25.09.2014) entire document	5, 16
Y	US 2014/0356958 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 04 December 2014 (04.12.2014) entire document	7, 8, 10, 18, 22
Y	US 2015/0067922 A1 (THE PENN STATE RESEARCH FOUNDATION) 05 March 2015 (05.03.2015) entire document	14-16, 19-23
A	NISSIM et al. "Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells," Mol. Cell., 22 May 2014 (22.05.2014), Vol. 54, No. 4, Pgs. 698-710. entire document	1-23

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
03 August 2016	02 SEP 2016
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)