RNA-GUIDED HUMAN GENOME ENGINEERING

RELATED APPLICATION DATA

This application claims priority to U.S. Provisional Application No. 61/779,169, filed on March 13, 2013 and U.S. Provisional Application No. 61/738,355, filed on December 17, 2012 each of which is hereby incorporated herein by reference in their entirety for all purposes.

STATEMENT OF GOVERNMENT INTERESTS

This invention was made with government support under P50 HG005550 awarded by National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

Bacterial and archaeal CRISPR systems rely on crRNAs in complex with Cas proteins to direct degradation of complementary sequences present within invading viral and plasmid DNA (*1-3*). A recent *in vitro* reconstitution of the *S. pyogenes* type II CRISPR system demonstrated that crRNA fused to a normally trans-encoded tracrRNA is sufficient to direct Cas9 protein to sequence-specifically cleave target DNA sequences matching the crRNA (*4*).

SUMMARY

The present disclosure references documents numerically which are listed at the end of the present disclosure. The document corresponding to the number is incorporated by reference into the specification as a supporting reference corresponding to the number as if fully cited.

According to one aspect of the present disclosure, a eukaryotic cell is transfected with a two component system including RNA complementary to genomic DNA and an enzyme that interacts with the RNA. The RNA and the enzyme are expressed by the cell. The RNA of the RNA/enzyme complex then binds to complementary genomic DNA. The enzyme then performs a function, such as cleavage of the genomic DNA. The RNA includes between about 10 nucleotides to about 250 nucleotides. The RNA includes between about 20 nucleotides to about 100 nucleotides. According to certain aspects, the enzyme may perform any desired function in a site specific manner for which the enzyme has been engineered. According to one aspect, the eukaryotic cell is a yeast cell, plant cell or mammalian cell. According to one aspect, the enzyme cleaves genomic sequences targeted by RNA sequences (see references (*4-6*)), thereby creating a genomically altered eukaryotic cell.

According to one aspect, the present disclosure provides a method of genetically altering a human cell by including a nucleic acid encoding an RNA complementary to genomic DNA into the genome of the cell and a nucleic acid encoding an enzyme that performs a desired function on

genomic DNA into the genome of the cell. According to one aspect, the RNA and the enzyme are expressed, According to one aspect, the RNA hybridizes with complementary genomic DNA. According to one aspect, the enzyme is activated to perform a desired function, such as cleavage, in a site specific manner when the RNA is hybridized to the complementary genomic DNA. According to one aspect, the RNA and the enzyme are components of a bacterial Type II CRISPR system.

According to one aspect, a method of altering a eukaryotic cell is providing including transfecting the eukaryotic cell with a nucleic acid encoding RNA complementary to genomic DNA of the eukaryotic cell, transfecting the eukaryotic cell with a nucleic acid encoding an enzyme that interacts with the RNA and cleaves the genomic DNA in a site specific manner, wherein the cell expresses the RNA and the enzyme, the RNA binds to complementary genomic DNA and the enzyme cleaves the genomic DNA in a site specific manner. According to one aspect, the enzyme is Cas9 or modified Cas9 or a homolog of Cas9. According to one aspect, the eukaryotic cell is a yeast cell, a plant cell or a mammalian cell. According to one aspect, the RNA includes between about 10 to about 250 nucleotides. According to one aspect, the RNA includes between about 20 to about 100 nucleotides.

According to one aspect, a method of altering a human cell is provided including transfecting the human cell with a nucleic acid encoding RNA complementary to genomic DNA of the eukaryotic cell, transfecting the human cell with a nucleic acid encoding an enzyme that interacts with the RNA and cleaves the genomic DNA in a site specific manner, wherein the human cell expresses the RNA and the enzyme, the RNA binds to complementary genomic DNA and the enzyme cleaves the genomic DNA in a site specific manner. According to one aspect, the enzyme is Cas9 or modified Cas9 or a homolog of Cas9. According to one aspect, the RNA includes between about 10 to about 250 nucleotides. According to one aspect, the RNA includes between about 20 to about 100 nucleotides.

According to one aspect, a method of altering a eukaryotic cell at a plurality of genomic DNA sites is provided including transfecting the eukaryotic cell with a plurality of nucleic acids encoding RNAs complementary to different sites on genomic DNA of the eukaryotic cell, transfecting the eukaryotic cell with a nucleic acid encoding an enzyme that interacts with the RNA and cleaves the genomic DNA in a site specific manner, wherein the cell expresses the RNAs and the enzyme, the RNAs bind to complementary genomic DNA and the enzyme cleaves the genomic DNA in a site specific manner. According to one aspect, the enzyme is Cas9. According to one aspect, the eukaryotic cell is a yeast cell, a plant cell or a mammalian cell. According to one aspect, the RNA includes between about 10 to about 250 nucleotides. According to one aspect, the RNA includes between about 20 to about 100 nucleotides.

DETAILED DESCRIPTION

According to one aspect, a human codon-optimized version of the Cas9 protein bearing a C-terminus SV40 nuclear localization signal is synthetized and cloned into a mammalian expression system (Fig. 1A and Fig. 3A). Accordingly, Figure 1 is directed to genome editing in human cells using an engineered type II CRISPR system. As shown in Figure 1A, RNA-guided gene targeting in human cells involves co-expression of the Cas9 protein bearing a C-terminus SV40 nuclear localization signal with one or more guide RNAs (gRNAs) expressed from the human U6 polymerase III promoter. Cas9 unwinds the DNA duplex and cleaves both strands upon recognition of a target sequence by the gRNA, but only if the correct protospacer-adjacent motif (PAM) is present at the 3' end. Any genomic sequence of the form $GN_{20}GG$ can in principle be targeted. As shown in Figure 1B, a genomically integrated GFP coding sequence is disrupted by the insertion of a stop codon and a 68bp genomic fragment from the AAVS1 locus. Restoration of the GFP sequence by homologous recombination (HR) with an appropriate donor sequence results in GFP^+ cells that can be quantitated by FACS. T1 and T2 gRNAs target sequences within the AAVS1 fragment. Binding sites for the two halves of the TAL effector nuclease heterodimer (TALEN) are underlined. As shown in Figure 1C, bar graph depict HR efficiencies induced by T1, T2, and TALEN-mediated nuclease activity at the target locus, as measured by FACS. Representative FACS plots and microscopy images of the targeted cells are depicted below (scale bar is 100 microns). Data is mean +/- SEM (N=3).

According to one aspect, to direct Cas9 to cleave sequences of interest, crRNA-tracrRNA fusion transcripts are expressed, hereafter referred to as guide RNAs (gRNAs), from the human U6 polymerase III promoter. According to one aspect, gRNAs are directly transcribed by the cell. This aspect advantageously avoids reconstituting the RNA processing machinery employed by bacterial CRISPR systems (Fig. 1A and Fig. 3B) (see references (*4, 7-9*)). According to one aspect, a method is provided for altering genomic DNA using a U6 transcription initiating with G and a PAM (protospacer-adjacent motif) sequence -NGG following the 20 bp crRNA target. According to this aspect, the target genomic site is in the form of $GN_{20}GG$ (See Fig. 3C).

According to one aspect, a GFP reporter assay (Fig. 1B) in 293T cells was developed similar to one previously described (see reference (*10*)) to test the functionality of the genome engineering methods described herein. According to one aspect, a stable cell line was established bearing a genomically integrated GFP coding sequence disrupted by the insertion of a stop codon and a 68bp genomic fragment from the AAVS1 locus that renders the expressed protein fragment non-fluorescent. Homologous recombination (HR) using an appropriate repair donor can restore the normal GFP sequence, which allows one to quantify the resulting GFP^+ cells by flow activated cell sorting (FACS).

According to one aspect, a method is provided of homologous recombination (HR). Two gRNAs are constructed, T1 and T2, that target the intervening AAVS1 fragment (Fig. 1b). Their activity to that of a previously described TAL effector nuclease heterodimer (TALEN) targeting the same region (see reference (*11*)) was compared. Successful HR events were observed using all three targeting reagents, with gene correction rates using the T1 and T2 gRNAs approaching 3% and 8% respectively (Fig. 1C). This RNA-mediated editing process was notably rapid, with the first detectable GFP^+ cells appearing ~20 hours post transfection compared to ~40 hours for the AAVS1 TALENs. HR was observed only upon simultaneous introduction of the repair donor, Cas9 protein, and gRNA, confirming that all components are required for genome editing (Fig. 4). While no apparent toxicity associated with Cas9/crRNA expression was noted, work with ZFNs and TALENs has shown that nicking only one strand further reduces toxicity. Accordingly, a Cas9D10A mutant was tested that is known to function as a nickase *in vitro*, which yielded similar HR but lower non-homologous end joining (NHEJ) rates (Fig. 5) (see references (*4, 5*)). Consistent with (*4*) where a related Cas9 protein is shown to cut both strands 6 bp upstream of the PAM, NHEJ data confirmed that most deletions or insertions occurred at the 3' end of the target sequence (Fig. 5B). Also confirmed was that mutating the target genomic site prevents the gRNA from effecting HR at that locus, demonstrating that CRISPR-mediated genome editing is sequence specific (Fig. 6). It was showed that two gRNAs targeting sites in the GFP gene, and also three additional gRNAs targeting fragments from homologous regions of the DNA methyl transferase 3a (DNMT3a) and DNMT3b genes could sequence specifically induce significant HR in the engineered reporter cell lines (Fig. 7, 8). Together these results confirm that RNA-guided genome targeting in human cells induces robust HR across multiple target sites.

According to certain aspects, a native locus was modified. gRNAs were used to target the AAVS1 locus located in the PPP1R12C gene on chromosome 19, which is ubiquitously expressed across most tissues (Fig. 2A) in 293Ts, K562s, and PGP1 human iPS cells (see reference (*12*)) and analyzed the results by next-generation sequencing of the targeted locus. Accordingly, Figure 2 is directed to RNA-guided genome editing of the native AAVS1 locus in multiple cell types. As shown in Figure 2A, T1 (red) and T2 (green) gRNAs target sequences in an intron of the PPP1R12C gene within the chromosome 19 AAVS1 locus. As shown in Figure 2B, total count and location of deletions caused by NHEJ in 293Ts, K562s, and PGP1 iPS cells following expression of Cas9 and either T1 or T2 gRNAs as quantified by next-generation sequencing is provided. Red and green dash lines demarcate the boundaries of the T1 and T2 gRNA targeting sites. NHEJ frequencies for T1 and T2 gRNAs were 10% and 25% in 293T, 13% and 38% in K562, and 2% and 4% in PGP1 iPS cells, respectively. As shown in Figure 2C, DNA donor architecture for HR at the AAVS1 locus, and the locations of sequencing primers (arrows) for detecting successful targeted events, are depicted. As shown in Figure 2D, PCR assay three days

post transfection demonstrates that only cells expressing the donor, Cas9 and T2 gRNA exhibit successful HR events. As shown in Figure 2E, successful HR was confirmed by Sanger sequencing of the PCR amplicon showing that the expected DNA bases at both the genome-donor and donor-insert boundaries are present. As shown in Figure 2F, successfully targeted clones of 293T cells were selected with puromycin for 2 weeks. Microscope images of two representative GFP+ clones is shown (scale bar is 100 microns).

Consistent with results for the GFP reporter assay, high numbers of NHEJ events were observed at the endogenous locus for all three cell types. The two gRNAs T1 and T2 achieved NHEJ rates of 10 and 25% in 293Ts, 13 and 38% in K562s, and 2 and 4% in PGP1-iPS cells, respectively (Fig. 2B). No overt toxicity was observed from the Cas9 and crRNA expression required to induce NHEJ in any of these cell types (Fig. 9). As expected, NHEJ-mediated deletions for T1 and T2 were centered around the target site positions, further validating the sequence specificity of this targeting process (Fig. 9, 10, 11). Simultaneous introduction of both T1 and T2 gRNAs resulted in high efficiency deletion of the intervening 19bp fragment (Fig. 10), demonstrating that multiplexed editing of genomic loci is feasible using this approach.

According to one aspect, HR is used to integrate either a dsDNA donor construct (see reference (*13*)) or an oligo donor into the native AAVS1 locus (Fig. 2C, Fig. 12). HR-mediated integration was confirmed using both approaches by PCR (Fig. 2D, Fig. 12) and Sanger sequencing (Fig. 2E). 293T or iPS clones were readily derived from the pool of modified cells using puromycin selection over two weeks (Fig. 2F, Fig. 12). These results demonstrate that Cas9 is capable of efficiently integrating foreign DNA at endogenous loci in human cells. Accordingly, one aspect of the present disclosure includes a method of integrating foreign DNA into the genome of a cell using homologous recombination and Cas9.

According to one aspect, an RNA-guided genome editing system is provided which can readily be adapted to modify other genomic sites by simply modifying the sequence of the gRNA expression vector to match a compatible sequence in the locus of interest. According to this aspect, 190,000 specifically gRNA-targetable sequences targeting about 40.5% exons of genes in the human genome were generated. These target sequences were incorporated into a 200bp format compatible with multiplex synthesis on DNA arrays (see reference (*14*)) (Fig. 13). According to this aspect, a ready genome-wide reference of potential target sites in the human genome and a methodology for multiplex gRNA synthesis is provided.

According to one aspect, methods are provided for multiplexing genomic alterations in a cell by using one or more or a plurality of RNA/enzyme systems described herein to alter the genome of a cell at a plurality of locations. According to one aspect, target sites perfectly match the PAM sequence NGG and the 8-12 base "seed sequence" at the 3' end of the gRNA. According to certain aspects, perfect match is not required of the remaining 8-12 bases. According to certain

aspects, Cas9 will function with single mismatches at the 5' end. According to certain aspects, the target locus's underlying chromatin structure and epigenetic state may affect efficiency of Cas9 function. According to certain aspects, Cas9 homologs having higher specificity are included as useful enzymes. One of skill in the art will be able to identify or engineer suitable Cas9 homologs. According to one aspect, CRISPR-targetable sequences include those having different PAM requirements (see reference (*9*)), or directed evolution. According to one aspect, inactivating one of the Cas9 nuclease domains increases the ratio of HR to NHEJ and may reduce toxicity (Fig. 3A, Fig. 5) (*4, 5*), while inactivating both domains may enable Cas9 to function as a retargetable DNA binding protein. Embodiments of the present disclosure have broad utility in synthetic biology (see references (*21, 22*)), the direct and multiplexed perturbation of gene networks (see references (*13, 23*)), and targeted *ex vivo* (see references (*24-26*)) and *in vivo* gene therapy (see reference (*27*)).

According to certain aspects, a "re-engineerable organism" is provided as a model system for biological discovery and in vivo screening. According to one aspect, a "re-engineerable mouse" bearing an inducible Cas9 transgene is provided, and localized delivery (using adeno-associated viruses, for example) of libraries of gRNAs targeting multiple genes or regulatory elements allow one to screen for mutations that result in the onset of tumors in the target tissue type. Use of Cas9 homologs or nuclease-null variants bearing effector domains (such as activators) allow one to multiplex activate or repress genes in vivo. According to this aspect, one could screen for factors that enable phenotypes such as: tissue-regeneration, trans-differentiation etc. According to certain aspects, (a) use of DNA-arrays enables multiplex synthesis of defined gRNA libraries (refer Fig. 13); and (b) gRNAs being small in size (refer Fig. 3b) are packaged and delivered using a multitude of non-viral or viral delivery methods.

According to one aspect, the lower toxicities observed with "nickases" for genome engineering applications is achieved by inactivating one of the Cas9 nuclease domains, either the nicking of the DNA strand base-paired with the RNA or nicking its complement. Inactivating both domains allows Cas9 to function as a retargetable DNA binding protein. According to one aspect, the Cas9 retargetable DNA binding protein is attached

(a) to transcriptional activation or repression domains for modulating target gene expression, including but not limited to chromatin remodeling, histone modification, silencing, insulation, direct interactions with the transcriptional machinery;

(b) to nuclease domains such as FokI to enable 'highly specific' genome editing contingent upon dimerization of adjacent gRNA-Cas9 complexes;

(c) to fluorescent proteins for visualizing genomic loci and chromosome dynamics; or

(d) to other fluorescent molecules such as protein or nucleic acid bound organic fluorophores, quantum dots, molecular beacons and echo probes or molecular beacon replacements;

(e) to multivalent ligand-binding protein domains that enable programmable manipulation of genome-wide 3D architecture.

According to one aspect, the transcriptional activation and repression components can employ CRISPR systems naturally or synthetically orthogonal, such that the gRNAs only bind to the activator or repressor class of Cas. This allows a large set of gRNAs to tune multiple targets.

According to certain aspects, the use of gRNAs provide the ability to multiplex than mRNAs in part due to the smaller size -- 100 vs. 2000 nucleotide lengths respectively. This is particularly valuable when nucleic acid delivery is size limited, as in viral packaging. This enables multiple instances of cleavage, nicking, activation, or repression – or combinations thereof. The ability to easily target multiple regulatory targets allows the coarse-or-fine-tuning or regulatory networks without being constrained to the natural regulatory circuits downstream of specific regulatory factors (e.g. the 4 mRNAs used in reprogramming fibroblasts into IPSCs). Examples of multiplexing applications include:

1. Establishing (major and minor) histocompatibility alleles, haplotypes, and genotypes for human (or animal) tissue/organ transplantation. This aspect results e.g. in HLA homozygous cell lines or humanized animal breeds – or – a set of gRNAs capable of superimposing such HLA alleles onto an otherwise desirable cell lines or breeds.

2. Multiplex cis-regulatory element (CRE = signals for transcription, splicing, translation, RNA and protein folding, degradation, etc.) mutations in a single cell (or a collection of cells) can be used for efficiently studying the complex sets of regulatory interaction that can occur in normal development or pathological, synthetic or pharmaceutical scenarios. According to one aspect, the CREs are (or can be made) somewhat orthogonal (i.e. low cross talk) so that many can be tested in one setting -- e.g. in an expensive animal embryo time series. One exemplary application is with RNA fluorescent in situ sequencing (FISSeq).

3. Multiplex combinations of CRE mutations and/or epigenetic activation or repression of CREs can be used to alter or reprogram iPSCs or ESCs or other stem cells or non-stem cells to any cell type or combination of cell types for use in organs-on-chips or other cell and organ cultures for purposes of testing pharmaceuticals (small molecules, proteins, RNAs, cells, animal, plant or microbial cells, aerosols and other delivery methods), transplantation strategies, personalization strategies, etc.

4. Making multiplex mutant human cells for use in diagnostic testing (and/or DNA sequencing) for medical genetics. To the extent that the chromosomal location and context of a human genome allele (or epigenetic mark) can influence the accuracy of a clinical genetic diagnosis, it is important to have alleles present in the correct location in a reference genome – rather than in an ectopic (aka transgenic) location or in a separate piece of synthetic DNA. One embodiment is a series of independent cell lines one per each diagnostic human SNP, or structural variant. Alternatively, one

embodiment includes multiplex sets of alleles in the same cell. In some cases multiplex changes in one gene (or multiple genes) will be desirable under the assumption of independent testing. In other cases, particular haplotype combinations of alleles allows testing of sequencing (genotyping) methods which accurately establish haplotype phase (i.e. whether one or both copies of a gene are affected in an individual person or somatic cell type.

5. Repetitive elements or endogenous viral elements can be targeted with engineered Cas + gRNA systems in microbes, plants, animals, or human cells to reduce deleterious transposition or to aid in sequencing or other analytic genomic/transcriptomic/proteomic/diagnostic tools (in which nearly identical copies can be problematic).

The following references identified by number in the foregoing section are hereby incorporated by reference in their entireties.

1. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, *Nature* 482, 331 (Feb 16, 2012).
2. D. Bhaya, M. Davison, R. Barrangou, *Annual review of genetics* 45, 273 (2011).
3. M. P. Terns, R. M. Terns, *Current opinion in microbiology* 14, 321 (Jun, 2011).
4. M. Jinek *et al.*, *Science* 337, 816 (Aug 17, 2012).
5. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579 (Sep 25, 2012).
6. R. Sapranauskas *et al.*, *Nucleic acids research* 39, 9275 (Nov, 2011).
7. T. R. Brummelkamp, R. Bernards, R. Agami, *Science* 296, 550 (Apr 19, 2002).
8. M. Miyagishi, K. Taira, *Nature biotechnology* 20, 497 (May, 2002).
9. E. Deltcheva *et al.*, *Nature* 471, 602 (Mar 31, 2011).
10. J. Zou, P. Mali, X. Huang, S. N. Dowey, L. Cheng, *Blood* 118, 4599 (Oct 27, 2011).
11. N. E. Sanjana *et al.*, *Nature protocols* 7, 171 (Jan, 2012).
12. J. H. Lee *et al.*, *PLoS Genet* 5, e1000718 (Nov, 2009).
13. D. Hockemeyer *et al.*, *Nature biotechnology* 27, 851 (Sep, 2009).
14. S. Kosuri *et al.*, *Nature biotechnology* 28, 1295 (Dec, 2010).
15. V. Pattanayak, C. L. Ramirez, J. K. Joung, D. R. Liu, *Nature methods* 8, 765 (Sep, 2011).
16. N. M. King, O. Cohen-Haguenauer, *Molecular therapy : the journal of the American Society of Gene Therapy* 16, 432 (Mar, 2008).
17. Y. G. Kim, J. Cha, S. Chandrasegaran, *Proceedings of the National Academy of Sciences of the United States of America* 93, 1156 (Feb 6, 1996).
18. E. J. Rebar, C. O. Pabo, *Science* 263, 671 (Feb 4, 1994).
19. J. Boch *et al.*, *Science* 326, 1509 (Dec 11, 2009).
20. M. J. Moscou, A. J. Bogdanove, *Science* 326, 1501 (Dec 11, 2009).
21. A. S. Khalil, J. J. Collins, *Nature reviews. Genetics* 11, 367 (May, 2010).
22. P. E. Purnick, R. Weiss, *Nature reviews. Molecular cell biology* 10, 410 (Jun, 2009).

23. J. Zou *et al.*, *Cell stem cell* 5, 97 (Jul 2, 2009).
24. N. Holt *et al.*, *Nature biotechnology* 28, 839 (Aug, 2010).
25. F. D. Urnov *et al.*, *Nature* 435, 646 (Jun 2, 2005).
26. A. Lombardo *et al.*, *Nature biotechnology* 25, 1298 (Nov, 2007).
27. H. Li *et al.*, *Nature* 475, 217 (Jul 14, 2011).

The following examples are set forth as being representative of the present disclosure. These examples are not to be construed as limiting the scope of the present disclosure as these and other equivalent embodiments will be apparent in view of the present disclosure, figures and accompanying claims.

EXAMPLE I

The Type II CRISPR-Cas System

According to one aspect, embodiments of the present disclosure utilize short RNA to identify foreign nucleic acids for activity by a nuclease in a eukaryotic cell. According to a certain aspect of the present disclosure, a eukaryotic cell is altered to include within its genome nucleic acids encoding one or more short RNA and one or more nucleases which are activated by the binding of a short RNA to a target DNA sequence. According to certain aspects, exemplary short RNA / enzyme systems may be identified within bacteria or archaea, such as (CRISPR)/CRISPR-associated (Cas) systems that use short RNA to direct degradation of foreign nucleic acids. CRISPR ("clustered regularly interspaced short palindromic repeats") defense involves acquisition and integration of new targeting "spacers" from invading virus or plasmid DNA into the CRISPR locus, expression and processing of short guiding CRISPR RNAs (crRNAs) consisting of spacer-repeat units, and cleavage of nucleic acids (most commonly DNA) complementary to the spacer.

Three classes of CRISPR systems are generally known and are referred to as Type I, Type II or Type III). According to one aspect, a particular useful enzyme according to the present disclosure to cleave dsDNA is the single effector enzyme, Cas9, common to Type II. (See reference (*1*)). Within bacteria, the Type II effector system consists of a long pre-crRNA transcribed from the spacer-containing CRISPR locus, the multifunctional Cas9 protein, and a tracrRNA important for gRNA processing. The tracrRNAs hybridize to the repeat regions separating the spacers of the pre-crRNA, initiating dsRNA cleavage by endogenous RNase III, which is followed by a second cleavage event within each spacer by Cas9, producing mature crRNAs that remain associated with the tracrRNA and Cas9. According to one aspect, eukaryotic cells of the present disclosure are engineered to avoid use of RNase III and the crRNA processing in general. See reference *(2)*.

According to one aspect, the enzyme of the present disclosure, such as Cas9 unwinds the DNA duplex and searches for sequences matching the crRNA to cleave. Target recognition occurs upon detection of complementarity between a "protospacer" sequence in the target DNA and the remaining spacer sequence in the crRNA. Importantly, Cas9 cuts the DNA only if a correct protospacer-adjacent motif (PAM) is also present at the 3' end. According to certain aspects, different protospacer-adjacent motif can be utilized. For example, the *S. pyogenes* system requires an NGG sequence, where N can be any nucleotide. *S. thermophilus* Type II systems require NGGNG (see reference (*3*)) and NNAGAAW (see reference (*4*)), respectively, while different *S. mutans* systems tolerate NGG or NAAR (see reference (*5*)). Bioinformatic analyses have generated extensive databases of CRISPR loci in a variety of bacteria that may serve to identify additional useful PAMs and expand the set of CRISPR-targetable sequences (see references (*6, 7*)). In *S. thermophilus,* Cas9 generates a blunt-ended double-stranded break 3bp prior to the 3' end of the protospacer (see reference (*8*)), a process mediated by two catalytic domains in the Cas9 protein: an HNH domain that cleaves the complementary strand of the DNA and a RuvC-like domain that cleaves the non-complementary strand (See Figure 1A and Figure 3). While the *S. pyogenes* system has not been characterized to the same level of precision, DSB formation also occurs towards the 3' end of the protospacer. If one of the two nuclease domains is inactivated, Cas9 will function as a nickase *in vitro* (see reference (*2*)) and in human cells (see Figure 5).

According to one aspect, the specificity of gRNA-directed Cas9 cleavage is used as a mechanism for genome engineering in a eukaryotic cell. Accordign to one aspect, hybridization of the gRNA need not be 100 percent in order for the enzyme to recognize the gRNA/DNA hybrid and affect cleavage. Some off-target activity could occur. For example, the *S. pyogenes* system tolerates mismatches in the first 6 bases out of the 20bp mature spacer sequence *in vitro*. According to one aspect, greater stringency may be beneficial *in vivo* when potential off-target sites matching (last 14 bp) NGG exist within the human reference genome for the gRNAs. The effect of mismatches and enzyme activity in general are described in references (*9*), (*2*), (*10*), and (*4*).

According to certain aspects, specificity may be improved. When interference is sensitive to the melting temperature of the gRNA-DNA hybrid, AT-rich target sequences may have fewer off-target sites. Carefully choosing target sites to avoid pseudo-sites with at least 14bp matching sequences elsewhere in the genome may improve specificity. The use of a Cas9 variant requiring a longer PAM sequence may reduce the frequency of off-target sites. Directed evolution may improve Cas9 specificity to a level sufficient to completely preclude off-target activity, ideally requiring a perfect 20bp gRNA match with a minimal PAM. Accordingly, modification to the Cas9 protein is a representative embodiment of the present disclosure. As such, novel methods

permitting many rounds of evolution in a short timeframe (see reference (*11*) and envisioned. CRISPR systems useful in the present disclosure are described in references (*12, 13*).

EXAMPLE II

Plasmid construction

The Cas9 gene sequence was human codon optimized and assembled by hierarchical fusion PCR assembly of 9 500bp gBlocks ordered from IDT. Fig. 3A for the engineered type II CRISPR system for human cells shows the expression format and full sequence of the cas9 gene insert. The RuvC-like and HNH motifs, and the C-terminus SV40 NLS are respectively highlighted by blue, brown and orange colors. Cas9_D10A was similarly constructed. The resulting full-length products were cloned into the pcDNA3.3-TOPO vector (Invitrogen). The target gRNA expression constructs were directly ordered as individual 455bp gBlocks from IDT and either cloned into the pCR-BluntII-TOPO vector (Invitrogen) or pcr amplified. Fig. 3B shows the U6 promoter based expression scheme for the guide RNAs and predicted RNA transcript secondary structure. The use of the U6 promoter constrains the 1st position in the RNA transcript to be a 'G' and thus all genomic sites of the form $GN_{20}GG$ can be targeted using this approach. Fig. 3C shows the 7 gRNAs used.

The vectors for the HR reporter assay involving a broken GFP were constructed by fusion PCR assembly of the GFP sequence bearing the stop codon and 68bp AAVS1 fragment (or mutants thereof; see Fig. 6), or 58bp fragments from the DNMT3a and DNMT3b genomic loci (see Fig. 8) assembled into the EGIP lentivector from Addgene (plasmid #26777). These lentivectors were then used to establish the GFP reporter stable lines. TALENs used in this study were constructed using the protocols described in (*14*). All DNA reagents developed in this study are available at Addgene.

EXAMPLE III

Cell culture

PGP1 iPS cells were maintained on Matrigel (BD Biosciences)-coated plates in mTeSR1 (Stemcell Technologies). Cultures were passaged every 5–7 d with TrypLE Express (Invitrogen). K562 cells were grown and maintained in RPMI (Invitrogen) containing 15% FBS. HEK 293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) high glucose supplemented with 10% fetal bovine serum (FBS, Invitrogen), penicillin/streptomycin (pen/strep,

Invitrogen), and non-essential amino acids (NEAA, Invitrogen). All cells were maintained at 37°C and 5% CO_2 in a humidified incubator.

EXAMPLE IV

Gene targeting of PGP1 iPS, K562 and 293Ts

PGP1 iPS cells were cultured in Rho kinase (ROCK) inhibitor (Calbiochem) 2h before nucleofection. Cells were harvest using TrypLE Express (Invitrogen) and 2×10^6 cells were resuspended in P3 reagent (Lonza) with 1μg Cas9 plasmid, 1μg gRNA and/or 1μg DNA donor plasmid, and nucleofected according to manufacturer's instruction (Lonza). Cells were subsequently plated on an mTeSR1-coated plate in mTeSR1 medium supplemented with ROCK inhibitor for the first 24h. For K562s, 2×10^6 cells were resuspended in SF reagent (Lonza) with 1μg Cas9 plasmid, 1μg gRNA and/or 1μg DNA donor plasmid, and nucleofected according to manufacturer's instruction (Lonza). For 293Ts, 0.1×10^6 cells were transfected with 1μg Cas9 plasmid, 1μg gRNA and/or 1μg DNA donor plasmid using Lipofectamine 2000 as per the manufacturer's protocols. The DNA donors used for endogenous AAVS1 targeting were either a dsDNA donor (Fig. 2C) or a 90mer oligonucleotide. The former has flanking short homology arms and a SA-2A-puromycin-CaGGS-eGFP cassette to enrich for successfully targeted cells.

The targeting efficiency was assessed as follows. Cells were harvested 3 days after nucleofection and the genomic DNA of ~1 X 10^6 cells was extracted using prepGEM (ZyGEM). PCR was conducted to amplify the targeting region with genomic DNA derived from the cells and amplicons were deep sequenced by MiSeq Personal Sequencer (Illumina) with coverage >200,000 reads. The sequencing data was analyzed to estimate NHEJ efficiencies. The reference AAVS1 sequence analyzed is:

CACTTCAGGACAGCATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGCTGGGACCA
CCTTATATTCCCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGTACTTTTATCTGTCCCC
TCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTT
AGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCCCACCTCCTGTTAGGCAG
ATTCCTTATCTGGTGACACACCCCCATTTCCTGGA

The PCR primers for amplifying the targeting regions in the human genome are:

AAVS1-R CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTacaggaggtgggggttagac

AAVS1-F.1

ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATtatattcccagggccggtta

AAVS1-F.2

ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGtatattcccagggccggtta

AAVS1-F.3

ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCTAAtatattcccagggccggtta

AAVS1-F.4

ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTCAtatattcccagggccggtta

AAVS1-F.5

ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACTGTtatattcccagggccggtta

AAVS1-F.6

ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGGCtatattcccagggccggtta

AAVS1-F.7

ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCTGtatattcccagggccggtta

AAVS1-F.8

ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAAGTtatattcccagggccggtta

AAVS1-F.9

ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGATCtatattcccagggccggtta

AAVS1-F.10

ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGCTAtatattcccagggccggtta

AAVS1-F.11

ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAGCCtatattcccagggccggtta

AAVS1-F.12

ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACAAGtatattcccagggccggtta

To analyze the HR events using the DNA donor in Fig. 2C, the primers used were:

HR_AAVS1-F CTGCCGTCTCTCTCCTGAGT

HR_Puro-R GTGGGCTTGTACTCGGTCAT

EXAMPLE V

Bioinformatics approach for computing human exon CRISPR targets and methodology for their multiplexed synthesis

A set of gRNA gene sequences that maximally target specific locations in human exons but minimally target other locations in the genome were determined as follows. According to one aspect, maximally efficient targeting by a gRNA is achieved by 23nt sequences, the 5'-most 20nt of which exactly complement a desired location, while the three 3'-most bases must be of the form NGG. Additionally, the 5'-most nt must be a G to establish a pol-III transcription start site. However, according to (*2*), mispairing of the six 5'-most nt of a 20bp gRNA against its genomic target does not abrogate Cas9-mediated cleavage so long as the last 14nt pairs properly, but mispairing of the eight 5'-most nt along with pairing of the last 12 nt does, while the case of the

seven 5-most nt mispairs and 13 3' pairs was not tested. To be conservative regarding off-target effects, one condition was that the case of the seven 5'-most mispairs is, like the case of six, permissive of cleavage, so that pairing of the 3'-most 13nt is sufficient for cleavage. To identify CRISPR target sites within human exons that should be cleavable without off-target cuts, all 23bp sequences of the form 5'-GBBBB BBBBB BBBBB BBBBB NGG-3' (*form 1*) were examined, where the B's represent the bases at the exon location, for which no sequence of the form 5'-NNNNN NNBBB BBBBB BBBBB NGG-3' (*form 2*) existed at any other location in the human genome. Specifically, (i) a BED file of locations of coding regions of all RefSeq genes the GRCh37/hg19 human genome from the UCSC Genome Browser (*15-17*) was downloaded. Coding exon locations in this BED file comprised a set of 346089 mappings of RefSeq mRNA accessions to the hg19 genome. However, some RefSeq mRNA accessions mapped to multiple genomic locations (probable gene duplications), and many accessions mapped to subsets of the same set of exon locations (multiple isoforms of the same genes). To distinguish apparently duplicated gene instances and consolidate multiple references to the same genomic exon instance by multiple RefSeq isoform accessions, (ii) unique numerical suffixes to 705 RefSeq accession numbers that had multiple genomic locations were added, and (iii) the mergeBed function of BEDTools (*18*) (v2.16.2-zip-87e3926) was used to consolidate overlapping exon locations into merged exon regions. These steps reduced the initial set of 346089 RefSeq exon locations to 192783 distinct genomic regions. The hg19 sequence for all merged exon regions were downloaded using the UCSC Table Browser, adding 20bp of padding on each end. (iv) Using custom perl code, 1657793 instances of *form 1* were identified within this exonic sequence. (v) These sequences were then filtered for the existence of off-target occurrences of *form* 2: For each merged exon *form 1* target, the 3'-most 13bp specific (B) "core" sequences were extracted and, for each core generated the four 16bp sequences 5'-BBB BBBBB BBBBB NGG-3' (N = A, C, G, and T), and searched the entire hg19 genome for exact matches to these 6631172 sequences using Bowtie version 0.12.8 (*19*) using the parameters -l 16 -v 0 -k 2. Any exon target site for which there was more than a single match was rejected. Note that because any specific 13bp core sequence followed by the sequence NGG confers only 15bp of specificity, there should be on average ~5.6 matches to an extended core sequence in a random ~3Gb sequence (both strands). Therefore, most of the 1657793 initially identified targets were rejected; however 189864 sequences passed this filter. These comprise the set of CRISPR-targetable exonic locations in the human genome. The 189864 sequences target locations in 78028 merged exonic regions (~40.5% of the total of 192783 merged human exon regions) at a multiplicity of ~2.4 sites per targeted exonic region. To assess targeting at a gene level, RefSeq mRNA mappings were clustered so that any two RefSeq accessions (including the gene duplicates distinguished in (ii)) that overlap a merged exon region are counted as a single gene cluster, the 189864 exonic specific CRISPR sites

target 17104 out of 18872 gene clusters (~90.6% of all gene clusters) at a multiplicity of ~11.1 per targeted gene cluster. (Note that while these gene clusters collapse RefSeq mRNA accessions that represent multiple isoforms of a single transcribed gene into a single entity, they will also collapse overlapping distinct genes as well as genes with antisense transcripts.) At the level of original RefSeq accessions, the 189864 sequences targeted exonic regions in 30563 out of a total of 43726 (~69.9%) mapped RefSeq accessions (including distinguished gene duplicates) at a multiplicity of ~6.2 sites per targeted mapped RefSeq accession.

According to one aspect, the database can be refined by correlating performance with factors, such as base composition and secondary structure of both gRNAs and genomic targets (*20, 21*), and the epigenetic state of these targets in human cell lines for which this information is available (*22*).

EXAMPLE VI

Multiplex Synthesis

The target sequences were incorporated into a 200bp format that is compatible for multiplex synthesis on DNA arrays (*23, 24*). According to one aspect the method allows for targeted retrieval of a specific or pools of gRNA sequences from the DNA array based oligonucleotide pool and its rapid cloning into a common expression vector (Fig. 13A). Specifically, a 12k oligonucleotide pool from CustomArray Inc. was synthesized. Furthermore, gRNAs of choice from this library (Fig. 13B) were successfully retrieved. We observed an error rate of ~4 mutations per 1000bp of synthesized DNA.

EXAMPLE VII

RNA-guided genome editing requires both Cas9 and guide RNA for successful targeting

Using the GFP reporter assay described in Fig. 1B, all possible combinations of the repair DNA donor, Cas9 protein, and gRNA were tested for their ability to effect successful HR (in 293Ts). As shown in Figure 4, GFP+ cells were observed only when all the 3 components were present, validating that these CRISPR components are essential for RNA-guided genome editing. Data is mean +/- SEM (N=3).

EXAMPLE VIII

Analysis of gRNA and Cas9 mediated genome editing

The CRISPR mediated genome editing process was examined using either (A) a GFP reporter assay as described earlier results of which are shown in Figure 5A, and (B) deep

sequencing of the targeted loci (in 293Ts), results of which are shown in Figure 5B. As comparison, a D10A mutant for Cas9 was tested that has been shown in earlier reports to function as a nickase in *in vitro* assays. As shown in Figure 5, both Cas9 and Cas9D10A can effect successful HR at nearly similar rates. Deep sequencing however confirms that while Cas9 shows robust NHEJ at the targeted loci, the D10A mutant has significantly diminished NHEJ rates (as would be expected from its putative ability to only nick DNA). Also, consistent with the known biochemistry of the Cas9 protein, NHEJ data confirms that most base-pair deletions or insertions occurred near the 3' end of the target sequence: the peak is ~3-4 bases upstream of the PAM site, with a median deletion frequency of ~9-10bp. Data is mean +/- SEM (N=3).

EXAMPLE IX

RNA-guided genome editing is target sequence specific

Similar to the GFP reporter assay described in Fig. 1B, 3 293T stable lines each bearing a distinct GFP reporter construct were developed. These are distinguished by the sequence of the AAVS1 fragment insert (as indicated in the Figure 6). One line harbored the wild-type fragment while the two other lines were mutated at 6bp (highlighted in red). Each of the lines was then targeted by one of the following 4 reagents: a GFP-ZFN pair that can target all cell types since its targeted sequence was in the flanking GFP fragments and hence present in along cell lines; a AAVS1 TALEN that could potentially target only the wt-AAVS1 fragment since the mutations in the other two lines should render the left TALEN unable to bind their sites; the T1 gRNA which can also potentially target only the wt-AAVS1 fragment, since its target site is also disrupted in the two mutant lines; and finally the T2 gRNA which should be able to target all 3 cell lines since, unlike the T1 gRNA, its target site is unaltered among the 3 lines. ZFN modified all 3 cell types, the AAVS1 TALENs and the T1 gRNA only targeted the wt-AAVS1 cell type, and the T2 gRNA successfully targets all 3 cell types. These results together confirm that the guide RNA mediated editing is target sequence specific. Data is mean +/- SEM (N=3).

EXAMPLE X

Guide RNAs targeted to the GFP sequence enable robust genome editing

In addition to the 2 gRNAs targeting the AAVS1 insert, two additional gRNAs targeting the flanking GFP sequences of the reporter described in Fig. 1B (in 293Ts) were tested. As shown in Figure 7, these gRNAs were also able to effect robust HR at this engineered locus. Data is mean +/- SEM (N=3).

EXAMPLE XI

RNA-guided genome editing is target sequence specific, and demonstrates similar targeting efficiencies as ZFNs or TALENs

Similar to the GFP reporter assay described in Fig. 1B, two 293T stable lines each bearing a distinct GFP reporter construct were developed. These are distinguished by the sequence of the fragment insert (as indicated in Figure 8). One line harbored a 58bp fragment from the DNMT3a gene while the other line bore a homologous 58bp fragment from the DNMT3b gene. The sequence differences are highlighted in red. Each of the lines was then targeted by one of the following 6 reagents: a GFP-ZFN pair that can target all cell types since its targeted sequence was in the flanking GFP fragments and hence present in along cell lines; a pair of TALENs that potentially target either DNMT3a or DNMT3b fragments; a pair of gRNAs that can potentially target only the DNMT3a fragment; and finally a gRNA that should potentially only target the DNMT3b fragment. As indicated in Figure 8, the ZFN modified all 3 cell types, and the TALENs and gRNAs only their respective targets. Furthermore the efficiencies of targeting were comparable across the 6 targeting reagents. These results together confirm that RNA-guided editing is target sequence specific and demonstrates similar targeting efficiencies as ZFNs or TALENs. Data is mean +/- SEM (N=3).

EXAMPLE XII

RNA-guided NHEJ in human iPS cells

Human iPS cells (PGP1) were nucleofected with constructs indicated in the left panel of Figure 9. 4 days after nucleofection, NHEJ rate was measured by assessing genomic deletion and insertion rate at double-strand breaks (DSBs) by deep sequencing. *Panel 1*: Deletion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. The deletion incidence at each nucleotide position was plotted in black lines and the deletion rate as the percentage of reads carrying deletions was calculated. *Panel 2*: Insertion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. The incidence of insertion at the genomic location where the first insertion junction was detected was plotted in black lines and the insertion rate as the percentage of reads carrying insertions was calculated. *Panel 3*: Deletion size distribution. The frequencies of different size deletions among the whole NHEJ population was plotted. *Panel 4*: insertion size distribution. The frequencies of different sizes insertions among the whole NHEJ population was plotted. iPS targeting by both gRNAs is efficient (2-4%), sequence specific (as shown by the shift in position of the NHEJ deletion distributions), and reaffirming the

results of Figure 4, the NGS-based analysis also shows that both the Cas9 protein and the gRNA are essential for NHEJ events at the target locus.

EXAMPLE XIII

RNA-guided NHEJ in K562 cells

K562 cells were nucleated with constructs indicated in the left panel of Figure 10. 4 days after nucleofection, NHEJ rate was measured by assessing genomic deletion and insertion rate at DSBs by deep sequencing. *Panel 1*: Deletion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. The deletion incidence at each nucleotide position was plotted in black lines and the deletion rate as the percentage of reads carrying deletions was calculated. *Panel 2*: Insertion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. The incidence of insertion at the genomic location where the first insertion junction was detected was plotted in black lines and the insertion rate as the percentage of reads carrying insertions was calculated. *Panel 3*: Deletion size distribution. The frequencies of different size deletions among the whole NHEJ population was plotted. *Panel 4*: insertion size distribution. The frequencies of different sizes insertions among the whole NHEJ population was plotted. K562 targeting by both gRNAs is efficient (13-38%) and sequence specific (as shown by the shift in position of the NHEJ deletion distributions). Importantly, as evidenced by the peaks in the histogram of observed frequencies of deletion sizes, simultaneous introduction of both T1 and T2 guide RNAs resulted in high efficiency deletion of the intervening 19bp fragment, demonstrating that multiplexed editing of genomic loci is also feasible using this approach.

EXAMPLE XIV

RNA-guided NHEJ in 293T cells

293T cells were transfected with constructs indicated in the left panel of Figure 11. 4 days after nucleofection, NHEJ rate was measured by assessing genomic deletion and insertion rate at DSBs by deep sequencing. *Panel 1*: Deletion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. The deletion incidence at each nucleotide position was plotted in black lines and the deletion rate as the percentage of reads carrying deletions was calculated. *Panel 2*: Insertion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. The incidence of insertion at the genomic location where the first insertion junction was detected was plotted in black lines and the insertion rate as the percentage of reads carrying

insertions was calculated. *Panel 3*: Deletion size distribution. The frequencies of different size deletions among the whole NHEJ population was plotted. *Panel 4*: insertion size distribution. The frequencies of different sizes insertions among the whole NHEJ population was plotted. 293T targeting by both gRNAs is efficient (10-24%) and sequence specific (as shown by the shift in position of the NHEJ deletion distributions).

EXAMPLE XV

HR at the endogenous AAVS1 locus using either a dsDNA donor or a short oligonucleotide donor

As shown in Figure 12A, PCR screen (with reference to Figure 2C) confirmed that 21/24 randomly picked 293T clones were successfully targeted. As shown in Figure 12B, similar PCR screen confirmed 3/7 randomly picked PGP1-iPS clones were also successfully targeted. As shown in Figure 12C, short 90mer oligos could also effect robust targeting at the endogenous AAVS1 locus (shown here for K562 cells).

EXAMPLE XVI

Methodology for multiplex synthesis, retrieval and U6 expression vector cloning of guide RNAs targeting genes in the human genome

A resource of about 190k bioinformatically computed unique gRNA sites targeting ~40.5% of all exons of genes in the human genome was generated. As shown in Figure 13A, the gRNA target sites were incorporated into a 200bp format that is compatible for multiplex synthesis on DNA arrays. Specifically, the design allows for (i) targeted retrieval of a specific or pools of gRNA targets from the DNA array oligonucleotide pool (through 3 sequential rounds of nested PCR as indicated in the figure schematic); and (ii) rapid cloning into a common expression vector which upon linearization using an AflII site serves as a recipient for Gibson assembly mediated incorporation of the gRNA insert fragment. As shown in Figure 13B, the method was used to accomplish targeted retrieval of 10 unique gRNAs from a 12k oligonucleotide pool synthesized by CustomArray Inc.

EXAMPLE XVII

CRISPR Mediated RNA-Guided Transcriptional Activation

The CRISPR-Cas system has an adaptive immune defense system in bacteria and functions to 'cleave' invading nucleic acids. According to one aspect, the CRISPR-CAS system is

engineered to function in human cells, and to 'cleave' genomic DNA. This is achieved by a short guide RNA directing a Cas9 protein (which has nuclease function) to a target sequence complementary to the spacer in the guide RNA. The ability to 'cleave' DNA enables a host of applications related to genome editing, and also targeted genome regulation. Towards this, the Cas9 protein was mutated to make it nuclease-null by introducing mutations that are predicted to abrogate coupling to Mg2+ (known to be important for the nuclease functions of the RuvC-like and HNH-like domains): specifically, combinations of D10A, D839A, H840A and N863A mutations were introduced. The thus generated Cas9 nuclease-null protein (as confirmed by its ability to not cut DNA by sequencing analysis) and hereafter referred to as Cas9R-H-, was then coupled to a transcriptional activation domain, here VP64, enabling the CRISPR-cas system to function as a RNA guided transcription factor (see Figure 14). The Cas9R-H-+VP64 fusion enables RNA-guided transcriptional activation at the two reporters shown. Specifically, both FACS analysis and immunofluorescence imaging demonstrates that the protein enables gRNA sequence specific targeting of the corresponding reporters, and furthermore, the resulting transcription activation as assayed by expression of a dTomato fluorescent protein was at levels similar to those induced by a convention TALE-VP64 fusion protein.

EXAMPLE XVIII

gRNA Sequence Flexibility and Applications Thereof

Flexibility of the gRNA scaffold sequence to designer sequence insertions was determined by systematically assaying for a range of the random sequence insertions on the 5', middle and 3' portions of the gRNA: specifically, 1bp, 5bp, 10bp, 20bp, and 40bp inserts were made in the gRNA sequence at the 5', middle, and 3' ends of the gRNA (the exact positions of the insertion are highlighted in 'red' in Figure 15). This gRNA was then tested for functionality by its ability to induce HR in a GFP reporter assay (as described herein). It is evident that gRNAs are flexible to sequence insertions on the 5' and 3' ends (as measured by retained HR inducing activity). Accordingly, aspects of the present disclosure are directed to tagging of small-molecule responsive RNA aptamers that may trigger onset of gRNA activity, or gRNA visualization. Additionally, aspects of the present disclosure are directed to tethering of ssDNA donors to gRNAs via hybridization, thus enabling coupling of genomic target cutting and immediate physical localization of repair template which can promote homologous recombination rates over error-prone non-homologous end-joining.

The following references identified in the Examples section by number are hereby incorporated by reference in their entireties for all purposes.

References

1. K. S. Makarova *et al.*, Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology* 9, 467 (Jun, 2011).
2. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816 (Aug 17, 2012).
3. P. Horvath, R. Barrangou, CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167 (Jan 8, 2010).
4. H. Deveau *et al.*, Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology* 190, 1390 (Feb, 2008).
5. J. R. van der Ploeg, Analysis of CRISPR in Streptococcus mutans suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* 155, 1966 (Jun, 2009).
6. M. Rho, Y. W. Wu, H. Tang, T. G. Doak, Y. Ye, Diverse CRISPRs evolving in human microbiomes. *PLoS genetics* 8, e1002441 (2012).
7. D. T. Pride *et al.*, Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome research* 21, 126 (Jan, 2011).
8. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579 (Sep 25, 2012).
9. R. Sapranauskas *et al.*, The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic acids research* 39, 9275 (Nov, 2011).
10. J. E. Garneau *et al.*, The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67 (Nov 4, 2010).
11. K. M. Esvelt, J. C. Carlson, D. R. Liu, A system for the continuous directed evolution of biomolecules. *Nature* 472, 499 (Apr 28, 2011).
12. R. Barrangou, P. Horvath, CRISPR: new horizons in phage resistance and strain identification. *Annual review of food science and technology* 3, 143 (2012).
13. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331 (Feb 16, 2012).
14. N. E. Sanjana *et al.*, A transcription activator-like effector toolbox for genome engineering. *Nature protocols* 7, 171 (Jan, 2012).
15. W. J. Kent *et al.*, The human genome browser at UCSC. *Genome Res* 12, 996 (Jun, 2002).

16. T. R. Dreszer *et al.*, The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res* 40, D918 (Jan, 2012).

17. D. Karolchik *et al.*, The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32, D493 (Jan 1, 2004).

18. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841 (Mar 15, 2010).

19. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).

20. R. Lorenz *et al.*, ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB* 6, 26 (2011).

21. D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology* 288, 911 (May 21, 1999).

22. R. E. Thurman *et al.*, The accessible chromatin landscape of the human genome. *Nature* 489, 75 (Sep 6, 2012).

23. S. Kosuri *et al.*, Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nature biotechnology* 28, 1295 (Dec, 2010).

24. Q. Xu, M. R. Schlabach, G. J. Hannon, S. J. Elledge, Design of 240,000 orthogonal 25mer DNA barcode probes. *Proceedings of the National Academy of Sciences of the United States of America* 106, 2289 (Feb 17, 2009).

Claims

1. An *in vitro* or *ex vivo* method of altering a eukaryotic cell comprising

providing to the eukaryotic cell guide RNA sequence GN_{19} GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGUGCUUUU,

wherein the GN_{19} sequence is complementary to a target nucleic acid sequence and binds to the target nucleic acid sequence and

providing to the eukaryotic cell a Cas 9 protein that interacts with the guide RNA sequence, and

wherein the method does not comprise a process for modifying the germ line genetic identity of human beings and wherein the cell is not a human embryo.

2. The method of claim 1 wherein the Cas 9 protein is a Cas9 enzyme and the target nucleic acid sequence is cleaved.

3. The method of any one of claims 1 to 2 wherein the guide RNA sequence is provided to the cell by introducing to the cell a first nucleic acid encoding the guide RNA sequence,

wherein the Cas 9 protein is provided to the cell by introducing to the cell a second nucleic acid encoding the Cas9 protein, and

wherein the cell produces the guide RNA and the Cas9 protein.

4. The method of claim 3 wherein the first nucleic acid is introduced into the cell using a viral delivery method.

5. The method of claim 3 wherein the first and/or second nucleic acid are introduced into the cell by transfection.

6. The method of claim 3 wherein the first nucleic acid is introduced into the cell using an adeno-associated virus.

7. The method of any one of claims 1 to 6 wherein the Cas9 protein is a Cas9 enzyme, a Cas9 nickase, a nuclease null Cas9, a modified Cas9 or a homolog of Cas9.

8. The method of any one of claims 1 to 7 wherein the eukaryotic cell is a yeast cell, a plant cell or a mammalian cell.

9. The method of any one of claims 1 to 8 wherein the cell is a human cell.

10. The method of any one of claims 1 to 9 wherein a plurality of guide RNA sequences complementary to different target nucleic acid sequences are provided to the eukaryotic cell, wherein the guide RNA sequence binds to the different target nucleic acid sequence,
and the Cas 9 protein interacts with the plurality of guide RNA sequences.

11. The method of any one of claims 1 to 10 wherein the guide RNA is a crRNA-tracrRNA fusion.

12. The method of any one of claims 1 to 11 wherein the Cas9 protein is human codon-optimized.

13. The method of any one of claims 1 to 12 wherein guide RNA sequences complementary to target nucleic acid sequences in a locus are provided to the cell, wherein the guide RNA sequences bind to the target nucleic acid sequences and the Cas9 protein is a Cas9 enzyme that cleaves the target nucleic acid sequences and an intervening nucleic acid sequence is deleted from the locus.

14. The method of any one of claims 1 to 13 wherein the Cas9 protein is a Cas9 enzyme that cleaves the target nucleic acid sequence and a donor nucleic acid sequence provided to the eukaryotic cell is inserted into the target nucleic acid sequence.

15. The method of claim 1 wherein the Cas9 protein is nuclease null and interacts with the guide RNA sequence and binds to the target nucleic acid sequence,
wherein the Cas 9 protein has a FokI nuclease domain attached thereto,
wherein a target nucleic acid sequence is cleaved upon FokI nuclease domain dimerization of adjacent guide RNA sequence and Cas9 protein co-localization complexes.

16. The method of claim 1 wherein the Cas9 protein is nuclease null and includes a transcriptional activation or repression domain and wherein expression of a target gene is modulated.

17. The method of claim 1 wherein the Cas9 protein is nuclease null and includes a fluorescent protein and wherein the fluorescent protein is visualized.

18. The method of claim 1 wherein the Cas9 protein is nuclease null and includes a protein bound organic fluorophore, a nucleic acid bound organic fluorophore, a quantum dot, a molecular beacon, an echo probe or a multivalent ligand-binding protein domain.

19. An RNA-guided genome editing system comprising:
guide RNA sequence GN_{19} GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGA

AAAAGUGGCACCGAGUCGGUGCUUUU, wherein the GN_{19} sequence is complementary to a target nucleic acid sequence and binds to the target nucleic acid sequence; or a first nucleic acid sequence encoding the guide RNA sequence; and

a Cas9 protein that interacts with the guide RNA sequence, or a second nucleic acid sequence encoding the Cas9 protein, wherein the Cas9 protein includes a C-terminus SV40 nuclear localization signal.

20. The RNA-guided genome editing system of claim 19 for use in gene therapy.

21. An eukaryotic cell containing the RNA-guided genome editing system of claim 19, wherein the cell is not a human germ cell and not a human embryo.