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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 December 2018 (20.12.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/229521 A1

(51) International Patent Classification:
C12N 15/90 (2006.01) *C12N 15/63* (2006.01)
C12N 15/10 (2006.01)

(21) International Application Number:
PCT/IB2017/000967

(22) International Filing Date:
16 June 2017 (16.06.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/350,941 16 June 2017 (16.06.2017) US

(71) Applicant: **OSLO UNIVERSITETSSYKEHUS HF** [NO/NO]; Postboks 450, Nydalen, 0424 Oslo (NO).

(72) Inventors: **ROBERTSON, Adam, Brian**; c/o Oslo Universitetssykehus HF Postboks, 450, Nydalen, 0424 Oslo (NO). **ROBERTSON, Julia**; c/o Oslo Universitetssykehus HF Postboks, 450, Nydalen, 0424 Oslo (NO). **REIFSCHNEIDER, Anika**; c/o Oslo Universitetssykehus HF Postboks, 450, Nydalen, 0424 Osol (NO). **KLUNGLAND, Arne**; c/o Oslo Universitetssykehus HF Postboks, 450, Nydalen, 0424 Oslo (NO).

(74) Agent: **JONES, Mitchell J.**; Casimir Jones, S.C., 2275 Deming Way, Ste 310, Middleton, WI 53562 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: IMPROVED GENE EDITING

(57) Abstract: The present invention generally relates to systems, methods and compositions used for the control of gene expression involving sequence targeting, such as genome perturbation or gene-editing, that may use vector systems related to recombinases and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and components thereof.

WO 2018/229521 A1

IMPROVED GENE EDITING

FIELD OF THE INVENTION

The present invention generally relates to systems, methods and compositions used for the control of gene expression involving sequence targeting, such as genome perturbation or gene-editing, that may use vector systems related to recombinases and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and components thereof.

BACKGROUND OF THE INVENTION

Genome editing is a powerful technology that allows for the specific and often precise addition or removal of genetic material. Genome editing is initiated by making double stranded DNA breaks in the target cell. These double stranded DNA breaks can be created by several methods – including; meganucleases, Zinc-Finger Nucleases, TALE-nucleases, and/or the CRISPR/Cas9 restriction modification system. Each of these systems creates a dsDNA break at a user designated genomic location. After the creation of the dsDNA break, the cellular machinery acts quickly to repair this dsDNA using either by the non-homologus end joining (NHEJ) pathway or by homologus recombination (HDR). While, the NHEJ pathway efficiently repairs this break, repair is frequently imperfect resulting in insertions and deletions. If these insertions and deletions created by NHEJ repair occur within open reading frames, the most common result is a frame-shift mutation. This frame shift often results in the inactivation of that particular gene. Repair of the dsDNA break by HDR pathway not only can result in precise repair but also allows for the introduction of experimentally designed genomic elements. The correction of many diseases, successful gene therapy, can be achieved by forcing the cell to correct the dsDNA break using HDR. Unfortunately for gene therapy researchers, clinicians, and patients, most human cells strongly prefer to correct dsDNA breaks the error-prone NHEJ pathway as opposed to the more precise HDR pathway. Using endogenous cellular machinery 95% of dsDNA breaks are repaired using NHEJ, while only 5% of dsDNA breaks are repaired using HDR. This statistic represents the best-case scenario; many cell types lack HDR machinery altogether resulting in no repair using the precise HDR pathway. For precise gene therapy to be successful, a cells ability to use the HDR pathway must be improved.

SUMMARY OF THE INVENTION

The present inventors have developed a method that significantly improves the cells ability to utilize the HDR pathway in conjunctions with techniques such as CAS-CRISPR gene editing. The data presented herein demonstrates that the system can skew the cell's preference of dsDNA repair pathways away from the error prone NHEJ pathway and towards the HDR pathway – between 5 and 125 fold improvement. This improvement in HDR is achieved by the addition of viral and/or bacterial recombinases – UvsX or RecA (from either *E. coli* or *S. Pneumoniae*). The expression of these recombinases in combination with a site-specific nuclease and a homologus recombination (HR) substrate results in improved HDR and allows for the HDR in cells that have no HDR machinery.

This improvement in HDR mediated dsDNA repair has substantial implications not only in research applications but also in gene therapy. This technology will have immediate applicability in basic research laboratories. In the near future, this technology can be used to reactivate enzymes that are deficient in many metabolic diseases such as Gauchers disease, Fabry disease, and Pompe disease. These diseases are currently treated by enzyme replacement therapies (ERT). ERTs are often extremely expensive (200,000 USD per year) and require regular injections. Additionally, these treatments can result in undesired immune responses. As an example, Pompe disease (Glycogen storage disease type II) results from a defective copy of acid alpha-glucosidase (GAA). Most commonly this defective copy of GAA results from a single thymine to guanosine transversion. As the injection of recombinant GAA significantly improves patients' lives, using our technology to faithfully correct this mutation in a small fraction of a Pompe disease patient's cells would allow the secretion of active GAA. This genetic change would not only result in a substantial cost savings but also a significant improvement in Pompe disease patient's prognosis. Similar genetic therapies could be introduced using our technology to treat patients that are currently dependent on ERTs.

These diseases are debilitating for the patient and the treatments for these disorders are costly. Even though ERTs work well to alleviate suffering, ERTs are so costly that some governments and health insurance providers refuse to provide them. Our gene replacement technology, once fully developed, will produce similar results as ERTs with no need for life long infusions and at a fraction of the cost. While our work to improve HDR is at an early stage, we have successfully shown that our technology works many fold better than current technologies; therefore, now is the time to invest resources into our project; a project that,

when successful, has the immediate propensity to improve the lives of hundreds of thousands of people.

Accordingly, in one aspect, the invention provides a method for altering or modifying expression of one or more gene products. The said method may comprise introducing into a eukaryotic cell containing and expressing DNA molecules encoding the one or more gene products an engineered, non-naturally occurring vector system comprising one or more vectors comprising: a) a first regulatory element operably linked to a nucleic acid sequence encoding one or more Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)—CRISPR associated (Cas) system guide RNAs that hybridize with target sequences in genomic loci of the DNA molecules encoding the one or more gene products, b) a second regulatory element operably linked to a nucleic acid sequence encoding a Cas9 protein (e.g., Type-II Cas9 protein or a Cas9 nickase), and c) a third regulatory element operably linked to a nucleic acid sequence encoding a recombinase, wherein components (a), (b) and (c) are located on same or different vectors of the system, whereby the guide RNAs target the genomic loci of the DNA molecules encoding the one or more gene products and the Cas9 protein cleaves the genomic loci of the DNA molecules encoding the one or more gene products, whereby expression of the one or more gene products is altered; and, wherein the Cas9 protein and the guide RNAs do not naturally occur together. The invention comprehends the expression of two or more gene products being altered and the vectors of the system further comprising one or more nuclear localization signal(s) (NLS(s)). The invention comprehends the guide RNAs comprising a guide sequence fused to a tracr sequence. The invention further comprehends the Cas9 protein being codon optimized for expression in the eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell or a human cell. In a further embodiment of the invention, the expression of one or more of the gene products is decreased. In aspects of the invention cleaving the genomic loci of the DNA molecule encoding the gene product encompasses cleaving either one or both strands of the DNA duplex.

In one aspect, the invention provides an engineered, programmable, non-naturally occurring CRISPR-Cas system comprising a Cas9 protein, a recombinase and one or more guide RNAs that target the genomic loci of DNA molecules encoding one or more gene products in a eukaryotic cell and the Cas9 protein cleaves the genomic loci of the DNA molecules encoding the one or more gene products, whereby expression of the one or more gene products is altered; and, wherein the Cas9 protein and the guide RNAs do not naturally occur together. The invention comprehends the expression of two or more gene products

being altered and the CRISPR-Cas system further comprising one or more NLS(s). The invention comprehends the guide RNAs comprising a guide sequence fused to a tracr sequence. The invention further comprehends the Cas9 protein being codon optimized for expression in the eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell or a human cell. In aspects of the invention cleaving the genomic loci of the DNA molecule encoding the gene product encompasses cleaving either one or both strands of the DNA duplex.

In another aspect, the invention provides an engineered, non-naturally occurring vector system comprising one or more vectors comprising a) a first regulatory element operably linked to a nucleic acid sequence encoding one or more Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)—CRISPR associated (Cas) system guide RNAs that hybridize with target sequences in genomic loci of the DNA molecules encoding the one or more gene products, b) a second regulatory element operably linked to a nucleic acid sequence encoding a Cas9 protein (e.g., Type-II Cas9 protein or a Cas9 nickase), and c) a third regulatory element operably linked to a nucleic acid sequence encoding a recombinase, wherein components (a), (b) and (c) are located on same or different vectors of the system, whereby the guide RNAs target the genomic loci of the DNA molecules encoding the one or more gene products in a eukaryotic cell and the Cas9 protein cleaves the genomic loci of the DNA molecules encoding the one or more gene products, whereby expression of the one or more gene products is altered; and, wherein the Cas9 protein and the guide RNAs do not naturally occur together. The invention comprehends the expression of two or more gene products being altered and the vectors of the system further comprising one or more nuclear localization signal(s) (NLS(s)). The invention comprehends the guide RNAs comprising a guide sequence fused to a tracr sequence. The invention further comprehends the Cas9 protein being codon optimized for expression in the eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell or a human cell. In a further embodiment of the invention, the expression of one or more of the gene products is decreased. In aspects of the invention cleaving the genomic loci of the DNA molecule encoding the gene product encompasses cleaving either one or both strands of the DNA duplex.

In one aspect, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracr mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic

cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence; (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence; and (c) a third regulatory element operably linked to a nucleic acid sequence encoding a recombinase, wherein components (a), (b) and (c) are located on the same or different vectors of the system. In some embodiments, component (a) further comprises the tracr sequence downstream of the tracr mate sequence under the control of the first regulatory element. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the system comprises the tracr sequence under the control of a third regulatory element, such as a polymerase III promoter. In some embodiments, the tracr sequence exhibits at least 50%, 60%, 70%, 80%, 90%, 95%, or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned.

Determining optimal alignment is within the purview of one of skill in the art. For example, there are publically and commercially available alignment algorithms and programs such as, but not limited to, ClustalW, Smith-Waterman in matlab, Bowtie, Geneious, Biopython and SeqMan. In some embodiments, the CRISPR complex comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR complex in a detectable amount in the nucleus of a eukaryotic cell. Without wishing to be bound by theory, it is believed that a nuclear localization sequence is not necessary for CRISPR complex activity in eukaryotes, but that including such sequences enhances activity of the system, especially as to targeting nucleic acid molecules in the nucleus. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is *S. pneumoniae, S. pyogenes*, or *S. thermophilus* Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the recombinase is a bacterial recombinase. In some embodiments, the recombinase does not occur naturally in the cell type transformed or transduced with the vector system. In some embodiments, the

recombinase is selected from the group consisting of Rad51, RecA recombinase and UvsX recombinase or any protein that contains a RecA or RadA domain.

In some embodiments, the first regulatory element is a cytomegalovirus promoter (CMV), polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 15, 16, 17, 18, 19, 20, 25 nucleotides, or between 10-30, or between 15-25, or between 15-20 nucleotides in length. In general, and throughout this specification, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked.

Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in

vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g. transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g. liver, pancreas), or particular cell types (e.g. lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g. 1, 2, 3, 4, 5, or more pol I promoters), one or more pol II promoters (e.g. 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g. 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al, Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5′ segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.).

Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

In some alternative embodiments of the above described methods and systems, the Cas9 enzyme and/or recombinase may be introduced as proteins rather than as encoded by nucleic acid vectors. For example, it is contemplated to introduce these molecules as protein and nucleic acid complexes, e.g., Cas9-RNA complexes and DNA-RecA complexes. In some embodiments, these complexes are assembled in vitro prior to transfection, e.g., by electroporation.

In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The organism also may be a plant. Further, the organism may be a fungus.

In one aspect, the invention provides a kit comprising one or more of the components described herein. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracr mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence. In some embodiments, the kit comprises components (a) and (b) located on the same or different vectors of the system. In some embodiments, component (a) further comprises the tracr sequence downstream of the tracr mate sequence under the control of the first regulatory element. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the system further comprises a third regulatory element, such as a polymerase III promoter, operably linked to said tracr sequence. In some embodiments, the

tracr sequence exhibits at least 50%, 60%, 70%, 80%, 90%, 95%, or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. In some embodiments, the CRISPR enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is *S. pneumoniae, S. pyogenes* or *S. thermophilus* Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 15, 16, 17, 18, 19, 20, 25 nucleotides, or between 10-30, or between 15-25, or between 15-20 nucleotides in length.

In one aspect, the invention provides a method of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cell, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence. In some embodiments, said vectors are delivered to the eukaryotic cell in a

subject. In some embodiments, said modifying takes place in said eukaryotic cell in a cell culture. In some embodiments, the method further comprises isolating said eukaryotic cell from a subject prior to said modifying. In some embodiments, the method further comprises returning said eukaryotic cell and/or cells derived therefrom to said subject.

In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cells, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence.

In one aspect, the invention provides a method of generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) introducing one or more vectors into a eukaryotic cell, wherein the one or more vectors drive expression of one or more of: a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, and a tracr sequence; and (b) allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said disease gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized to the tracr sequence, thereby generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expression from a gene comprising the target sequence.

In one aspect, the invention provides a method for developing a biologically active agent that modulates a cell signaling event associated with a disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) contacting a test compound with a model cell of any one of the described embodiments; and (b) detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with said mutation in said disease gene, thereby developing said biologically active agent that modulates said cell signaling event associated with said disease gene.

In one aspect, the invention provides a recombinant polynucleotide comprising a guide sequence upstream of a tracr mate sequence, wherein the guide sequence when expressed directs sequence-specific binding of a CRISPR complex to a corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

In one aspect the invention provides for a method of selecting one or more cell(s) by introducing one or more mutations in a gene in the one or more cell (s), the method comprising: introducing one or more vectors into the cell (s), wherein the one or more vectors drive expression of one or more of: a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, a tracr sequence, and an editing template; wherein the editing template comprises the one or more mutations that abolish CRISPR enzyme cleavage; allowing homologous recombination of the editing template with the target polynucleotide in the cell(s) to be selected; allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized to the tracr sequence, wherein binding of the CRISPR complex to the target polynucleotide induces cell death, thereby allowing one or more cell(s) in which one or more mutations have been introduced to be selected. In a preferred embodiment, the CRISPR enzyme is Cas9. In another preferred embodiment of the invention the cell to be selected may be a eukaryotic cell. Aspects of the invention allow for selection of specific cells without requiring a selection marker or a two-step process that may include a counter-selection system Accordingly, it is an object of the invention not to encompass within the invention any previously known product, process of making the product, or method of using the product such that Applicants reserve the right and hereby disclose a disclaimer of any

previously known product, process, or method. It is further noted that the invention does not intend to encompass within the scope of the invention any product, process, or making of the product or method of using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. §112, first paragraph) or the EPO (Article 83 of the EPC), such that Applicants reserve the right and hereby disclose a disclaimer of any previously described product, process of making the product, or method of using the product.

In some embodiments, the present invention provides methods of altering expression of at least one gene product and/or genome editing comprising: introducing into a cell having a genome a nucleic acid filament comprising a single stranded nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into the genome, the single stranded nucleic acid molecule having bound thereto a multimeric recombinase complex and wherein the nucleic of interest comprises 5' and 3' flanking regions that are homologous to a genomic target sequence encoding the gene product, the 5' and 3' flanking sequences flank an insert sequence that is different from the genomic target sequence, and wherein the nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into the genome and the multimeric recombinase complex do not naturally occur together; whereby the nucleic acid sequence of interest is inserted by homologous recombination into the genome to effect altered expression of at the at least one gene product and/or genome editing of the gene encoding the gene product of interest. In some embodiments, the flanking sequences are substantially homologous to a target region in the genome. In some embodiments, the flanking regions are at least 95%, 96%, 97%, 98%, 99% or 100% homologous to a target region in the genome.

In some embodiments, the nucleic acid filament is synthesized in vitro by incubating the single stranded nucleic acid of interest with a recombinase so that a multimeric recombinase complex is formed on the single stranded nucleic acid of interest. In some embodiments, the methods further comprise incubating the single stranded nucleic acid of interest with a nucleotide. In some embodiments, the nucleotide is selected from the group consisting of a nucleotide triphosphate or analog. In some embodiments, the nucleotide triphosphate or analog is selected from the group consisting of adenosine triphosphate, adenosine monophosphate, adenosine diphosphate, adenosine triphosphate-γS, adenosine monophosphate-PNP, and adenosine diphosphate-AlF_4.

In some embodiments, the recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase. In some

embodiments, the recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

In some embodiments, the 5' and 3' flanking sequences are greater than 50 bases in length. In some embodiments, the 5' and 3' flanking sequences are greater than 100 bases in length. In some embodiments, the 5' and 3' flanking sequences are greater than 200 bases in length. In some embodiments, the 5' and 3' flanking sequences are greater than 500 bases in length. In some embodiments, the 5' and 3' flanking sequences are greater than 1000 bases in length. In some embodiments, the 5' and 3' flanking sequences are from about 20 to about 1000 bases in length. In some embodiments, the 5' and 3' flanking sequences are from about 100 to about 1000 bases in length.

In some embodiments, the single stranded nucleic acid is single stranded DNA. In some embodiments, the methods further comprise introducing a break in the targeted sequence in a gene encoding the gene product in the genome of the cell. In some embodiments, the break is a double stranded break or a single stranded break. In some embodiments, the break is introduced by an enzyme selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases, and a Type I restriction endonuclease, Type II restriction endonuclease, Type III, restriction endonuclease, Type IV restriction endonuclease or nickase.

In some embodiments, the methods further comprise introducing into the cell a CRISPR -Cas system guide RNA that hybridizes with the target sequence. In some embodiments, the CRISPR -Cas system comprises a trans-activating cr (tracr) sequence. In some embodiments, the guide RNAs comprise a guide sequence fused to a tracr sequence.

In some embodiments, the filament is introduced into the cell by electroporation. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian cell. In some embodiments, the mammalian cell is a human cell. In some embodiments, the expression of one or more gene products is increased. In some embodiments, the expression of one or more gene products is decreased. In some embodiments, the expression of two or more gene products is altered.

In some embodiments, the present invention provides an engineered, non-naturally occurring system for altering expression of a gene product and/or genome editing comprising: a nucleic acid filament comprising a single stranded nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into the genome, the single stranded nucleic acid molecule having bound thereto a multimeric recombinase complex and wherein the nucleic of interest comprises 5' and 3' flanking regions that are homologous to a genomic target

sequence encoding the gene product, the 5' and 3' flanking sequences flank an insert sequence that is different from the genomic target sequence, and wherein the single stranded nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into the genome and the multimeric recombinase complex do not naturally occur together; whereby when introduced into a cell having a genome the nucleic acid sequence of interest is inserted by homologous recombination into the genome to effect altered expression of at the at least one gene product and/or genome editing of the gene encoding the gene product of interest.

In some embodiments, the nucleic acid filament is synthesized in vitro by incubating the single stranded nucleic acid of interest with a recombinase so that a multimeric recombinase complex is formed on the single stranded nucleic acid of interest. In some embodiments, the single stranded nucleic acid of interest is further comprises a nucleotide. In some embodiments, the nucleotide is selected from the group consisting of a nucleotide triphosphate or analog. In some embodiments, the nucleotide triphosphate or analog is selected from the group consisting of adenosine triphosphate, adenosine monophosphate, adenosine diphosphate, adenosine triphosphate-γS, adenosine monophosphate-PNP, and adenosine diphosphate-AlF_4.

In some embodiments, the recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase. In some embodiments, the recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

In some embodiments, the 5' and 3' flanking sequences are greater than 50 bases in length. In some embodiments, the 5' and 3' flanking sequences are greater than 100 bases in length. In some embodiments, the 5' and 3' flanking sequences are greater than 200 bases in length. In some embodiments, the 5' and 3' flanking sequences are greater than 500 bases in length. In some embodiments, the 5' and 3' flanking sequences are greater than 1000 bases in length. In some embodiments, the 5' and 3' flanking sequences are from about 20 to about 1000 bases in length. In some embodiments, the 5' and 3' flanking sequences are from about 100 to about 1000 bases in length.

In some embodiments, the single stranded nucleic acid is single stranded DNA. In some embodiments, the systems further comprise reagents for introducing a break in the targeted sequence in a gene encoding the gene product in the genome of the cell. In some embodiments, the break is a double stranded break or a single stranded break. In some embodiments, the break is introduced by an enzyme selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like

effector (TALE)-nucleases, and a Type I restriction endonuclease, Type II restriction endonuclease, Type III, restriction endonuclease, Type IV restriction endonuclease or nickase.

In some embodiments, the systems further comprise a CRISPR -Cas system guide RNA that hybridizes with the target sequence. In some embodiments, the CRISPR -Cas system comprises a trans-activating cr (tracr) sequence. In some embodiments, the guide RNAs comprise a guide sequence fused to a tracr sequence.

In some embodiments, the present invention provides a cell comprising a system as described above. In some embodiments, the present invention provides for use of a system described above to treat a disease by altering expression of gene in a target cell or editing the genome of a target cell.

In some embodiments, the present invention provides methods of altering expression of at least one gene product and/or genome editing comprising: introducing into a cell a) a nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into the genome at a break in the genome by homologous recombination; and b) a recombinase; wherein when components (a) and (b) are introduced or expressed in the cell, the nucleic acid molecule encoding a nucleic acid sequence of interest is inserted by homologous recombination at the break to effect altered expression of at the at least one gene product and/or genome editing of the gene encoding the gene product of interest and wherein the nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into the genome at the break and the recombinase do not naturally occur together.

In some embodiments, the recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase. In some embodiments, the recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are greater than 50 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences greater than 100 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are greater than 200 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid

sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are greater than 500 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are greater than 1000 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences from about 20 to about 1000 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are from about 100 to about 1000 bases in length. In some embodiments, the flanking sequences are substantially homologous to a target region in the genome. In some embodiments, the flanking regions are at least 95%, 96%, 97%, 98%, 99% or 100% homologous to a target region in the genome.

In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest is single stranded DNA. In some embodiments, the break is a double stranded break or a single stranded break. In some embodiments, the break is introduced by an enzyme selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases. In some embodiments, the break is introduced by a Type I, II, III or IV restriction endonuclease or nickase.

In some embodiments, the methods further comprise introducing into the cell a CRISPR -Cas system guide RNA that hybridizes with the target sequence. In some embodiments, the CRISPR-Cas system comprises a trans-activating cr (tracr) sequence. In some embodiments, the guide RNAs comprise a guide sequence fused to a tracr sequence.

In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest is introduced into the cell by electroporation. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian cell. In some embodiments, the mammalian cell is a human cell.

In some embodiments, the expression of one or more gene products is increased. In some embodiments, the expression of one or more gene products is decreased. In some embodiments, the expression of two or more gene products is altered.

In some embodiments, the present invention provides an engineered, non-naturally occurring system for altering expression of a gene product and/or genome editing comprising:

a) a nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into the genome at a break in the genome by homologous recombination; and b) a recombinase; wherein when components (a) and (b) are introduced or expressed in the cell, the nucleic acid molecule encoding a nucleic acid sequence of interest is inserted by homologous recombination at the break to effect altered expression of at the at least one gene product and/or genome editing of the gene encoding the gene product of interest and wherein the nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into the genome at the break and the recombinase do not naturally occur together.

In some embodiments, the recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase. In some embodiments, the recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are greater than 50 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences greater than 100 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are greater than 200 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are greater than 500 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are greater than 1000 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences from about 20 to about 1000 bases in length. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from the genomic target sequence, wherein the 5' and 3' flanking sequences are from about 100 to about 1000 bases in length. In some

embodiments, the flanking sequences are substantially homologous to a target region in the genome. In some embodiments, the flanking regions are at least 95%, 96%, 97%, 98%, 99% or 100% homologous to a target region in the genome.

In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest is single stranded DNA. In some embodiments, the break is a double stranded break or a single stranded break. In some embodiments, the break is introduced by an enzyme selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases. In some embodiments, the break is introduced by a Type I, II, III or IV restriction endonuclease or nickase.

In some embodiments, the systems further comprise a CRISPR-Cas system guide RNA that hybridizes with the target sequence. In some embodiments, the CRISPR-Cas system comprises a trans-activating cr (tracr) sequence. In some embodiments, the guide RNAs comprise a guide sequence fused to a tracr sequence.

In some embodiments, the present invention provides a cell comprising a system as described above. In some embodiments, the present invention provides for use of a system described above to treat a disease by altering expression of gene in a target cell or editing the genome of a target cell.

It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as “comprises”, “comprised”, “comprising” and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean “includes”, “included”, “including”, and the like; and that terms such as “consisting essentially of” and “consists essentially of” have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention. These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

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

FIG 1. Precise gene repair of a damaged DNA base (Red) using the Cas9/gRNA RNP (not depicted) and the recombinase enzymes. Recombinases form a filament with (more strongly) or without an Adenosine triphosphate cofactor.

FIG. 2A-C. Improvement of HDR in Hela Cells by expressing either the RecA recombinase or the UvsX recombinase. (A) Raw data showing increased recombination. (B) Quantification of the increase in HDR.

DETAILED DESCRIPTION OF THE INVENTION

The terms "polynucleotide", "nucleotide", "nucleotide sequence", "nucleic acid" and "oligonucleotide" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

In aspects of the invention the terms "chimeric RNA", "chimeric guide RNA", "guide RNA", "single guide RNA" and "synthetic guide RNA" are used interchangeably and refer to the polynucleotide sequence comprising the guide sequence, the tracr sequence and the tracr mate sequence. The term "guide sequence" refers to the about 20 bp sequence within the guide RNA that specifies the target site and may be used interchangeably with the terms "guide" or "spacer". The term "tracr mate sequence" may also be used interchangeably with the term "direct repeat(s)". Exemplary CRISPR-Cas system are provided in US 8697359 and US 20140234972, both of which are incorporated herein by reference in their entirety.

As used herein, the term "filament" refers to a single stranded nucleic acid having a multimeric recombinase complex bound thereto. In some embodiments, the filament may be "isolated" and provided in a biologically compatible solution such as a buffered solution.

As used herein the term "wild type" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms.

As used herein the term "variant" should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature.

The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature.

"Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%. 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions.

As used herein, "stringent conditions" for hybridization refer to conditions under which a nucleic acid having complementarity to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Stringent conditions are generally sequence-dependent, and vary depending on a number of factors. In general, the longer the sequence, the higher the temperature at which the sequence specifically hybridizes to its target sequence. Non-limiting examples of stringent conditions are described in detail in Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part 1, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y.

"Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide

residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self 17 hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PCR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence.

As used herein, "expression" refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics.

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

The terms "therapeutic agent", "therapeutic capable agent" or "treatment agent" are used interchangeably and refer to a molecule or compound that confers some beneficial effect upon administration to a subject. The beneficial effect includes enablement of diagnostic determinations; amelioration of a disease, symptom, disorder, or pathological condition; reducing or preventing the onset of a disease, symptom, disorder or condition; and generally counteracting a disease, symptom, disorder or pathological condition.

As used herein, "treatment" or "treating," or "palliating" or "ameliorating" are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results

including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested.

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

The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M. J. MacPherson, B. D. Hames and G. R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R. I. Freshney, ed. (1987)).

Several aspects of the invention relate to vector systems comprising one or more vectors, or vectors as such. Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990).

Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

Vectors may be introduced and propagated in a prokaryote. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus of the recombinant protein. Such fusion vectors may serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A. respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, et al., 1987. *EMBO J.* 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

In some embodiments, a vector drives protein expression in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in

cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In some embodiments, a vector is capable of driving expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, et al., 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Baneiji, et al., 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, et al., 1985. *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR system so as to drive expression of the one or more elements of the CRISPR system. In general, CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats), also known as SPIDRs (SPacer Interspersed Direct Repeats), constitute a family of DNA loci that are usually specific to a particular bacterial species. The CRISPR locus comprises a distinct class of interspersed short sequence repeats (SSRs) that were recognized

in *E. coli* (Ishino et al., J. Bacteriol., 169:5429-5433 [1987]; and Nakata et al., J. Bacteriol., 171:3553-3556 [1989]), and associated genes. Similar interspersed SSRs have been identified in *Haloferax mediterranei, Streptococcus pyogenes, Anabaena*, and *Mycobacterium tuberculosis* (See, Groenen et al., Mol. Microbiol., 10:1057-1065 [1993]; Hoe et al., Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl et al., Biochim. Biophys. Acta 1307:26-30 [1996]; and Mojica et al., Mol. Microbiol., 17:85-93 [1995]). The CRISPR loci typically differ from other SSRs by the structure of the repeats, which have been termed short regularly spaced repeats (SRSRs) (Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; and Mojica et al., Mol. Microbiol., 36:244-246 [2000]). In general, the repeats are short elements that occur in clusters that are regularly spaced by unique intervening sequences with a substantially constant length (Mojica et al., [2000], supra). Although the repeat sequences are highly conserved between strains, the number of interspersed repeats and the sequences of the spacer regions typically differ from strain to strain (van Embden et al., J. Bacteriol., 182:2393-2401 [2000]). CRISPR loci have been identified in more than 40 prokaryotes (See e.g., Jansen et al., Mol. Microbiol., 43:1565-1575 [2002]; and Mojica et al., [2005]) including, but not limited to *Aeropyrum, Pyrobaculum, Sulfolobus, Archaeoglobus, Halocarcula, Methanobacteriumn, Methanococcus, Methanosarcina, Methanopyrus, Pyrococcus, Picrophilus, Thernioplasnia, Corynebacterium, Mycobacterium, Streptomyces, Aquifrx, Porphvromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoanaerobacter, Mycoplasma, Fusobacterium, Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myrococcus, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema*, and *Thermotoga.*

In general, "CRISPR system" refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or other sequences and transcripts from a CRISPR locus. In some embodiments, one or more elements of a CRISPR system is derived from a type I, type II, or type III CRISPR system. In some embodiments, one or more elements of a CRISPR system is derived from a particular organism comprising an endogenous CRISPR system, such as *Streptococcus pyogenes*. In general, a CRISPR system is characterized by elements

that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell. In some embodiments, the target sequence may be within an organelle of a eukaryotic cell, for example, mitochondrion or chloroplast. A sequence or template that may be used for recombination into the targeted locus comprising the target sequences is referred to as an "editing template" or "editing polynucleotide" or "editing sequence". In aspects of the invention, an exogenous template polynucleotide may be referred to as an editing template. In an aspect of the invention the recombination is homologous recombination.

Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. Without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence. In some embodiments, the tracr sequence has sufficient complementarity to a tracr mate sequence to hybridize and participate in formation of a CRISPR complex. As with the target sequence, it is believed that complete complementarity is not needed, provided there is sufficient to be functional. In some embodiments, the tracr sequence has at least 50%, 60%, 70%, 80%, 90%, 95% or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. In some embodiments, one or more vectors driving expression of one or more elements of a CRISPR system are introduced into a host cell such that expression of the elements of the CRISPR system direct formation of a CRISPR complex at one or more target sites. For example, a Cas enzyme, a guide sequence linked to a tracr-mate sequence, and a tracr sequence could each be operably linked to separate regulatory elements on separate

vectors. Alternatively, two or more of the elements expressed from the same or different regulatory elements, may be combined in a single vector, with one or more additional vectors providing any components of the CRISPR system not included in the first vector. CRISPR system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5′ with respect to ("upstream" of) or 3′ with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a CRISPR enzyme and one or more of the guide sequence, tracr mate sequence (optionally operably linked to the guide sequence), and a tracr sequence embedded within one or more intron sequences (e.g. each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the CRISPR enzyme, guide sequence, tracr mate sequence, and tracr sequence are operably linked to and expressed from the same promoter.

In preferred embodiments, a nucleic acid sequence encoding exogenous recombinase is co-expressed in the host cell with the other CRISPR system components. Without being limited to any theory, expression of the exogenous recombinase with the other CRISPR system components increasing the frequency or efficiency of use of the HDR pathway in a cell transformed or transduced with the system. Accordingly, in some preferred embodiments, the systems of the present invention include a vector comprising a regulatory element operably linked to a nucleic acid sequence encoding a recombinase. In some embodiments, the recombinase is a bacterial recombinase. In some embodiments, the recombinase does not occur naturally in the cell type transformed or transduced with the vector system. In some embodiments, the recombinase is selected from the group consisting of Rad51, RecA recombinase and UvsX recombinase or proteins containing a RecA or RadA domain. In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a "cloning site"). In some embodiments, one or more insertion sites (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. In some embodiments, a vector comprises an insertion site upstream of a tracr mate sequence, and optionally downstream of a regulatory element operably linked to the tracr mate sequence, such that following insertion of a guide sequence into the insertion site and upon expression the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell. In some

embodiments, a vector comprises two or more insertion sites, each insertion site being located between two tracr mate sequences so as to allow insertion of a guide sequence at each site. In such an arrangement, the two or more guide sequences may comprise two or more copies of a single guide sequence, two or more different guide sequences, or combinations of these. When multiple different guide sequences are used, a single expression construct may be used to target CRISPR activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell.

In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, such as a Cas protein. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologs thereof, or modified versions thereof. These enzymes are known; for example, the amino acid sequence of *S. pyogenes* Cas9 protein may be found in the SwissProt database under accession number Q99ZW2. In some embodiments, the unmodified CRISPR enzyme has DNA cleavage activity, such as Cas9. In some embodiments the CRISPR enzyme is Cas9, and may be Cas9 from *S. pyogenes* or *S. pneumoniae*. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence. In some embodiments, a vector encodes a CRISPR enzyme that is mutated to with respect to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from *S. pyogenes* converts Cas9 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). Other examples of mutations that render Cas9 a nickase include, without limitation, H840A, N854A, and N863A. In aspects of the invention, nickases may be used for genome editing via homologous recombination.

In some embodiments, a Cas9 nickase may be used in combination with guide sequence(s), e.g., two guide sequences, which target respectively sense and antisense strands of the DNA target. This combination allows both strands to be nicked and used to induce NHEJ. Applicants have demonstrated (data not shown) the efficacy of two nickase targets (i.e., sgRNAs targeted at the same location but to different strands of DNA) in inducing mutagenic NHEJ. A single nickase (Cas9-D10A with a single sgRNA) is unable to induce NHEJ and create indels but Applicants have shown that double nickase (Cas9-D01A and two sgRNAs targeted to different strands at the same location) can do so in human embryonic stem cells (hESCs). The efficiency is about 50% of nuclease (i.e., regular Cas9 without D10 mutation) in hESCs.

As a further example, two or more catalytic domains of Cas9 (RuvC I, RuvC II, and RuvC III) may be mutated to produce a mutated Cas9 substantially lacking all DNA cleavage activity. In some embodiments, a D10A mutation is combined with one or more of H840A, N854A, or N863A mutations to produce a Cas9 enzyme substantially lacking all DNA cleavage activity. In some embodiments, a CRISPR enzyme is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is less than about 25%, 10%, 5%, 1%, 0.1%, 0.01%, or lower with respect to its non-mutated form. Other mutations may be useful; where the Cas9 or other CRISPR enzyme is from a species other than *S. pyogenes*, mutations in corresponding amino acids may be made to achieve similar effects.

In some embodiments, an enzyme coding sequence encoding a CRISPR enzyme is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, mouse, rat, rabbit, dog, or non-human primate. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g. about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most

frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database", and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, Pa.), are also available. In some embodiments, one or more codons (e.g. 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a CRISPR enzyme correspond to the most frequently used codon for a particular amino acid.

In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control

guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

A guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell. Exemplary target sequences include those that are unique in the target genome. For example, for the *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGG (SEQ ID NO: 9) where NNNNNNNNNNNNXGG (SEQ ID NO: 10) (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. pyogenes* Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNXGG (SEQ ID NO: 11) where NNNNNNNNNNNXGG (SEQ ID NO: 12) (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. For the *S. thermophilus* CRISPR1 Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 1) where NNNNNNNNNNNNXXAGAAW (SEQ ID NO: 2) (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. thermophilus* CRISPR1 Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 3) where NNNNNNNNNNNXXAGAAW (SEQ ID NO: 4) (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. For the *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGGXG (SEQ ID NO: 13) where NNNNNNNNNNNNXGGXG (SEQ ID NO: 14) (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. pyogenes* Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNXGGXG (SEQ ID NO: 15) where NNNNNNNNNNNXGGXG (SEQ ID NO: 16) (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. In each of these sequences "M" may be A, G, T, or C, and need not be considered in identifying a sequence as unique.

In some embodiments, a guide sequence is selected to reduce the degree of secondary structure within the guide sequence. Secondary structure may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online

webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g. A. R. Gruber et al., 2008, *Cell* 106(1): 23-24; and PA Carr and GM Church, 2009, *Nature Biotechnology* 27(12): 1151-62). Further algorithms may be found in U.S. application Ser. No. 61/836,080; incorporated herein by reference.

In general, a tracr mate sequence includes any sequence that has sufficient complementarity with a tracr sequence to promote one or more of: (1) excision of a guide sequence flanked by tracr mate sequences in a cell containing the corresponding tracr sequence; and (2) formation of a CRISPR complex at a target sequence, wherein the CRISPR complex comprises the tracr mate sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracr sequence or tracr mate sequence. In some embodiments, the degree of complementarity between the tracr sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. Example illustrations of optimal alignment between a tracr sequence and a tracr mate sequence are provided in US 8697359, incorporated herein by reference in its entirety. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and tracr mate sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. Preferred loop forming sequences for use in hairpin structures are four nucleotides in length, and most preferably have the sequence GAAA. However, longer or shorter loop sequences may be used, as may alternative sequences. The sequences preferably include a nucleotide triplet (for example, AAA), and an additional nucleotide (for example C or G). Examples of loop forming sequences include CAAA and AAAG. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In some embodiments, the single transcript further includes a transcription termination sequence; preferably this is a polyT sequence, for example six T nucleotides. Further non-limiting examples of single polynucleotides comprising a guide sequence, a tracr

mate sequence, and a tracr sequence are as follows (listed 5′ to 3′), where "N" represents a base of a guide sequence, the first block of lower case letters represent the tracr mate sequence, and the second block of lower case letters represent the tracr sequence, and the final poly-T sequence represents the transcription terminator:

(1) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGA
AAtaaatcttgcagaagctacaaagataaggcttcatgccgaaa
tcaacaccctgtcattttatggcagggtgttttcgttatttaaT
TTTTT (SEQ ID NO: 5);

(2) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcag
aagctacaaagataaggcttcatgccgaaatcaacaccctgtca
ttttatggcagggtgttttcgttatttaaTTTTTT (SEQ ID
NO: 6);

(3) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcag
aagctacaaagataaggcttcatgccgaaatcaacaccctgtca
ttttatggcagggtgtTTTTTT (SEQ ID NO: 7);

(4) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtagca
agttaaaataaggctagtccgttatcaacttgaaaaagtggcac
cgagtcggtgcTTTTTT (SEQ ID NO: 8);

(5) NNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagt
taaaataaggctagtccgttatcaacttgaaaaagtgTTTTTTT
(SEQ ID NO: 9);
and

(6) NNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagt
taaaataaggctagtccgctatcaTTTTTTTT (SEQ ID NO:
10).

In some embodiments, sequences (1) to (3) are used in combination with Cas9 from *S. thermophilus* CRISPR1. In some embodiments, sequences (4) to (6) are used in combination with Cas9 from *S. pyogenes*. In some embodiments, the tracr sequence is a separate transcript from a transcript comprising the tracr mate sequence.

In some embodiments, the CRISPR enzyme is part of a fusion protein comprising one or more heterologous protein domains (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the CRISPR enzyme). A CRISPR enzyme fusion protein may comprise any additional protein sequence, and optionally a linker sequence between any two domains. Examples of protein domains that may be fused to a CRISPR enzyme include, without limitation, epitope tags, reporter gene sequences, and protein domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity and nucleic acid binding activity. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-5-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). A CRISPR enzyme may be fused to a gene sequence encoding a protein or a fragment of a protein that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4A DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions. Additional domains that may form part of a fusion protein comprising a CR ISPR enzyme are described in US20110059502, incorporated herein by reference. In some embodiments, a tagged CRISPR enzyme is used to identify the location of a target sequence.

In an aspect of the invention, a reporter gene which includes but is not limited to glutathione-5-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP), may be introduced into a cell to encode a gene product which serves as a marker by which to measure the alteration or modification of expression of the gene product. In a further embodiment of the invention, the DNA molecule encoding the gene product may be introduced into the cell via a vector. In a preferred embodiment of the invention the gene product is luciferase. In a further embodiment of the invention the expression of the gene product is decreased.

In some embodiments, the present invention provides a nucleic acid filament comprising a single stranded nucleic acid of interest encoding a nucleic acid sequence of interest to be inserted into a genome. In some preferred embodiments, single stranded nucleic acid sequence is single stranded DNA. In some embodiments, the single stranded nucleic acid molecule has bound thereto a multimeric recombinase complex. In some embodiments, the nucleic of interest comprises 5' and 3' flanking regions, preferably which are about 100% homologous to a genomic target sequence encoding a gene product. In some embodiments, the 5' and 3' flanking sequences flank an insert sequence that is different from said genomic target sequence. In some embodiments, the insert sequence may differ from the genomic target sequence by a single (one) base. In some embodiments, the insert sequence may differ from the genomic target sequence by a more than one bases (i.e., 2, 3, 4, 5, 6, 7, 8, 9, `10, 20, 30, 40, 50, 100 or more bases. In some embodiments, the insert sequence is designed to insert a sequence of a defined length (for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, `10, 20, 30, 40, 50, 100 or more bases) at a target site. In some embodiments, the insert sequence is designed to delete a sequence of a defined length (for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, `10, 20, 30, 40, 50, 100 or more bases) at a target site.

In some embodiments, the filament is synthesized in vitro by incubating said single stranded nucleic acid of interest with a recombinase so that a multimeric recombinase complex is formed on said single stranded nucleic acid of interest. In some embodiments, the filament further comprises nucleotides associated therewith. In some embodiments, the nucleotide is selected from the group consisting of a nucleotide triphosphate or analog. In some embodiments, the nucleotide triphosphate or analog is selected from the group consisting of adenosine triphosphate, adenosine monophosphate, adenosine diphosphate, adenosine triphosphate-γS, adenosine monophosphate-PNP, and adenosine diphosphate-AlF_4.

The systems may be collectively referred to as Recombinase Assisted Cas9-mediated gene Repair (RACeR systems). It will be understood that the filaments may be used with or without other CRISPR system components, such as Cas9. It will be further understood that the system may be used to edit a genome in combination with a break in the target genomic sequence or in the absence of a break in the target genomic sequence.

As described above, the recombinase may be selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase. In some embodiments, the recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

In some embodiments, the 5' and 3' flanking sequences are greater than 20, 50, 100, 200, 500, or 1000 bases in length, and can preferably range from about 20 to about 1000 bases in length, from about 100 to about 1000 bases in length, from about 20 to 5000 bases in length, from about 100 to 5000 bases in length, or from about 100 to 10,000 bases in length.

In some embodiments, filament systems of the present invention further comprise CRISPR or CRISPR-type components. Accordingly, in some embodiments, the filament systems comprise reagents for introducing a break in said targeted sequence in a gene encoding the gene product in said genome of the cell. In some embodiments, the break is a double stranded break. In some embodiments, the break is introduced by an enzyme selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases, or by type I, II, III or IV restriction enzymes or nickases. In some embodiments, the systems further comprise a CRISPR-Cas system guide RNA that hybridizes with the target sequence. In some embodiments, the CRISPR-Cas system comprises a trans-activating cr (tracr) sequence. In some embodiments, the guide RNAs comprise a guide sequence fused to a tracr sequence. In some embodiments, the filament is introduced into said cell by electroporation.

In some aspects, the invention provides methods comprising delivering one or more polynucleotides, such as or one or more vectors, systems or filaments as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some aspects, the invention further provides cells produced by such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells. In some embodiments, a CRISPR enzyme in combination with (and optionally complexed with) a guide sequence or filament is delivered to a cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a CRISPR or RACeR system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon. TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10): 1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin

51(1):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology Doerfler and Bohm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995): Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

The use of RNA or DNA viral based systems for the delivery of nucleic acids take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic

gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommnerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700). In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line may also be infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which

adenovirus is more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panel, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr −/−, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-MeI 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT cell lines, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a CRISPR system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, is used

to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

In some embodiments, one or more vectors described herein are used to produce a non-human transgenic animal or transgenic plant. In some embodiments, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. In certain embodiments, the organism or subject is a plant. In certain embodiments, the organism or subject or plant is algae. Methods for producing transgenic plants and animals are known in the art, and generally begin with a method of cell transfection, such as described herein. Transgenic animals are also provided, as are transgenic plants, especially crops and algae. The transgenic animal or plant may be useful in applications outside of providing a disease model. These may include food or feed production through expression of, for instance, higher protein, carbohydrate, nutrient or vitamins levels than would normally be seen in the wildtype. In this regard, transgenic plants, especially pulses and tubers, and animals, especially mammals such as livestock (cows, sheep, goats and pigs), but also poultry and edible insects, are preferred.

Transgenic algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell, which may be in vivo, ex vivo or in vitro. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal or plant (including micro-algae), and modifying the cell or cells. Culturing may occur at any stage ex vivo. The cell or cells may even be re-introduced into the non-human animal or plant (including micro-algae).

In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence.

In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In another aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a RACeR filament to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide.

With recent advances in crop genomics, the ability to use CRISPR-Cas or RACeR systems to perform efficient and cost effective gene editing and manipulation will allow the rapid selection and comparison of single and multiplexed genetic manipulations to transform such genomes for improved production and enhanced traits. In this regard reference is made to U.S. patents and publications: U.S. Pat. No. 6,603,061—*Agrobacterium*-Mediated Plant Transformation Method; U.S. Pat. No. 7,868,149—Plant Genome Sequences and Uses Thereof and US 2009/0100536—Transgenic Plants with Enhanced Agronomic Traits, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Morrell et al "Crop genomics:advances and applications" Nat Rev Genet. 2011 Dec. 29; 13(2):85-96 are also herein incorporated by reference in their entirety.

In plants, pathogens are often host-specific. For example, *Fusarium oxysporum* f. sp. *lycopersici* causes tomato wilt but attacks only tomato, and *F. oxysporn f. dianthii Puccinia graminis* f. sp. *tritici* attacks only wheat. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible. There can also be Horizontal Resistance, e.g., partial resistance against all races of a pathogen, typically controlled by many genes and Vertical Resistance, e.g., complete resistance to some races of a pathogen but not to other races, typically controlled by a few genes. In a Gene-for-Gene level, plants and pathogens evolve together, and the genetic changes in one balance changes in other. Accordingly, using Natural Variability, breeders combine most useful genes for Yield, Quality, Uniformity, Hardiness, Resistance. The sources of resistance genes include native or foreign Varieties, Heirloom Varieties, Wild

Plant Relatives, and Induced Mutations, e.g., treating plant material with mutagenic agents. Using the present invention, plant breeders are provided with a new tool to induce mutations. Accordingly, one skilled in the art can analyze the genome of sources of resistance genes, and in Varieties having desired characteristics or traits employ the present invention to induce the rise of resistance genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

In one aspect, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting a guide sequence upstream of the tracr mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence. In some embodiments, the kits may comprise a single stranded DNA sequence for use in a RACeR system, recombinase enzyme, and/or suitable nucleotide cofactors. Elements may be provided individually or in combinations, and may be provided in any suitable container, such as a vial, a bottle, or a tube. In some embodiments, the kit includes instructions in one or more languages, for example in more than one language.

In some embodiments, a kit comprises one or more reagents for use in a process utilizing one or more of the elements described herein. Reagents may be provided in any suitable container. For example, a kit may provide one or more reaction or storage buffers. Reagents may be provided in a form that is usable in a particular assay, or in a form that requires addition of one or more other components before use (e.g. in concentrate or lyophilized form). A buffer can be any buffer, including but not limited to a sodium carbonate buffer, a sodium bicarbonate buffer, a borate buffer, a Tris buffer, a MOPS buffer, a HEPES buffer, and combinations thereof. In some embodiments, the buffer is alkaline. In some embodiments, the buffer has a pH from about 7 to about 10. In some embodiments, the kit comprises one or more oligonucleotides corresponding to a guide sequence for insertion into a vector so as to operably link the guide sequence and a regulatory element. In some embodiments, the kit comprises a homologous recombination template polynucleotide.

In one aspect, the invention provides methods for using one or more elements of a CRISPR or RACeR system. The CRISPR or RACeR complex of the invention provides an effective means for modifying a target polynucleotide. The CRISPR or RACeR complex of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target polynucleotide in a multiplicity of cell types. As such the CRISPR or RACeR complex of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the target polynucleotide. The guide sequence is linked to a tracr mate sequence, which in turn hybridizes to a tracr sequence.

The target polynucleotide of a CRISPR or RACeR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence) Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR enzyme.

The target polynucleotide of a CRISPR or RACeR complex may include a number of disease-associated genes and polynucleotides as well as signaling biochemical pathway-associated genes and polynucleotides as listed in U.S. provisional patent applications 61/736,527 and 61/748,427, both entitled SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION filed on Dec. 12, 2012 and Jan. 2, 2013, respectively, the contents of all of which are herein incorporated by reference in their entirety.

Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A “disease-associated” gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or

cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

Examples of disease-associated genes and polynucleotides are available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web.

Examples of disease-associated genes and polynucleotides are listed in Tables A and B. Disease specific information is available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web. Examples of signaling biochemical pathway-associated genes and polynucleotides are listed in Table C.

Mutations in these genes and pathways can result in production of improper proteins or proteins in improper amounts which affect function. genes, proteins and pathways may be the target polynucleotide of a CRISPR complex.

DISEASE	
DISORDERS	GENE(S)
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alpha; PPAR gamma; WT1 (Wilms Tumor); FGF Receptor Family members (5 members: 1, 2, 3, 4, 5); CDKN2a; APC; RB (retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR (Androgen Receptor); TSG101; IGF; IGF Receptor; Igf1 (4 variants); Igf2 (3 variants); Igf 1 Receptor; Igf 2 Receptor; Bax; Bcl2; caspases family (9 members: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc

Age-related Macular Degeneration	Aber; Ccl2; Cc2; cp (ceruloplasmin); Timp3; cathepsinD; Vldlr; Ccr2
Schizophrenia	Neuregulin1 (Nrg1); Erb4 (receptor for Neuregulin); Complexin1 (Cplx1); Tph1 Tryptophan hydroxylase; Tph2 Tryptophan hydroxylase 2; Neurexin 1; GSK3; GSK3a; GSK3b
Disorders	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA; DTNBP1; Dao (Dao1)
Trinucleotide Repeat Disorders	HTT (Huntington's Dx); SBMA/SMAX1/AR (Kennedy's Dx); FXN/X25 (Friedrich's Ataxia); ATX3 (Machado-Joseph's Dx); ATXN1 and ATXN2 (spinocerebellar ataxias); DMPK (myotonic dystrophy); Atrophin-1 and Atn 1 (DRPLA Dx); CBP (Creb-BP-global instability); VLDLR (Alzheimer's); Atxn7; Atxn10
Fragile X Syndrome	FMR2; FXR1; FXR2; mGLUR5
Secretase Related. Disorders	APH-1 (alpha and beta); Presenilin (Psen1); nicastrin (Ncstn); PEN-2
Others	Nos1; Parp1; Nat1; Nat2
Prion-related disorders	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; V EGF-c)
Drug addiction	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (alcohol)
Autism	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexin 1; Fragile X (FMR2 (AFF2); FXR1; FXR2; Mglur5)
Alzheimer's Disease	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterin; PS1; SORL1; CR1; Vld1r; Uba1; Uba3; CHIP28 (Aqp1, Aquaporin 1); Uchl1; Uchl3; APP

Inflammation	1L-10; IL-1 (1L-1a; IL-1b); 1L-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); II-23; Cx3er1; ptpn22; TNFa; NOD2/CARD15 for IBD; IL-6; 1L-12 (1L-12a; 1L-12b); CTLA4; Cx3cl1
Parkinson's Disease	x-Synuclein; DJ-1; LRRK2; Parkin; PINK1

TABLE B	
Blood and coagulation diseases and disorders	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); Bare lymphocyte syndrome (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Bleeding disorders (TBXA2R, P2RX1, P2X1); Factor H and factor H-like 1 (HF1, CFH, HUS); Factor V and factor VIII (MCFD2); Factor VII deficiency (F7); Factor X deficiency (F10); Factor XI deficiency (F11); Factor XII deficiency (F12, HAF); Factor XIIIA deficiency (F13A1, F13A); Factor XIIIB deficiency (F13B); Fanconi anemia (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCJ, PHF9, FANCL, FANCM, KIAA1596); Hemophagocytic lymphohistiocytosis disorders (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemophilia A (F8, F8C, HEMA); Hemophilia B (F9, HEMB), Hemorrhagic disorders (PI, ATT, F5); Leukocyde deficiencies and disorders (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5,

	LVWM, CACH, CLE, EIF2B4); Sickle cell anemia
	(HBB); Thalassemia (HBA2, HBB, HBD, LCRB,
	HBA1).
Cell dysregulation	B-cell non-Hodgkin lymphoma (BCL7A, BCL7);
and oncology	Leukemia (TAL1 TCL5, SCL, TAL2, FLT3, NBS1,
diseases and	NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B,
disorders	BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2,
	GMPS, AF10, ARHGEF12, LARG, KIAA0382,
	CALM, CLTH, CEBPA, CEBP, CHIC2, BTL,
	FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E,
	CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1,
	NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145,
	PLZF, PML, MYL, STAT5B, AF10, CALM,
	CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR,
	CML, PHL, ALL, GRAF, NF1, VRNF, WSS,
	NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2,
	CCND1, PRAD1, BCL1, TCRA, GATA1, GF1,
	ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1,
	NUP214, D9S46E, CAN, CAIN).
Inflammation and	AIDS (KIR3DL1, NKAT3, NKB1, AMB11,
immune related	KIR3DS1, IFNG, CXCL12, SDF1); Autoimmune
diseases and	lymphoproliferative syndrome (TNFRSF6, APT1,
disorders	FAS, CD95, ALPS1A); Combined immuno-
	deficiency, (IL2RG, SCIDX1, SCIDX, IMD4);
	HIV-1 (CCL5, SCYA5, D17S136E, TCP228), HIV
	susceptibility or infection (IL10, CSIF, CMKBR2,
	CCR2, CMKBR5, CCCKR5 (CCR5)); Immuno-
	deficiencies (CD3E, CD3G, AICDA, AID, HIGM2,
	TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5,
	CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID,
	XPID, PIDX, TNFRSF14B, TACI); Inflammation
	(IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a

	(CTLA8), IL-17b, IL-17c, IL-17d, IL-17f, II-23,
	Cx3cr1, ptpn22, TNFa, NOD2/CARD15 for IBD,
	IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1);
	Severe combined immunodeficiencies (SCIDs)
	(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA,
	RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R,
	CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
Metabolic, liver,	Amyloid neuropathy (TTR, PALB); Amyloidosis
kidney and protein	(APOA1, APP, AAA, CVAP, AD1, GSN, FGA,
diseases and	LYZ, TTR, PALB); Cirrhosis (KRT18, KRT8,
disorders	CIRH1A, NAIC, TEX292, KIAA1988); Cystic
	fibrosis (CFTR, ABCC7, CF, MRP7); Glycogen
	storage diseases (SLC2A2, GLUT2, G6PC, G6PT,
	G6PT1, GAA, LAMP2, LAMPB, AGL, GDE,
	GBE1, GYS2, PYGL, PFKM); Hepatic adenoma,
	142330 (TCF1, HNF1A, MODY3), Hepatic failure,
	early onset, and neurologic disorder (SCOD1,
	SCO1), Hepatic lipase deficiency (LIPC), Hepato-
	blastoma, cancer and carcinomas (CTNNB1,
	PDGFRL, PDGRL, PRLTS, AXIN1, AXIN,
	CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET,
	CASP8, MCH5; Medullary cystic kidney disease
	(UMOD, HNFJ, FJHN, MCKD2, ADMCKD2);
	Phenylketonuria (PAH, PKU1, QDPR, DHPR,
	PTS); Polycystic kidney and hepatic disease
	(FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4,
	PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Muscular/Skeletal	Becker muscular dystrophy (DMD, BMD, MYF6),
diseases and	Duchenne Muscular Dystrophy (DMD, BMD);
disorders	Emery-Dreifuss muscular dystrophy (LMNA,
	LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B,
	LMNA, LMN1, EMD2, FPLD, CMD1A); Facio-

	scapulohumeral muscular dystrophy (FSHMD1A,
	FSHD1A); Muscular dystrophy (FKRP, MDC1C,
	LGMD2I, LAMA2, LAMM, LARGE, KIAA0609,
	MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3,
	DYSF, LGMD2B, SGCG, LGMD2C, DMDA1,
	SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2,
	SGCB, LGMD2E, SGCD, SGD, LGMD2F,
	CMD1L, TCAP, LGMD2G, CMD1N, TRIM32,
	HT2A, LGMD2H, FKRP, MDC1C, LGMD2I,
	TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3,
	LGMD1C, SEPN1, SELN, RSMD1, PLEC1,
	PLTN, EBS1); Osteopetrosis (LRP5, BMND1,
	LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7,
	OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116,
	OPTB1); Muscular atrophy (VAPB, VAPC, ALS8,
	SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2,
	SPG17, GARS, SMAD1, CMT2D, HEXB,
	IGHMBP2, SMUBP2, CATF1, SMARD1).
Neurological and	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF
neuronal diseases	(VEGF-a, VEGF-b, VEGF-c); Alzheimer disease
and disorders	(APP, AAA, CVAP, AD1, APOE, AD2, PSEN2,
	AD4, STM2, APBB2, FE65L1, NOS3, PLAU,
	URK, ACE, DCP1, ACE1, MPO, PACIP1,
	PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1,
	AD3); Autism (Mecp2, BZRAP1, MDGA2,
	Sema5A, Neurexin 1, GLO1, MECP2, RTT,
	PPMX, MRX16, MRX79, NLGN3, NLGN4,
	KIAA1260, AUTSX2); Fragile X Syndrome
	(FMR2, FXR1, FXR2, mGLUR5); Huntington's
	disease and disease like disorders (HD, IT15,
	PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17);
	Parkinson disease (NR4A2, NURR1, NOT, TINUR,

	SNCAIP, TBP, SCA17, SNCA, NACP, PARK1,
	PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1,
	PARK6, UCHL1, PARK5, SNCA, NACP, PARK1,
	PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2);
	Rett syndrome (MECP2, RTT, PPMX, MRX16,
	MRX79, CDKL5, STK9, MECP2, RTT, PPMX,
	MRX16, MRX79, x-Synuclein, DJ-1); Schizo-
	phrenia (Neuregulin1 (Nrg1), Erb4 (receptor for
	Neuregulin), Complexin1 (Cplx1), Tph1 Trypto-
	phan hydroxylase, Tph2, Tryptophan hydroxylase 2,
	Neurexin 1, GSK3, GSK3a, GSK3b, 5-HTT
	(Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA,
	DTNBP1, Dao (Dao1)); Secretase Related Dis-
	orders (APH-1 (alpha and beta), Presenilin (Psen1),
	nicastrin, (Ncstn), PEN-2, Nos1, Parp1, Nat1,
	Nat2); Trinucleotide Repeat Disorders (HTT
	(Huntington's Dx), SBMA/SMAX1/AR (Kennedy's
	Dx), FXN/X25 (Friedrich's Ataxia), ATX3
	(Machado- Joseph's Dx), ATXN1 and ATXN2
	(spinocerebellar ataxias), DMPK (myotonic
	dystrophy), Atrophin-1 and Atn1 (DRPLA Dx),
	CBP (Creb-BP - global instability), VLDLR
	(Alzheimer's), Atxn7, Atxn10).
Occular diseases	Age-related macular degeneration (Aber, Ccl2, Cc2,
and disorders	cp (ceruloplasmin), Timp3, cathepsinD, Vldlr,
	Ccr2); Cataract (CRYAA, CRYA1, CRYBB2,
	CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA,
	CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1,
	CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD,
	CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4,
	CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2,
	CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2,

	CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8,
	CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1,
	CAM, KRIT1); Corneal clouding and dystrophy
	(APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3,
	CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX,
	PPCD, PPD, KTCN, COL8A2, FECD, PPCD2,
	PIP5K3, CFD); Cornea plana congenital (KERA,
	CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG,
	GPOA, OPTN, GLC1E, FIP2, HYPL, NRP,
	CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1,
	GLC3A); Leber congenital amaurosis (CRB1,
	RP12, CRX, CORD2, CRD, RPGRIP1, LCA6,
	CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D,
	GUC2D, LCA1, CORD6, RDH12, LCA3);
	Macular dystrophy (ELOVL4, ADMD, STGD2,
	STGD3, RDS, RP7, PRPH2, PRPH, AVMD,
	AOFMD, VMD2).
Epilepsy, myoclonic,	EPM2A, MELF, EPM2
Lafora type, 254780	
Epilepsy, myoclonic,	NHLRC1, EPM2A, EPM2B
Lafora type, 254780	
Duchenne muscular	DMD, BMD
dystrophy, 310200 (3)	
AIDS, delayed/rapid	KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1
progression to (3)	
AIDS, rapid	IFNG
progression to,	
609423 (3)	
AIDS, resistance to	CXCL12, SDF1
(3)	
Alpha 1-Antitrypsin	SERPINA1 [serpin peptidase inhibitor, clade A
Deficiency	(alpha-1 antiproteinase, antitrypsin), member 1];

	SERPINA2 [serpin peptidase inhibitor, clade A
	(alpha-1 antiproteinase, antitrypsin), member 2];
	SERPINA3 [serpin peptidase inhibitor, clade A
	(alpha-1 antiproteinase, antitrypsin), member 3];
	SERPINA5 [serpin peptidase inhibitor, clade A
	(alpha-1 antiproteinase, antitrypsin), member 5];
	SERPINA6 [serpin peptidase inhibitor, clade A
	(alpha-1 antiproteinase, antitrypsin), member 6];
	SERPINA7 [serpin peptidase inhibitor, clade A
	(alpha-1 antiproteinase, antitrypsin), member 7];"
	AND "SERPLNA6 (serpin peptidase inhibitor,
	clade A (alpha-1 antiproteinase, antitrypsin),
	member 6)

TABLE C	
CELLULAR	
FUNCTION	GENES
PI3K/AKT Signaling	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2;
	EIF2AK2; PTEN; EIF4E; PRKCZ; GRK6;
	MAPK1; TSC1; PLK1; AKT2; IKBKB;
	PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;
	PIK3CB; PPP2R1A; MAPK8; BCL2L1;
	MAPK3; TSC2; ITGA1; KRAS; EIF4EBP1;
	RELA; PRKCD; NOS3; PRKAA1; MAPK9;
	CDK2; PPP2CA; PIM1; ITGB7; YWHAZ; ILK;
	TP53; RAF1.; IKBKG; RELB; DYRK1A;
	CDKN1A; ITGB1; MAP2K2; JAK1; AKT1;
	JAK2; PIK3R1; CHUK; PDPK1; PPP2R5C;
	CTNNB1.; MAP2K1; NFKB1; PAK3; ITGB3;
	CCND1; GSK3A; FRAP1; SFN; ITGA2; TTK;
	CSNK1A1; BRAF; GSK3B; AKT3; FOXO1;
	SGK; HSP90AA1; RPS6KB1
ERK/MAPK Signaling	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1;

	PRKAA2; EIF2AK2; RAC1; RAP1A; TLN1;
	EIF4E; ELK1; GRK6; MAPK1; RAC2; PLK1;
	AKT2; PIK3CA; CDK8; CREB1; PRKCI;
	PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; MAPK3; ITGA1; ETS1;
	KRAS; MYCN; EIF4EBP1; PPARG; PRKCD;
	PRKAA1; MAPK9; SRC; CDK2; PPP2CA;
	PIM1; PIK3C2A; ITGB7; YWHAZ; PPP1CC;
	KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C;
	MAP2K1; PAK3; ITGB3; ESR1; ITGA2; MYC;
	TTK; CSNK1A1; CRKL; BRAF; ATF4;
	PRKCA; SRF; STAT1; SGK
Glucocorticoid Receptor	RAC1; TAF4B; EP300; SMAD2; TRAF6;
Signaling	PCAF; ELK1; MAPK1; SMAD3; AKT2;
	IKBKB; NCOR2; UBE2I; PIK3CA; CREB1;
	FOS; HSPA5; NFKB2; BCL2; MAP3K14;
	STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1;
	MAPK3; TSC22D3; MAPK10; NRIP1; KRAS;
	MAPK13; RELA; STAT5A; MAPK9; NOS2A;
	PBX1; NR3C1; PIK3C2A; CDKN1C; TRAF2;
	SERPINE1; NCOA3; MAPK14; TNF; RAF1;
	IKBKG; MAP3K7; CREBBP; CDKN1A;
	MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2;
	PIK3R1; CHUK; STAT3; MAP2K1; NFKB1;
	TGFBR1; ESR1; SMAD4; CEBPB; JUN; AR;
	AKT3; CCL2; MMP1; STAT1; IL6; HSP90AA1
Axonal Guidance	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4;
Signaling	ADAM12; IGF1; RAC1; RAP1A; E1F4E;
	PRKCZ; NRP1; NTRK2; ARHGEF7; SMO;
	ROCK2; MAPK1; PGF; RAC2; PTPN11;
	GNAS; AKT2; PIK3CA; ERBB2; PRKC1;

	PTK2; CFL1; GNAQ; PIK3CB; CXCL12;
	PIK3C3; WNT11; PRKD1; GNB2L1; ABL1;
	MAPK3; ITGA1; KRAS; RHOA; PRKCD;
	PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1;
	FYN; ITGB1; MAP2K2; PAK4; ADAM17;
	AKT1; PIK3R1; GLI1; WNT5A; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA;
	ITGA2; EPHA8; CRKL; RND1; GSK3B;
	AKT3; PRKCA
Ephrin Receptor	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4;
Signaling	IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A;
	GRK6; ROCK2; MAPK1; PGF; RAC2;
	PTPN11; GNAS; PLK1; AKT2; DOK1; CDK8;
	CREB1; PTK2; CFL1; GNAQ; MAP3K14;
	CXCL12; MAPK8; GNB2L1; ABL1; MAPK3;
	ITGA1; KRAS; RHOA; PRKCD; PRKAA1;
	MAPK9; SRC; CDK2; PIM1; ITGB7; PXN;
	RAF1; FYN; DYRK1A; ITGB1; MAP2K2;
	PAK4, AKT1; JAK2; STAT3; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA;
	ITGA2; EPHA8; TTK; CSNK1A1; CRKL;
	BRAF; PTPN13; ATF4; AKT3; SGK
Actin Cytoskeleton	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5;
Signaling	IRAK1; PRKAA2; EIF2AK2; RAC1; INS;
	ARHGEF7; GRK6; ROCK2; MAPK1; RAC2;
	PLK1; AKT2; PIK3CA; CDK8; PTK2; CFL1;
	PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8;
	F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1;
	PIK3C2A; ITGB7; PPP1CC; PXN; VIL2; RAF1;
	GSN; DYRK1A; ITGB1; MAP2K2; PAK4;
	PIP5K1A; PIK3R1; MAP2K1; PAK3; ITGB3;

	CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL;
	BRAF; VAV3; SGK
Huntington's Disease	PRKCE; IGF1; EP300; RCOR1.; PRKCZ;
Signaling	HDAC4; TGM2; MAPK1; CAPNS1; AKT2;
	EGFR; NCOR2; SP1; CAPN2; PIK3CA;
	HDAC5; CREB1; PRKC1; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R;
	PRKD1; GNB2L1; BCL2L1; CAPN1; MAPK3;
	CASP8; HDAC2; HDAC7A; PRKCD; HDAC11;
	MAPK9; HDAC9; PIK3C2A; HDAC3; TP53;
	CASP9; CREBBP; AKT1; PIK3R1; PDPK1;
	CASP1; APAF1; FRAP1; CASP2; JUN; BAX;
	ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6;
	CASP3
Apoptosis Signaling	PRKCE; ROCK1; BID; IRAK1; PRKAA2;
	EIF2AK2; BAK1; BIRC4; GRK6; MAPK1;
	CAPNS1; PLK1; AKT2; IKBKB; CAPN2;
	CDK8; FAS; NFKB2; BCL2; MAP3K14;
	MAPK8; BCL2L1; CAPN1; MAPK3; CASP8;
	KRAS; RELA; PRKCD; PRKAA1; MAPK9;
	CDK2; PIM1; TP53; TNF; RAF1; IKBKG;
	RELB; CASP9; DYRK1A; MAP2K2; CHUK;
	APAF1; MAP2K1; NFKB1; PAK3; LMNA;
	CASP2; BIRC2; TTK; CSNK1A1; BRAF; BAX;
	PRKCA; SGK; CASP3; BIRC3; PARP1
B Cell Receptor	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2;
Signaling	PTPN11; AKT2; IKBKB; PIK3CA; CREB1;
	SYK; NFKB2; CAMK2A; MAP3K14; PIK3CB;
	PIK3C3; MAPK8; BCL2L1; ABL1; MAPK3;
	ETS1; KRAS; MAPK13; RELA; PTPN6;
	MAPK9; EGR1; PIK3C2A; BTK; MAPK14;
	RAF1; IKBKG; RELB; MAP3K7; MAP2K2;

	AKT1; PIK3R1; CHUK; MAP2K1; NFKB1;
	CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Leukocyte Extravasation	ACTN4; CD44; PRKCE; ITGAM; ROCK1;
Signaling	CXCR4; CYBA; RAC1; RAP1A; PRKCZ;
	ROCK2; RAC2; PTPN11; MMP14; PIK3CA;
	PRKCI; PTK2; PIK3CB; CXCL12; PIK3C3;
	MAPK8; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC;
	PIK3C2A; BTK; MAPK14; NOX1; PXN; VIL2;
	VASP; ITGB1; MAP2K2; CTNND1; PIK3R1;
	CTNNB1; CLDN1; CDC42; F11R; ITK; CRKL;
	VAV3; CTTN; PRKCA; MMP1; MMP9
Integrin Signaling	ACTN4; ITGAM; ROCK1; ITGA5; RAC1;
	PTEN; RAP1A; TLN1; ARHGEF7; MAPK1;
	RAC2; CAPNS1; AKT2; CAPN2; P1K3CA;
	PTK2; PIK3CB; PIK3C3; MAPK8; CAV1;
	CAPN1; ABL1; MAPK3; ITGA1; KRAS;
	RHOA; SRC; PIK3C2A; ITGB7; PPP1CC; ILK;
	PXN; VASP; RAF1; FYN; ITGB1; MAP2K2;
	PAK4; AKT1; PIK3R1; TNK2; MAP2K1;
	PAK3; ITGB3; CDC42; RND3; ITGA2; CRKL;
	BRAF; GSK3B; AKT3
Acute Phase Response	IRAK1; SOD2; MYD88; TRAF6; ELK1;
Signaling	MAPK1; PTPN11; AKT2; IKBKB; PIK3CA;
	FOS; NFKB2; MAP3K14; PIK3CB; MAPK8;
	RIPK1; MAPK3; IL6ST; KRAS; MAPK13;
	IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1;
	PDK1; IKBKG; RELB; MAP3K7; MAP2K2;
	AKT1; JAK2; PIK3R1; CHUK; STAT3;
	MAP2K1; NFKB1; FRAP1; CEBPB; JUN;

	AKT3; IL1R1; IL6
PTEN Signaling	ITGAM; ITGA5; RAC1; PTEN; PRKCZ;
	BCL2L11; MAPK1; RAC2; AKT2; EGFR;
	IKBKB; CBL; PIK3CA; CDKN1B; PTK2;
	NFKB2; BCL2; PIK3CB; BCL2L1; MAPK3;
	ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; ITGB1;
	MAP2K2; AKT1; PIK3R1; CHUK; PDGFRA;
	PDPK1; MAP2K1; NFKB1; ITGB3; CDC42;
	CCND1; GSK3A; ITGA2; GSK3B; AKT3;
	FOXO1; CASP3; RPS6KB1
p53 Signaling	PTEN; EP300; BBC3; PCAF; FASN; BRCA1;
	GADD45A; BIRC5; AKT2; PIK3CA; CHEK1;
	TP53INP1; BCL2; PIK3CB; PIK3C3; MAPK8;
	THBS1; ATR; BCL2L1; E2F1; PMAIP1;
	CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD;
	CDKN1A; HIPK2; AKT1; RIK3R1; RRM2B;
	APAF1; CTNNB1; SIRT1; CCND1; PRKDC;
	ATM; SFN; CDKN2A; JUN; SNAI2; GSK3B;
	BAX; AKT3
Aryl Hydrocarbon	HSPB1; EP300; FASN; TGM2; RXRA;
Receptor	MAPK1; NQO1; NCOR2; SP1; ARNT;
Signaling	CDKN1B; FOS; CHEK1; SMARCA4; NFKB2;
	MAPK8; ALDH1A1; ATR; E2F1; MAPK3;
	NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53;
	TNF; CDKN1A; NCOA2; APAF1; NFKB1;
	CCND1; ATM; ESR1; CDKN2A; MYC; JUN;
	ESR2; BAX; IL6; CYP1B1; HSP90AA1
Xenobiotic Metabolism	PRKCE; EP300; PRKCZ; RXRA; MAPK1;
Signaling	NQO1; NCOR2; PIK3CA; ARNT; PRKCI;

	NFKB2; CAMK2A; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; PRKD1; ALDH1A1;
	MAPK3; NRIP1; KRAS; MAPK13; PRKCD;
	GSTP1; MAPK9; NOS2A; ABCB1; AHR;
	PPP2CA; FTL; NFE2L2; PIK3C2A;
	PPARGC1A; MAPK14; TNF; RAF1; CREBBP;
	MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB1; KEAP1; PRKCA; EIF2AK3; IL6;
	CYP1B1; HSP90AA1
SAPK/JNK Signaling	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1;
	ELK1; GRK6; MAPK1; GADD45A; RAC2;
	PLK1; AKT2; PIK3CA; FADD; CDK8;
	PIK3CB; PIK3C3; MAPK8; RIPK1; GNB2L1;
	IRS1; MAPK3; MAPK10; DAXX; KRAS;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1;
	PIK3C2A; TRAF2; TP53; LCK; MAP3K7;
	DYRK1A; MAP2K2; PIK3R1; MAP2K1;
	PAK3; CDC42; JUN; TTK; CSNK1A1; CRKL;
	BRAF; SGK
PPAr/RXR Signaling	PRKAA2; EP300; INS; SMAD2; TRAF6;
	PPARA; FASN; RXRA; MAPK1; SMAD3;
	GNAS; IKBKB; NCOR2; ABCA1; GNAQ;
	NFKB2; MAP3K14; STAT5B; MAPK8; IRS1;
	MAPK3; KRAS; RELA; PRKAA1;
	PPARGC1A; NCOA3; MAPK14; INSR; RAF1;
	IKBKG; RELB; MAP3K7; CREBBP; MAP2K2;
	JAK2; CHUK; MAP2K1; NFKB1; TGFBR1;
	SMAD4; JUN; IL1R1; PRKCA; IL6;
	HSP90AA1; ADIPOQ
NF-KB Signaling	IRAK1; EIF2AK2; EP300; INS; MYD88;
	PRKCZ: TRAF6; TBK1; AKT2; EGFR; IKBKB;
	PIK3CA; BTRC; NFKB2; MAP3K14; PIK3CB;

	PIK3C3; MAPK8; RIPK1; HDAC2; KRAS;
	RELA; PIK3C2A; TRAF2; TLR4: PDGFRB;
	TNF; INSR; LCK; IKBKG; RELB; MAP3K7;
	CREBBP; AKT1; PIK3R1; CHUK; PDGFRA;
	NFKB1; TLR2; BCL10; GSK3B; AKT3;
	TNFAIP3; IL1R1
Neuregulin Signaling	ERBB4; PRKCE; ITGAM; ITGA5: PTEN;
	PRKCZ; ELK1; MAPK1; PTPN11; AKT2;
	EGFR; ERBB2; PRKCI; CDKN1B; STAT5B;
	PRKD1; MAPK3; ITGA1; KRAS; PRKCD;
	STAT5A; SRC; ITGB7; RAF1; ITGB1;
	MAP2K2; ADAM17; AKT1; PIK3R1; PDPK1;
	MAP2K1; ITGB3; EREG; FRAP1; PSEN1;
	ITGA2; MYC; NRG1; CRKL; AKT3; PRKCA;
	HSP90AA1; RPS6KB1
Wnt & Beta catenin	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1;
Signaling	SMO; AKT2; PIN1; CDH1; BTRC; GNAQ;
	MARK2; PPP2R1A; WNT11; SRC; DKK1;
	PPP2CA; SOX6; SFRP2: ILK; LEF1; SOX9;
	TP53; MAP3K7; CREBBP; TCF7L2; AKT1;
	PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1;
	CCND1; GSK3A; DVL1; APC; CDKN2A;
	MYC; CSNK1A1; GSK3B; AKT3; SOX2
Insulin Receptor	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1;
Signaling	TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKCI;
	PIK3CB; PIK3C3; MAPK8; IRS1; MAPK3;
	TSC2; KRAS; EIF4EBP1; SLC2A4; PIK3C2A;
	PPP1CC; INSR; RAF1; FYN; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; PDPK1; MAP2K1;
	GSK3A; FRAP1; CRKL; GSK3B; AKT3;
	FOXO1; SGK; RPS6KB1
IL-6 Signaling	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1;

	PTPN11; IKBKB; FOS; NFKB2: MAP3K14;
	MAPK8; MAPK3; MAPK10; IL6ST; KRAS;
	MAPK13; IL6R; RELA; SOCS1; MAPK9;
	ABCB1; TRAF2; MAPK14; TNF; RAF1;
	IKBKG; RELB; MAP3K7; MAP2K2; IL8;
	JAK2; CHUK; STAT3; MAP2K1; NFKB1;
	CEBPB; JUN; IL1R1; SRF; IL6
Hepatic Cholestasis	PRKCE; IRAK1; INS; MYD88; PRKCZ;
	TRAF6; PPARA; RXRA; IKBKB; PRKCI;
	NFKB2; MAP3K14; MAPK8; PRKD1;
	MAPK10; RELA; PRKCD; MAPK9; ABCB1;
	TRAF2; TLR4; TNF; INSR; IKBKG; RELB;
	MAP3K7; IL8; CHUK; NR1H2; TJP2; NFKB1;
	ESR1; SREBF1; FGFR4; JUN; IL1R1; PRKCA;
	IL6
IGF-1 Signaling	IGF1; PRKCZ; ELK1; MAPK1; PTPN11;
	NEDD4; AKT2; PIK3CA; PRKC1; PTK2; FOS;
	PIK3CB; PIK3C3; MAPK8; 1GF1R; IRS1;
	MAPK3; IGFBP7; KRAS; PIK3C2A; YWHAZ;
	PXN; RAF1; CASP9; MAP2K2; AKT1;
	PIK3R1; PDPK1; MAP2K1; IGFBP2; SFN;
	JUN; CYR61; AKT3; FOXO1; SRF; CTGF;
	RPS6KB1
NRF2-mediated	PRKCE; EP300; SOD2; PRKCZ; MAPK1;
Oxidative	SQSTM1; NQO1; PIK3CA; PRKC1; FOS;
Stress Response	PIK3CB; P1K3C3; MAPK8; PRKD1; MAPK3;
	KRAS; PRKCD; GSTP1; MAPK9; FTL;
	NFE2L2; PIK3C2A; MAPK14; RAF1;
	MAP3K7; CREBBP; MAP2K2; AKT1; PIK3R1;
	MAP2K1; PPIB; JUN; KEAP1; GSK3B; ATF4;
	PRKCA; EIF2AK3; HSP90AA1
Hepatic, Fibrosis/Hepatic	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1;

Stellate Cell Activation	MET; PGF; SMAD3; EGFR; FAS; CSF1;
	NFKB2; BCL2; MYH9; IGF1R; IL6R; RELA;
	TLR4; PDGFRB; TNF; RELB; IL8; PDGFRA;
	NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6;
	CTGF; MMP9
PPAR Signaling	EP300; INS; TRAF6; PPARA; RXRA; MAPK1;
	IKBKB; NCOR2; FOS; NFKB2; MAP3K14;
	STAT5B; MAPK3; NRIP1; KRAS; PPARG;
	RELA; STAT5A; TRAF2; PPARGC1A;
	PDGFRB; TNF; INSR; RAF1; IKBKG; RELB;
	MAP3K7; CREBBP; MAP2K2; CHUK;
	PDGFRA; MAP2K1; NFKB1; JUN; IL1R1;
	HSP90AA1
Fc Epsilon RI Signaling	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2;
	PTPN11; AKT2; PIK3CA; SYK; PRKCI;
	PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3;
	MAPK10; KRAS; MAPK13; PRKCD; MAPK9;
	PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1;
	AKT3; VAV3; PRKCA
G-Protein Coupled	PRKCE; RAP1A; RGS16; MAPK1; GNAS;
Receptor Signaling	AKT2; IKBKB; PIK3CA; CREB1; GNAQ;
	NFKB2; CAMK2A; PIK3CB; PIK3C3; MAPK3;
	KRAS; RELA; SRC; PIK3C2A; RAF1; IKBKG;
	RELB; FYN; MAP2K2; AKT1; PIK3R1;
	CHUK; PDPK1; STAT3; MAP2K1; NFKB1;
	BRAF; ATF4; AKT3; PRKCA
Inositol Phosphate	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN;
Metabolism	GRK6; MAPK1; PLK1; AKT2; PIK3CA; CDK8;
	PIK3CB; PIK3C3; MAPK8; MAPK3; PRKCD;
	PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;

	DYRK1A; MAP2K2; PIP5K1A; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1;
	BRAF; SGK
PDGF Signaling	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA;
	FOS; PIK3CB;PIK3C3; MAPK8; CAV1; ABL1;
	MAPK3; KRAS; SRC; PIK3C2A; PDGFRB;
	RAF1; MAP2K2; JAK1; JAK2; PIK3R1;
	PDGFRA; STAT3; SPHK1; MAP2K1; MYC;
	JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF Signaling	ACTN4; ROCK1; KDR; FLT1; ROCK2;
	MAPK1; PGF; AKT2; PIK3CA; ARNT; PTK2;
	BCL2; PIK3CB; PIK3C3; BCL2L1; MAPK3;
	KRAS; HIF1A; NOS3; PIK3C2A; PXN; RAF1;
	MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1;
	SFN; VEGFA; AKT3; FOXO1; PRKCA
Natural Killer Cell	PRKCE; RAC1; PRKCZ; MAPK1; RAC2;
Signaling	PTPN11; KIR2DL3; AKT2; PIK3CA; SYK;
	PRKCI; PIK3CB; PIK3C3; PRKD1; MAPK3;
	KRAS; PRKCD; PTPN6; PIK3C2A; LCK;
	RAF1; FYN; MAP2K2; PAK4; AKT1; PIK3R1;
	MAP2K1; PAK3; AKT3; VAV3; PRKCA
Cell Cycle: G1/S	HDAC4; SMAD3; SUV39H1; HDAC5;
Checkpoint Regulation	CDKN1B; BTRC; ATR; ABL1; E2F1; HDAC2;
	HDAC7A; RB1; HDAC11; HDAC9; CDK2;
	E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC;
	NRG1; GSK3B; RBL1; HDAC6
T Cell Receptor	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA;
Signaling	FOS; NFKB2; PIK3CB; PIK3C3; MAPK8;
	MAPK3; KRAS; RELA, PIK3C2A; BTK; LCK;
	RAF1; IKBKG; RELB, FYN; MAP2K2;
	PIK3R1; CHUK; MAP2K1; NFKB1; ITK;

	BCL10; JUN; VAV3
Death Receptor Signaling	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB;
	FADD; FAS; NFKB2; BCL2; MAP3K14;
	MAPK8; RIPK1; CASP8; DAXX; TNFRSF10B;
	RELA; TRAF2; TNF; IKBKG; RELB; CASP9;
	CHUK; APAF1; NFKB1; CASP2; BIRC2;
	CASP3; BIRC3
FGF Signaling	RAC1; FGFR1; MET; MAPKAPK2; MAPK1;
	PTPN11; AKT2; PIK3CA; CREB1; PIK3CB;
	PIK3C3; MAPK8; MAPK3; MAPK13; PTPN6;
	PIK3C2A; MAPK14; RAF1; AKT1; PIK3R1;
	STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
GM-CSF Signaling	LYN; ELK1; MAPK1; PTPN11; AKT2;
	PIK3CA; CAMK2A; STAT5B; PIK3CB;
	PIK3C3; GNB2L1; BCL2L1; MAPK3; ETS1;
	KRAS; RUNX1; PIM1; PIK3C2A; RAF1;
	MAP2K2; AKT1; JAK2; PIK3R1; STAT3;
	MAP2K1; CCND1; AKT3; STAT1
Amyotrophic Lateral	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1;
Sclerosis Signaling	CAPN2; PIK3CA; BCL2; PIK3CB; PIK3C3;
	BCL2L1; CAPN1; PIK3C2A; TP53; CASP9;
	PIK3R1; RAB5A; CASP1; APAF1; VEGFA;
	BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat Signaling	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA;
	STAT5B; PIK3CB; PIK3C3; MAPK3; KRAS;
	SOCS1; STAT5A; PTPN6; PIK3C2A; RAF1;
	CDKN1A; MAP2K2; JAK1; AKT1; JAK2;
	PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;
	STAT1
Nicotinate and	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6;
Nicotinamide	MAPK1; PLK1; AKT2; CDK8; MAPK8;

Metabolism	MAPK3; PRKCD; PRKAA1; PBEF1; MAPK9;
	CDK2; PIM1; DYRK1A; MAP2K2; MAP2K1;
	PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Chemokine Signaling	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1;
	GNAQ; CAMK2A; CXCL12; MAPK8; MAPK3;
	KRAS; MAPK13; RHOA; CCR3; SRC;
	PPP1CC; MAPK14; NOX1; RAF1; MAP2K2;
	MAP2K1; JUN; CCL2; PRKCA
IL-2 Signaling	ELK1; MAPK1; PTPN11; AKT2; PIK3CA;
	SYK; FOS; STAT5B; PIK3CB; PIK3C3;
	MAPK8; MAPK3; KRAS; SOCS1; STAT5A;
	PIK3C2A: LCK; RAF1; MAP2K2; JAK1;
	AKT1; PIK3R1; MAP2K1; JUN; AKT3
Synaptic Long Term	PRKCE; IGF1; PRKCZ; PRDX6; LYN;
Depression	MAPK1; GNAS; PRKC1; GNAQ; PPP2R1A;
	IGF1R; PRKID1; MAPK3; KRAS; GRN;
	PRKCD; NOS3; NOS2A; PPP2CA; YWHAZ;
	RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Estrogen Receptor	TAF4B; EP300; CARM1; PCAF; MAPK1;
Signaling	NCOR2; SMARCA4; MAPK3; NRIP1; KRAS;
	SRC; NR3C1; HDAC3; PPARGC1A; RBM9;
	NCOA3; RAF1; CREBBP; MAP2K2; NCOA2;
	MAP2K1; PRKDC; ESR1; ESR2
Protein Ubiquitination	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1;
Pathway	NEDD4; CBL; UBE2I; BTRC; HSPA5; USP7;
	USP10; FBXW7; USP9X; STUB1; USP22;
	B2M; BIRC2; PARK2; USP8; USP1; VHL;
	HSP90AA1; BIRC3
IL-10 Signaling	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS;
	NFKB2; MAP3K14; MAPK8; MAPK13; RELA;
	MAPK14; TNF; IKBKG; RELB; MAP3K7;
	JAK1; CHUK; STAT3; NFKB1; JUN; IL1R1;

	IL6
VDR/RXR Activation	PRKCE; EP300; PRKCZ; RXRA; GADD45A;
	HES1; NCOR2; SP1; PRKC1; CDKN1B;
	PRKD1; PRKCD; RUNX2; KLF4; YY1;
	NCOA3; CDKN1A; NCOA2; SPP1; LRP5;
	CEBPB; FOXO1; PRKCA
TGF-beta Signaling	EP300; SMAD2; SMURF1; MAPK1; SMAD3;
	SMAD1; FOS; MAPK8; MAPK3; KRAS;
	MAPK9; RUNX2; SERPINE1; RAF1;
	MAP3K7; CREBBP; MAP2K2; MAP2K1;
	TGFBR1; SMAD4; JUN; SMAD5
Toll-like Receptor	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA;
Signaling	ELK1; IKBKB; FOS; NFKB2; MAP3K14;
	MAPK8; MAPK13; RELA; TLR4; MAPK14;
	IKBKG; RELB; MAP3K7; CHUK; NFKB1;
	TLR2; JUN
p38 MAPK Signaling	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1;
	FADD; FAS; CREB1; DDIT3; RPS6KA4;
	DAXX; MAPK13; TRAF2; MAPK14; TNF;
	MAP3K7; TGFBR1; MYC; ATF4; IL1R1; SRF;
	STAT1
Neurotrophin/TRK	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1;
Signaling	FOS; PIK3CB; PIK3C3; MAPK8; MAPK3;
	KRAS; PIK3C2A; RAF1; MAP2K2; AKT1;
	PIK3R1; PDPK1; MAP2K1; CDC42; JUN;
	ATF4
FXR/RXR Activation	INS; PPARA; FASN; RXRA; AKT2; SDC1;
	MAPK8; APOB; MAPK10; PPARG; MTTP;
	MAPK9; PPARGC1A; TNF; CREBBP; AKT1;
	SREBF1; FGFR4; AKT3; FOXO1
Synaptic Long Term	PRKCE; RAP1A; EP300; PRKCZ; MAPK1;
Potentiation	CREB1; PRKC1; GNAQ; CAMK2A; PRKD1;

	MAPK3; KRAS; PRKCD; PPP1CC; RAF1;
	CREBBP; MAP2K2; MAP2K1; ATF4; PRKCA
Calcium Signaling	RAP1A; EP300; HDAC4; MAPK1; HDAC5;
	CREB1; CAMK2A; MYH9; MAPK3; HDAC2;
	HDAC7A; HDAC11; HDAC9; HDAC3;
	CREBBP; CALR; CAMKK2; ATF4; HDAC6
EGF Signaling	ELK1; MAPK1; EGFR; PIK3CA; FOS;
	PIK3CB; PIK3C3; MAPK8; MAPK3; PIK3C2A;
	RAF1; JAK1; PIK3R1; STAT3; MAP2K1; JUN;
	PRKCA; SRF; STAT1
Hypoxia Signaling in the	EDN1; PTEN; EP300; NQO1; UBE21; CREB1;
Cardiovascular System	ARNT; HIF1A; SLC2A4; NOS3; TP53; LDHA;
	AKT1; ATM; VEGFA; JUN; ATF4; VHL;
	HSP90AA1
LPS/IL-1 Mediated	IRAK1; MYD88; TRAF6; PPARA; RXRA;
Inhibition	ABCA1, MAPK8; ALDH1A1; GSTP1; MAPK9;
of RXR Function	ABCB1; TRAF2; TLR4; TNF; MAP3K7;
	NR1H2; SREBF1; JUN; IL1R1
LXR/RXR Activation	FASN; RXRA; NCOR2; ABCA1; NFKB2;
	IRF3; RELA; NOS2A; TLR4; TNF; RELB;
	LDLR; NR1H2; NFKB1; SREBF1; IL1R1;
	CCL2; IL6; MMP9
Amyloid Processing	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2;
	CAPN2; CAPN1; MAPK3; MAPK13; MAPT;
	MAPK14; AKT1; PSEN1; CSNK1A1; GSK3B;
	AKT3; APP
IL-4 Signaling	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1;
	KRAS; SOCS1; PTPN6; NR3C1; PIK3C2A;
	JAK1; AKT1; JAK2; PIK3R1; FRAP1; AKT3;
	RPS6KB1
Cell Cycle: G2/M DNA	EP300; PCAF; BRCA1; GADD45A; PLK1;
Damage Checkpoint	BTRC; CHEK1; ATR; CHEK2; YWHAZ; TP53;

Regulation	CDKN1A; PRKDC; ATM; SFN; CDKN2A
Nitric Oxide Signaling in	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB;
the	PIK3C3; CAV1; PRKCD; NOS3; PIK3C2A;
Cardiovascular System	AKT1; PIK3R1; VEGFA; AKT3; HSP90AA1
Purine Metabolism	NME2; SMARCA4; MYH9; RRM2; ADAR;
	EIF2AK4; PKM2; ENTPD1; RAD51; RRM2B;
	TJP2; RAD51C; NT5E; POLD1; NME1
cAMP-mediated	RAP1A; MAPK1; GNAS; CREB1; CAMK2A;
Signaling	MAPK3; SRC; RAF1; MAP2K2; STAT3;
	MAP2K1; BRAF; ATF4
Mitochondrial	SOD2; MAPK8; CASP8; MAPK10; MAPK9;
Dysfunction	CASP9; PARK7; PSEN1; PARK2; APP; CASP3
Notch Signaling	HES1; JAG1; NUMB; NOTCH4; ADAM17;
	NOTCH2; PSEN1; NOTCH3; NOTCH1; DLL4
Endoplasmic Reticulum	HSPA5; MAPK8; XBP1; TRAF2; ATF6;
Stress Pathway	CASP9; ATF4; EIF2AK3; CASP3
Pyrimidine Metabolism	NME2; AICDA; RRM2; EIF2AK4; ENTPD1;
	RRM2B; NT5E; POLD1; NME1
Parkinson's Signaling	UCHL1; MAPK8; MAPK13; MAPK14; CASP9;
	PARK7; PARK2; CASP3
Cardiac & Beta	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA;
Adrenergic Signaling	PPP1CC; PPP2R5C
Glycolysis/Gluco-	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA;
neogenesis	HK1
Interferon Signaling	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1;
	IFIT3
Sonic Hedgehog	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B;
Signaling	DYRKIB
Glycerophospholipid	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Metabolism	
Phospholipid	PRDX6; PLD1; GRN; YWHAZ; SPHK1;
Degradation	SPHK2

Tryptophan Metabolism	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1;
	SIAH1
Lysine Degradation	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Nucleotide Excision	ERCC5; ERCC4; XPA; XPC; ERCC1
Repair Pathway	
Starch and Sucrose	UCHL1; HK2; GCK; GPI; HK1
Metabolism	
Aminosugars Metabolism	NQO1; HK2; GCK; HK1
Arachidonic Acid	PRDX6; GRN; YWHAZ; CYP1B1
Metabolism	
Circadian Rhythm	CSNK1E; CREB1; ATF4; NR1D1
Signaling	
Coagulation System	BDKRB1; F2R; SERPINE1; F3
Dopamine Receptor	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Signaling	
Glutathione Metabolism	IDH2; GSTP1; ANPEP; IDH1
Glycerolipid Metabolism	ALDH1A1; GPAM; SPHK1; SPHK2
Linoleic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Methionine Metabolism	DNMT1; DNMT3B; AHCY; DNMT3A
Pyruvate Metabolism	GLO1; ALDH1A1; PKM2; LDHA
Arginine and Proline	ALDH1A1; NOS3; NOS2A
Metabolism	
Eicosanoid Signaling	PRDX6; GRN; YWHAZ
Fructose and Mannose	HK2; GCK; HK1
Metabolism	
Galactose Metabolism	HK2; GCK; HK1
Stilbene, Coumarine and	PRDX6; PRDX1; TYR
Lignin Biosynthesis	
Antigen Presentation	CALR; B2M
Pathway	
Biosynthesis of Steroids	NQO1; DHCR7
Butanoate Metabolism	ALDH1A1; NLGN1

Citrate Cycle	IDH2; IDH1
Fatty Acid Metabolism	ALDH1A1; CYP1B1
Glycerophospholipid	PRDX6; CHKA
Metabolism	
Histidine Metabolism	PRMT5; ALDH1A1
Inositol Metabolism	ERO1L; APEX1
Metabolism of	GSTP1; CYP1B1
Xenobiotics	
by Cytochrome p450	
Methane Metabolism	PRDX6; PRDX1
Phenylalanine	PRDX6; PRDX1
Metabolism	
Propanoate Metabolism	ALDH1A1; LDHA
Selenoamino Acid	PRMT5; AHCY
Metabolism	
Sphingolipid Metabolism	SPHK1; SPHK2
Aminophosphonate	PRMT5
Metabolism	
Androgen and Estrogen	PRMT5
Metabolism	
Ascorbate and Aldarate	ALDH1A1
Metabolism	
Bile Acid Biosynthesis	ALDH1A1
Cysteine Metabolism	LDHA
Fatty Acid Biosynthesis	FASN
Glutamate Receptor	GNB2L1
Signaling	
NRF2-mediated	PRDX1
Oxidative	
Stress Response	
Pentose Phosphate	GPI
Pathway	

Pentose and Glucuronate	UCHL1
Interconversions	
Retinol Metabolism	ALDH1A1
Riboflavin Metabolism	TYR
Tyrosine Metabolism	PRMT5, TYR
Ubiquinone Biosynthesis	PRMT5
Valine, Leucine and	ALDH1A1
Isoleucine Degradation	
Glycine, Serine and	CHKA
Threonine Metabolism	
Lysine Degradation	ALDH1A1
Pain/Taste	TRPM5; TRPA1
Pain	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2;
	Grk2; Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b;
	TRPM5; Prkaca; Prkacb; Prkar1a; Prkar2a
Mitochondrial Function	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Developmental	BMP-4; Chordin (Chrd); Noggin (Nog); WNT
Neurology	(Wnt2; Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6;
	Wnt7b; Wnt8b; Wnt9a; Wnt9b; Wnt10a;
	Wnt10b; Wnt16); beta-catenin; Dkk-1; Frizzled
	related proteins; Otx-2; Gbx2; FGF-8; Reelin;
	Dab1; unc-86 (Pou4f1 or Brn3a); Numb; Reln

Embodiments of the invention also relate to methods and compositions related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct. 13, 2011-Medical). Specific aspects of tandem repeat sequences have been found to be responsible for more than twenty human diseases (New insights into repeat instability: role of RNA*DNA hybrids. McIvor E I, Polak U, Napierala M. RNA Biol. 2010 September-October; 7(5):551-8). The CRISPR-Cas system may be harnessed to correct these defects of genomic instability.

Several further aspects of the invention relate to correcting defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at health.nih.gov/topic/GeneticDisorders). The genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease. Alzheimer's Disease, Barth Syndrome, Batten Disease, CADASIL, Cerebellar Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly. These diseases are further described on the website of the National Institutes of Health under the subsection Genetic Brain Disorders.

In some embodiments, the condition may be neoplasia. In some embodiments, where the condition is neoplasia, the genes to be targeted are any of those listed in Table A (in this case PTEN asn so forth). In some embodiments, the condition may be Age-related Macular Degeneration. In some embodiments, the condition may be a Schizophrenic Disorder. In some embodiments, the condition may be a Trinucleotide Repeat Disorder. In some embodiments, the condition may be Fragile X Syndrome. In some embodiments, the condition may be a Secretase Related Disorder. In some embodiments, the condition may be a Prion—related disorder. In some embodiments, the condition may be ALS. In some embodiments, the condition may be a drug addiction. In some embodiments, the condition may be Autism. In some embodiments, the condition may be Alzheimer's Disease. In some embodiments, the condition may be inflammation. In some embodiments, the condition may be Parkinson's Disease.

Examples of proteins associated with Parkinson's disease include but are not limited to α-synuclein, DJ-1, LRRK2, PINK1, Parkin, UCHL1, Synphilin-1, and NURR1.

Examples of addiction-related proteins may include ABAT for example.

Examples of inflammation-related proteins may include the monocyte chemoattractant protein-1 (MCP1) encoded by the Ccr2 gene, the C—C chemokine receptor type 5 (CCR5) encoded by the Ccr5 gene, the IgG receptor IIB (FCGR2b, also termed CD32) encoded by the Fcgr2b gene, or the Fc epsilon R1g (FCER1g) protein encoded by the Fcer1g gene, for example.

Examples of cardiovascular diseases associated proteins may include IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin 12 (prostacyclin) synthase), MB (myoglobin), IL4 (interleukin 4), ANGPT1

(angiopoietin 1), ABCG8 (ATP-binding cassette, sub-family G (WHITE), member 8), or CTSK (cathepsin K), for example.

Examples of Alzheimer's disease associated proteins may include the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, or the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBA3 gene, for example.

Examples of proteins associated Autism Spectrum Disorder may include the benzodiazapine receptor (peripheral) associated protein 1 (BZRAP1) encoded by the BZRAP1 gene, the AF4/FMR2 family member 2 protein (AFF2) encoded by the AFF2 gene (also termed MFR2), the fragile X mental retardation autosomal homolog 1 protein (FXR1) encoded by the FXR1 gene, or the fragile X mental retardation autosomal homolog 2 protein (FXR2) encoded by the FXR2 gene, for example.

Examples of proteins associated Macular Degeneration may include the ATP-binding cassette, sub-family A (ABC1) member 4 protein (ABCA4) encoded by the ABCR gene, the apolipoprotein E protein (APOE) encoded by the APOE gene, or the chemokine (C—C motif) Ligand 2 protein (CCL2) encoded by the CCL2 gene, for example.

Examples of proteins associated Schizophrenia may include NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B, and combinations thereof.

Examples of proteins involved in tumor suppression may include ATM (ataxia telangiectasia mutated), ATR (ataxia telangiectasia and Rad3 related), EGFR (epidermal growth factor receptor), ERBB2 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2), ERBB3 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 3), ERBB4 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 4), Notch 1, Notch2, Notch 3, or Notch 4, for example.

Examples of proteins associated with a secretase disorder may include PSENEN (presenilin enhancer 2 homolog (*C. elegans*)), CTSB (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APH1B (anterior pharynx defective 1 homolog B (*C. elegans*)), PSEN2 (presenilin 2 (Alzheimer disease 4)), or BACE1 (beta-site APP-cleaving enzyme 1), for example.

Examples of proteins associated with Amyotrophic Lateral Sclerosis may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

Examples of proteins associated with prion diseases may include SODI (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

Examples of proteins related to neurodegenerative conditions in prion disorders may include A2M (Alpha-2-Macroglobulin), AATF (Apoptosis antagonizing transcription factor), ACPP (Acid phosphatase prostate), ACTA2 (Actin alpha 2 smooth muscle aorta), ADAM22 (ADAM metallopeptidase domain), ADORA3 (Adenosine A3 receptor), or ADRA1D (Alpha-1D adrenergic receptor for Alpha-1D adrenoreceptor), for example.

Examples of proteins associated with Immunodeficiency may include A2M [alpha-2-macroglobulin]; AANAT [arylalkylamine N-acetyltransferase]; ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1]; ABCA2 [ATP-binding cassette, sub-family A (ABC1), member 2]; or ABCA3 [ATP-binding cassette, sub-family A (ABC1), member 3]; for example.

Examples of proteins associated with Trinucleotide Repeat Disorders include AR (androgen receptor), FMR1 (fragile X mental retardation 1), HTT (huntingtin), or DMPK (dystrophia myotonica-protein kinase), FXN (frataxin), ATXN2 (ataxin 2). for example.

Examples of proteins associated with Neurotransmission Disorders include SST (somatostatin), NOS1 (nitric oxide synthase 1 (neuronal)), ADRA2A (adrenergic, alpha-2A-, receptor), ADRA2C (adrenergic, alpha-2C-, receptor), TACR1 (tachykinin receptor 1), or HTR2c (5-hydroxytryptamine (serotonin) receptor 2C), for example.

Examples of neurodevelopmental-associated sequences include A2BP1 [ataxin 2-binding protein 1], AADAT [aminoadipate aminotransferase], AANAT [arylalkylamine N-acetyltransferase], ABAT [4-aminobutyrate aminotransferase], ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1], or ABCA13 [ATP-binding cassette, sub-family A (ABC1), member 13], for example.

Further examples of preferred conditions treatable with the present system include may be selected from: Aicardi-Goutières Syndrome; Alexander Disease; Allan-Herndon-Dudley Syndrome; POLG-Related Disorders; Alpha-Mannosidosis (Type II and III); Alström Syndrome; Angelman; Syndrome; Ataxia-Telangiectasia; Neuronal Ceroid-Lipofuscinoses; Beta-Thalassemia; Bilateral Optic Atrophy and (Infantile) Optic Atrophy Type 1; Retinoblastoma (bilateral); Canavan Disease; Cerebrooculofacioskeletal Syndrome 1 [COFS1]; Cerebrotendinous Xanthomatosis; Cornelia de Lange Syndrome; MAPT-Related

Disorders; Genetic Prion Diseases; Dravet Syndrome; Early-Onset Familial Alzheimer Disease; Friedreich Ataxia [FRDA]; Fryns Syndrome; Fucosidosis; Fukuyama Congenital Muscular Dystrophy; Galactosialidosis; Gaucher Disease; Organic Acidemias; Hemophagocytic Lymphohistiocytosis; Hutchinson-Gilford Progeria Syndrome; Mucolipidosis II; Infantile Free Sialic Acid Storage Disease; PLA2G6-Associated Neurodegeneration; Jervell and Lange-Nielsen Syndrome; Junctional Epidermolysis Bullosa; Huntington Disease; Krabbe Disease (Infantile); Mitochondrial DNA-Associated Leigh Syndrome and NARP; Lesch-Nyhan Syndrome; LIS1-Associated Lissencephaly; Lowe Syndrome; Maple Syrup Urine Disease; MECP2 Duplication Syndrome; ATP7A-Related Copper Transport Disorders; LAMA2-Related Muscular Dystrophy; Arylsulfatase A Deficiency; Mucopolysaccharidosis Types I, II or III; Peroxisome Biogenesis Disorders, Zellweger Syndrome Spectrum; Neurodegeneration with Brain Iron Accumulation Disorders; Acid Sphingomyelinase Deficiency; Niemann-Pick Disease Type C; Glycine Encephalopathy; ARX-Related Disorders; Urea Cycle Disorders; COL1A 1/2-Related Osteogenesis Imperfecta; Mitochondrial DNA Deletion Syndromes; PLP1-Related Disorders; Perry Syndrome; Phelan-McDermid Syndrome; Glycogen Storage Disease Type II (Pompe Disease) (Infantile); MAPT-Related Disorders; MECP2-Related Disorders; Rhizomelic Chondrodysplasia Punctata Type 1; Roberts Syndrome; Sandhoff Disease; Schindler Disease—Type 1; Adenosine Deaminase Deficiency; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy, Infantile-Onset Spinocerebellar Ataxia; Hexosaminidase A Deficiency; Thanatophoric Dysplasia Type 1; Collagen Type VI-Related Disorders; Usher Syndrome Type I; Congenital Muscular Dystrophy; Wolf-Hirschhorn Syndrome; Lysosomal Acid Lipase Deficiency; and Xeroderma Pigmentosum.

As will be apparent, it is envisaged that the present system can be used to target any polynucleotide sequence of interest. Some examples of conditions or diseases that might be usefully treated using the present system are included in the Tables above and examples of genes currently associated with those conditions are also provided there. However, the genes exemplified are not exhaustive.

While the present invention has been described above in relation to CRISPR-Cas9 systems, the present invention also contemplates the use of other systems for introducing double stranded breaks into a target sequence is host cell genome followed by insertion of a sequence of interest by homologous recombination. As above, these systems include co-expression of an exogenous recombinase to increase the efficiency of homologous recombination.

In some embodiments, targeted zinc finger nucleases (ZFNs) are utilized to introduce double stranded breaks as a site for homologous recombination. See, e.g., Carroll et al., Genetics (2011) 188:773-782; Meyer et al., Proc. Nat'l. Acad. Sci. (2010) 107(34):15022-15026; Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. Nat Biotechnol 23:967–973; Geurts AM, et al. (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325:433; Mashimo T, et al. (2010) Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. PLoS One 5:e8870; Meng X, Noyes MB, Zhu LJ, Lawson ND,Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol 26:695–701; Rouet P, Smih F, Jasin M (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci USA 91:6064–6068; Hockemeyer D, et al. (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol 27:851–857; Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. Science 300:763; Santiago Y, et al. (2008) Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proc Natl Acad Sci USA 105:5809–5814; Urnov FD, et al. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435:646–651; each of which is incorporated herein by reference in its entirety. Zinc-finger nucleases (ZFN) link a DNA binding domain of the zinc-finger type to the nuclease domain of Fok I and enable the induction of double-strand breaks (DSBs) at preselected genomic sites. DSBs closed by the error-prone, nonhomologous end-joining (NHEJ) DNA repair pathway frequently exhibit nucleotide deletions and insertions at the cleavage site. The present invention addresses this problem by co-expression of an exogenous recombinase.

In some embodiments, targeted transcription activator-like effector (TALE) nucleases are utilized to introduce double stranded breaks as a site for homologous recombination. See, e.g., Shin et al., Development (2014) 141:3807-3818; Boch et al. (2009) Science 326, 1509-1512; and Moscou and Bogdanove (2009) Science 326, 1501; each of which is incorporated by reference herein in its entirety. In still other embodiments, targeted meganucleases are utilized. See, e.g., Mol Cell Biol. 1994 Dec;14(12):8096-106. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Rouet P1, Smih F, Jasin M.

Accordingly, in some embodiments the present invention provides methods of altering expression of at least one gene product and/or genome editing comprising:

introducing into a cell having a genome a) a nucleic acid molecule encoding an enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product in said genome of said cell, b) a nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break by homolgous recombinaation; and 3) a nucleic acid sequence encoding a recombinase; wherein components (a), (b) and (c) are expressed in said cell, whereby a targeted double stranded break is introduced into the genome of said host cell and said nucleic acid molecule encoding a nucleic acid sequence of interest is inserted by homologous recombination at said double stranded break to effect altered expression of at said at least one gene product and/or genome editing of said gene encoding said gene product of interest and wherein said enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product and said nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break do not naturally occur together. In some embodiments, the recombinase is a bacterial recombinase. In some embodiments, the bacterial recombinase is selected from the group consisting of RecA recombinase and UvsX recombinase. In some embodiments, the expression of two or more gene products is altered. In some embodiments, the enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product in said genome of said cell is selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases. In some embodiments, the Cas9 nuclease is part of a CRISPR-Cas system as described in detail above. The system may preferably further comprise a nucleotide sequence encoding a CRISPR-Cas system guide RNA that hybridizes with the target sequence. In some embodiments, components (a), (b) and (c) are operably associated with the same or different regulatory elements. In some embodiments, components (a), (b) and (c) are encoded by mRNA molecules. In some embodiments, components (a), (b) and (c) are located on the same or different expression vectors. In some embodiments, the expression vectors are one or more viral expression vectors. In some embodiments, the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors. In some embodiments, the nucleic acid sequences encoding components (a), (b) and (c) are codon optimized for expression in a eukaryotic cell. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian cell. In some embodiments, the mammalian cell is a human cell. In some embodiments, the expression of

one or more gene products is increased. In some embodiments, the expression of one or more gene products is decreased.

In further embodiments, the present invention provides an engineered, non-naturally occurring system for altering expression of a gene product and/or genome editing comprising:

one or more nucleic acid sequences comprising a) a nucleic acid molecule encoding an enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product in said genome of said cell, b) a nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break; and 3) a nucleic acid sequence encoding a recombinase; wherein components (a), (b) and (c) are expressed in a cell, whereby a targeted double stranded break is introduced into the genome of said cell and said nucleic acid molecule encoding a nucleic acid sequence of interest is inserted by homologous recombination at said double stranded break to effect altered expression of at said at least one gene product and/or genome editing of said gene encoding said gene product of interest and wherein said enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product and said nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break do not naturally occur together. In some embodiments, the recombinase is a bacterial recombinase. In some embodiments, the bacterial recombinase is selected from the group consisting of RecA recombinase and UvsX recombinase. In some embodiments, the expression of two or more gene products is altered. In some embodiments, the enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product in said genome of said cell is selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases. In some embodiments, the Cas9 nuclease is part of a CRISPR-Cas system as described in detail above. The system may preferably further comprise a nucleotide sequence encoding a CRISPR-Cas system guide RNA that hybridizes with the target sequence. In some embodiments, components (a), (b) and (c) are operably associated with the same or different regulatory elements. In some embodiments, components (a), (b) and (c) are encoded by mRNA molecules. In some embodiments, components (a), (b) and (c) are located on the same or different expression vectors. In some embodiments, the expression vectors are one or more viral expression vectors. In some embodiments, the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors. In some embodiments, the nucleic acid sequences encoding components (a), (b) and (c) are codon

optimized for expression in a eukaryotic cell. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian cell. In some embodiments, the mammalian cell is a human cell. In some embodiments, the expression of one or more gene products is increased. In some embodiments, the expression of one or more gene products is decreased. In some embodiments, the nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break by homologous recombination is at least 250, 500, or 1000 bases in length.

In some embodiments, the present invention provides a cell comprising the system described above. In some embodiments, the present invention provides for use of the system to treat a disease by altering expression of gene in a target cell or editing the genome of a target cell.

EXAMPLES

Example 1.

Currently, Cas9/CRISPR can be used to create insertions using the HDR pathway. Often the efficiency is limited to 5% in some cell types this efficiency is effectively 0%. This is largely because mammalian recombinases are not efficient at incorporating foreign DNA and some cell types do not encode any recombinases – this makes HDR practically impossible. When the RecA or UvsX recombinases are supplied in trans with Cas9 and a HR substrate, increased levels of HDR are observed. This discovery enables the development of novel customized therapeutic solutions.

A site-specific nuclease (Cas9 and the appropriate gRNA) was designed to target and catalyse the formation of dsDNA breaks at the EMX1 locus in the human genome (Figure 2A). In addition, an oligonucleotide with 5' and 3' regions that are homologus to the EMX1 genomic locus was designed. Between the 5' and 3' homology arms is a BamHI site. After transfection of the site-specific nuclease and the oligonucleotide substrate into HeLa cells, HDR occurrences are detected by performing PCR of the EMX1 region followed by a BamHI digest. NHEJ and uncleaved products are resistant to BamHI cleavage while HDR products are sensitive to BamHI digestion.

HeLa cells expressing either an empty vector, NLS-RecA, or NLS-UvsX (NLS refers to the nuclear localization signal from the SV40 large T-antigen) were plated on 6-well plates at a density of 4×10^5 cells/ well in 2 ml complete DMEM 24h prior to transfection. At the day of transfection the cells were 90% confluent. Reactions (250 µl) containing 2500 ng

plasmid DNA and either 0 pmol and 1000 pmol ss Oligonucleotides Emx1-1 U or Emx1-1 L (below) were diluted in 250 µl OPTI-MEM I reduced Serum Medium (Cat.no. 31985062, ThermoScientific, Waltham, MA, USA). Additionally 10 µl LIPOFECTAMINE 2000 transfection reagent (Cat.no. 11668500, ThermoScientific, Waltham, MA, USA) were diluted in 250 µl OPTI-MEM I reduced Serum Medium (Cat.no. 31985062, ThermoScientific, Waltham, MA, USA) and incubated for 5 min at room temperature to allow complexes to form. After the incubation both mixtures were combined and gently mixed. After an incubation of 20 min at room temperature the 500 µl mixture was added to each well. The plate was gently rocked and incubated at 37°C for 48 h. Fourty eight hours after transfection DNA from transfected cells was isolated. PCR reactions (50 µl) containing 150 ng genomic DNA, 500 µM dNTPs, 1 mM MgCl2, 1.5% dimethylsulfoxide, 25 pmol of each primer, and 1 unit PhusionHF DNA polymerase were combined and subjected in a thermocycler (TC512, Keison Products, Grants Pass, Oregon, USA) using the following conditions 98oC for 120 seconds, 40 cycles of 98°C for 5 sec, 60.6°C for 10 seconds, 72°C for 20 seconds; a final extension at 72°C for 420 sec.

The PCR reactions were purified using the QIAquick PCR Purification Kit (ID 28104, Qiagen, Hilden, Germany) and concentrations were measured using the NanoDrop (ND-2000, ThermoScientific, Waltham, MA, USA). Restriction digestions (50 µl) containing 500 ng DNA of the purified PCR reaction, 10 units BamHI-HF and 1x CutSmart Buffer were incubated for 2 h at 37°C. The entire reaction was resolved on an 8% polyacrylamid TBE gel (8% Polyacrylamide, 15% Glycerol, 1x TBE, 10% APS, TEMED) and electrophoresed overnight at 25 V. The following day gels were stained with in 200 ml (89 mM Tris borate, 2mM EDTA, pH 8.2-8.4) supplemented with 0.5 ug/ml Ethidiumbromide. Images were taken using an Alphalmager HP (Cat.no. 92-13824-00, ProteinSimple, San Jose, Ca, USA). The cleavage intensity was measured by measuring the intensity of cleavage bands and PCR amplicon by ImageJ compared to the marker.

HR_Oligo_Emx1-1U-BamHI (SEQ ID NO:5)
ATTGCCACGA AGCAGGCCAA TGGGGAGGAC
ATCGATGTCA CCTCCAATGA CTAGGGATCC
GGGCAACCAC AAACCCACGA GGGCAGAGTG
CTGCTTGCTG CTGGCCAGGC CCCTGCGTGG
HR_Oligo_Emx1-1L-BamHI (SEQ ID NO:6)
CCACGCAGGG GCCTGGCCAG CAGCAAGCAG

CACTCTGCCC TCGTGGGTTT GTGGTTGCCC
GGATCCCTAG TCATTGGAGG TGACATCGAT
GTCCTCCCCA TTGGCCTGCT TCGTGGCAAT

Emx1-1_Fwd2
CCATCCCCTTCTGTGAATGT (SEQ ID NO:7)
Emx1-1_Rev2
GGAGATTGGAGACACGGAGA (SEQ ID NO:8)

Transfection of the site-specific nuclease and the oligonucleotide substrate (WT) yields 0-5.84% HDR products in transfected cells (Figure 2B and 2C). When RecA or UvsX is co-expressed with the site-specific nuclease and the oligonucleotide, we observed 8.24-16.3% and 12.7-33.88% HDR products (Figure 2B and 2C). These results suggest that co-expression of RecA or UvsX substantially improves HDR. There are several methods to promote HDR; however, these results indicate that the present system is superior to these methods. These other methods inhibit NHEJ, while the present system targets the HDR pathway directly. In contrast to methods that inhibit NHEJ, the present system is unique because it can be used in combination with inhibitors of NHEJ. The present system, when used in combination with inhibitors of NHEJ, is likely improve HDR to an even greater extent. The demonstrated improvement in HDR suggests that the present system has potential utility in human gene therapy in the near future.

CLAIMS

What is claimed is:

1. A method of altering expression of at least one gene product and/or genome editing comprising:

introducing into a cell a) an enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product in said genome of said cell, b) a nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break by homologous recombination; and c) a recombinase;

wherein components (a), (b) and (c) are expressed in said cell, whereby a targeted double stranded break is introduced into the genome of said host cell and said nucleic acid molecule encoding a nucleic acid sequence of interest is inserted by homologous recombination at said double stranded break to effect altered expression of at said at least one gene product and/or genome editing of said gene encoding said gene product of interest and wherein said enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product and said nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break do not naturally occur together.

2. The method of claim 1, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

3. The method of claim 1, wherein said recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

4. The method of claim 1, wherein said recombinase is introduced via a nucleic acid vector.

5. The method of claim 1, wherein the expression of two or more gene products is altered.

6. The method of claim 1, wherein said enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product in said genome of said cell is selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases.

7. The method of claim 6, wherein said Cas9 nuclease is part of a CRISPR-Cas system further comprising a nucleotide sequence encoding a CRISPR-Cas system guide RNA that hybridizes with the target sequence.

8. The method of claim 1, wherein said enzyme that introduces a double stranded break in a specific targeted sequence is introduced by a nucleic acid vector.

9. The method of claim 7, wherein the CRISPR -Cas system further comprises one or more nuclear localization signal(s) (NLS(s)).

10. The method of claim 7, wherein the CRISPR -Cas system comprises a trans-activating cr (tracr) sequence.

11. The method of claim 7, wherein the guide RNAs comprise a guide sequence fused to a tracr sequence.

12. The method of claim 1, wherein components (a), (b) and (c) are operably associated with the same or different regulatory elements.

13. The method of claim 1, wherein components (a), (b) and (c) are encoded by mRNA molecules.

14. The method of claim 1, wherein components (a), (b) and (c) are located on the same or different expression vectors.

15. The method of claim 14, wherein said expression vectors are one or more viral expression vectors.

16. The method of claim 1, wherein the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

17. The method of claim 1, wherein the nucleic acid sequences encoding components (a), (b) and (c) are codon optimized for expression in a eukaryotic cell.

18. The method of claim 1, wherein said cell is a eukaryotic cell.

19. The method of claim 18, wherein said eukaryotic cell is a mammalian cell.

20. The method of claim 19, wherein said mammalian cell is a human cell.

21. The method of claim 1, wherein the expression of one or more gene products is increased.

22. The method of claim 1, wherein the expression of one or more gene products is decreased.

23. An engineered, non-naturally occurring system for altering expression of a gene product and/or genome editing comprising:

one or more nucleic acid sequences comprising a) a nucleic acid molecule encoding an enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product in said genome of said cell, b) a nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break; and c) a nucleic acid sequence encoding a recombinase;

wherein components (a), (b) and (c) are expressed in a cell, whereby a targeted double stranded break is introduced into the genome of said cell and said nucleic acid molecule encoding a nucleic acid sequence of interest is inserted by homologous recombination at said double stranded break to effect altered expression of at said at least one gene product and/or genome editing of said gene encoding said gene product of interest and wherein said enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product and said nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break do not naturally occur together.

24. The system of claim 23, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

25. The system of claim 23, wherein said bacterial recombinase is selected from the group consisting of Rad 51 recombinase, RecA recombinase and UvsX recombinase.

26. The system of claim 23, wherein the expression of two or more gene products is altered.

27. The system of claim 23, wherein said enzyme that introduces a double stranded break in a specific targeted sequence in a gene encoding said gene product in said genome of said cell is selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases.

28. The system of claim 27, wherein said Cas9 nuclease is part of a CRISPR-Cas system further comprising a nucleotide sequence encoding a CRISPR -Cas system guide RNA that hybridizes with the target sequence.

29. The system of claim 27, wherein the CRISPR -Cas system further comprises one or more nuclear localization signal(s) (NLS(s)).

30. The system of claim 27, wherein the CRISPR -Cas system comprises a trans-activating cr (tracr) sequence.

31. The system of claim 27, wherein the guide RNAs comprise a guide sequence fused to a tracr sequence.

32. The system of claim 23, wherein components (a), (b) and (c) are operably associated with the same or different regulatory elements.

33. The system of claim 23, wherein components (a), (b) and (c) are encoded by mRNA molecules.

34. The system of claim 23, wherein components (a), (b) and (c) are located on the same or different expression vectors.

35. The system of claim 34, wherein said expression vectors are one or more viral expression vectors.

36. The system of claim 23, wherein the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

37. The system of claim 23, wherein the nucleic acid sequences encoding components (a), (b) and (c) are codon optimized for expression in a eukaryotic cell.

38. The system of claim 23, wherein said cell is a eukaryotic cell.

39. The system of claim 38, wherein said eukaryotic cell is a mammalian cell.

40. The system of claim 39, wherein said mammalian cell is a human cell.

41. The system of claim 23, wherein the expression of one or more gene products is increased.

42. The system of claim 23, wherein the expression of one or more gene products is decreased.

43. The system of claim 23, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said double stranded break by homologous recombination is at least 250, 500, or 1000 bases in length.

44. A cell comprising the system of any one of claims 23 to 43.

45. Use of the system of any one of claims 23 to 43 to treat a disease by altering expression of gene in a target cell or editing the genome of a target cell.

46. A method of altering expression of at least one gene product comprising:

introducing into a cell containing and expressing a DNA molecule having a target sequence and encoding the gene product an engineered, non-naturally occurring Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CAS associated (Cas) system comprising one or more viral vectors comprising:

a) a first regulatory element operable in said cell operably linked to at least one nucleotide sequence encoding a CRISPR -Cas system guide RNA that hybridizes with the target sequence,

b) a second regulatory element operable in said cell operably linked to a nucleotide sequence encoding a Cas9 protein, and

c) a third regulatory element operable in said cell operably linked to a nucleotide sequence encoding a recombinase,

wherein components (a), (b), and (c) are located on same or different vectors of the system, whereby the guide RNA targets the target sequence and the Cas9 protein cleaves the DNA molecule, whereby expression of the at least one gene product is altered; and, wherein the Cas9 protein and the guide RNA do not naturally occur together.

47. The method of claim 46, wherein said Cas9 protein is a Type-II Cas9 protein.

48. The method of claim 46, wherein said Cas9 protein is a nickase.

49. The method of claim 46, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

50. The method of claim 49, wherein said bacterial recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

51. The method of claim 46, wherein the expression of two or more gene products is altered.

52. The method of claim 46, wherein the CRISPR -Cas system further comprises one or more nuclear localization signal(s) (NLS(s)).

53. The method of claim 46, wherein the CRISPR -Cas system comprises a trans-

activating cr (tracr) sequence.

54. The method of claim 46, wherein the guide RNAs comprise a guide sequence fused to a tracr sequence.

55. The method of claim 46, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.

56. The method of claim 46, wherein the cell is a eukaryotic cell.

57. The method of claim 56, wherein said eukaryotic cell is a mammalian cell.

58. The method of claim 57, wherein said mammalian cell is a human cell.

59. The method of claim 46, wherein the expression of one or more gene products is increased.

60. The method of claim 46, wherein the expression of one or more gene products is decreased.

61. The method of claim 46, wherein the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

62. A CRISPR-Cas system-mediated genome editing method comprising:

introducing into a cell containing and expressing a DNA molecule having a target sequence and encoding at least one gene product an engineered, non-naturally occurring CRISPR -Cas system comprising one or more vectors comprising:

a) a first regulatory element operable in said cell operably linked to at least one nucleotide sequence encoding a CRISPR -Cas system guide RNA that hybridizes with the target sequence,

b) a second regulatory element operable in said cell operably linked to a nucleotide sequence encoding a Cas9 protein, and

c) a third regulatory element operable in said cell operably linked to a nucleotide sequence encoding a recombinase,

wherein components (a), (b) and (c) are located on same or different vectors of the system, whereby expression of the at least one gene product is altered through the CRISPR -Cas system acting as to the DNA molecule comprising the guide RNA directing sequence-specific binding of the CRISPR -Cas system, whereby there is genome editing; and, wherein the Cas9 protein and the guide RNA do not naturally occur together.

63. The method of claim 62, wherein said Cas9 protein is a Type II Cas9 protein.

64. The method of claim 62, wherein said Cas9 protein is a nickase.

65. The method of claim 62, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

66. The method of claim 65, wherein said bacterial recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

67. The method of claim 62, wherein the expression of two or more gene products is altered.

68. The method of claim 62, wherein the CRISPR -Cas system further comprises one or more NLS(s).

69. The method of claim 62, wherein the CRISPR -Cas system comprises a tracr sequence.

70. The method of claim 62, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.

71. The method of claim 62, wherein the cell is a eukaryotic cell.

72. The method of claim 71, wherein said eukaryotic cell is a mammalian cell.

73. The method of claim 72, wherein said mammalian cell is a human cell.

74. The method of claim 62, wherein the expression of one or more gene products is increased.

75. The method of claim 62, wherein the expression of one or more gene products is decreased.

76. An engineered, non-naturally occurring CRISPR-Cas system comprising:

one or more vectors comprising:

a) a first regulatory element operable in a cell operably linked to at least one nucleotide sequence encoding a CRISPR -Cas system guide RNA that hybridizes with a target sequence of a DNA molecule in a eukaryotic cell that contains the DNA molecule, wherein the DNA molecule encodes and the eukaryotic cell expresses at least one gene product,

b) a second regulatory element operable in said cell operably linked to a nucleotide sequence encoding a Cas9 protein, and

c) a third regulatory element operable in said cell operably linked to a nucleotide sequence encoding a recombinase,

wherein components (a), (b) and (c) are located on same or different vectors of the system, whereby the guide RNA targets and hybridizes with the target sequence and the Cas9 protein cleaves the DNA molecule, whereby expression of the at least one gene product is altered; and, wherein the Cas9 protein and the guide RNA do not naturally occur together.

77. The system of claim 76, wherein said Cas9 protein is a Type II Cas9 protein.

78. The system of claim 76, wherein said Cas9 protein is a nickase.

79. The system of claim 76, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

80. The system of claim 79, wherein said bacterial recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

81. The system of claim 76, wherein said one or more vectors are viral vectors.

82. The system of claim 81, wherein the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

83. The system of claim 76, wherein the CRISPR -Cas system further comprises one or more NLS(s).

84. The system of claim 76, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.

85. The system of claim 76, wherein the CRISPR -Cas system comprises a tracr sequence.

86. The system of claim 76, wherein the cell is a eukaryotic cell.

87. The system of claim 86, wherein said eukaryotic cell is a mammalian cell.

88. The system of claim 86, wherein said mammalian cell is a human cell.

89. The system of claim 76, wherein the expression of one or more gene products is increased.

90. The system of claim 76, wherein the expression of one or more gene products is decreased.

91. A cell comprising the system of any one of claims 76 to 90.

92. Use of the system of any one of claims 76 to 90 to treat a disease by altering expression of gene in a target cell or editing the genome of a target cell.

93. An engineered, non-naturally occurring CRISPR-Cas system comprising:
one or more vectors comprising:

a) a first regulatory element operable in a cell operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA that hybridizes with a target sequence of a DNA molecule in a eukaryotic cell that contains the DNA molecule, wherein the DNA molecule encodes and the eukaryotic cell expresses at least one gene product,

b) a second regulatory element operable in said cell operably linked to a nucleotide sequence encoding a Cas9 protein, and

c) a third regulatory element operable in said cell operably linked to a nucleotide sequence encoding a recombinase,

wherein components (a), (b) and (c) are located on same or different vectors of the system, whereby expression of the at least one gene product is altered through the CRISPR-Cas system acting as to the DNA molecule comprising the guide RNA directing sequence-specific binding of the CRISPR-Cas system, whereby there is genome editing; and, wherein the Cas9 protein and the guide RNA do not naturally occur together.

94. The system of claim 93, wherein said Cas9 protein is a Type II Cas9 protein.

95. The system of claim 93, wherein said Cas9 protein is a nickase.

96. The system of claim 93, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

97. The system of claim 96, wherein said bacterial recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

98. The system of claim 93, wherein said one or more vectors are viral vectors.

99. The system of claim 98, wherein the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

100. The system of claim 93, wherein the CRISPR -Cas system further comprises one or more NLS(s).

101. The system of claim 93, wherein the Cas9 protein is codon optimized for expression in

the eukaryotic cell.

102. The system of claim 93, wherein the CRISPR -Cas system comprises a tracr sequence.

103. The system of claim 93, wherein the cell is a eukaryotic cell.

104. The system of claim 103, wherein said eukaryotic cell is a mammalian cell.

105. The system of claim 104, wherein said mammalian cell is a human cell.

106. The system of claim 93, wherein the expression of one or more gene products is increased.

107. The system of claim 93, wherein the expression of one or more gene products is decreased.

108. A cell comprising the system of any one of claims 93 to 107.

109. Use of the system of any one of claims 93 to 107 to treat a disease by altering expression of gene in a target cell or editing the genome of a target cell.

110. A method of altering expression of at least one gene product and/or genome editing comprising:

introducing into a cell having a genome a nucleic acid filament comprising a single stranded nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome, said single stranded nucleic acid molecule having bound thereto a multimeric recombinase complex and wherein said nucleic of interest comprises 5' and 3' flanking regions that are homologous to a genomic target sequence encoding said gene product, said 5' and 3' flanking sequences flank an insert sequence that is different from said genomic target sequence, and wherein said nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome and said multimeric recombinase complex do not naturally occur together;

whereby said nucleic acid sequence of interest is inserted by homologous recombination into said genome to effect altered expression of at said at least one gene product and/or genome editing of said gene encoding said gene product of interest.

111. The method of claim 110, wherein said filament nucleic acid filament is synthesized in vitro by incubating said single stranded nucleic acid of interest with a recombinase so that a multimeric recombinase complex is formed on said single stranded nucleic acid of interest.

112. The method of claim 111, further comprising incubating said single stranded nucleic acid of interest with a nucleotide.

113. The method of claim 112, wherein said nucleotide is selected from the group consisting of a nucleotide triphosphate or analog.

114. The method of claim 113, wherein said nucleotide triphosphate or analog is selected from the group consisting of adenosine triphosphate, adenosine monophosphate, adenosine diphosphate, adenosine triphosphate-γS, adenosine monophosphate-PNP, and adenosine diphosphate-AlF_4.

115. The method of claim 110, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

116. The method of claim 110, wherein said recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

117. The method of claim 110, wherein said 5' and 3' flanking sequences are greater than 50 bases in length.

118. The method of claim 110, wherein said 5' and 3' flanking sequences are greater than 100 bases in length.

119. The method of claim 110, wherein said 5' and 3' flanking sequences are greater than 200 bases in length.

120. The method of claim 110, wherein said 5' and 3' flanking sequences are greater than 500 bases in length.

121. The method of claim 110, wherein said 5' and 3' flanking sequences are greater than 1000 bases in length.

122. The method of claim 110, wherein said 5' and 3' flanking sequences are from about 20 to about 1000 bases in length.

123. The method of claim 110, wherein said 5' and 3' flanking sequences are from about 100 to about 1000 bases in length.

124. The method of claim 110, wherein said single stranded nucleic acid is single stranded DNA.

125. The method of claim 110, wherein the expression of two or more gene products is altered.

126. The method of claim 110, further comprising introducing a break in said targeted sequence in a gene encoding said gene product in said genome of said cell.

127. The method of claim 126, wherein said break is a double stranded break or a single stranded break.

128. The method of claim 126, wherein said break is introduced by an enzyme selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases, and a Type I restriction endonuclease, Type II restriction endonuclease, Type III, restriction endonuclease, Type IV restriction endonuclease or nickase.

129. The method of claim 110, further comprising introducing into said cell a CRISPR -Cas system guide RNA that hybridizes with the target sequence.

130. The method of claim 120, wherein the CRISPR -Cas system comprises a trans-

activating cr (tracr) sequence.

131. The method of claim 130, wherein the guide RNAs comprise a guide sequence fused to a tracr sequence.

132. The method of claim 110, wherein said filament is introduced into said cell by electroporation.

133. The method of claim 110, wherein said cell is a eukaryotic cell.

134. The method of claim 133, wherein said eukaryotic cell is a mammalian cell.

135. The method of claim 134, wherein said mammalian cell is a human cell.

136. The method of claim 110, wherein the expression of one or more gene products is increased.

137. The method of claim 110, wherein the expression of one or more gene products is decreased.

138. The method of claim 110, wherein the insert sequence is inserted into a coding region of a gene of interest.

139. An engineered, non-naturally occurring system for altering expression of a gene product and/or genome editing comprising:

a nucleic acid filament comprising a single stranded nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome, said single stranded nucleic acid molecule having bound thereto a multimeric recombinase complex and wherein said nucleic of interest comprises 5' and 3' flanking regions that are homologous to a genomic target sequence encoding said gene product, said 5' and 3' flanking sequences flank an insert sequence that is different from said genomic target sequence, and wherein said single stranded nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome and said multimeric recombinase complex do not naturally occur together;

whereby when introduced into a cell having a genome said nucleic acid sequence of interest is inserted by homologous recombination into said genome to effect altered expression of at said at least one gene product and/or genome editing of said gene encoding said gene product of interest.

140. The system of claim 139, wherein said filament nucleic acid filament is synthesized in vitro by incubating said single stranded nucleic acid of interest with a recombinase so that a multimeric recombinase complex is formed on said single stranded nucleic acid of interest.

141. The system of claim 140, further comprising incubating said single stranded nucleic acid of interest with a nucleotide.

142. The system of claim 141, wherein said nucleotide is selected from the group consisting of a nucleotide triphosphate or analog.

143. The system of claim 142, wherein said nucleotide triphosphate or analog is selected from the group consisting of adenosine triphosphate, adenosine monophosphate, adenosine diphosphate, adenosine triphosphate-γS, adenosine monophosphate-PNP, and adenosine diphosphate-AlF_4.

144. The system of claim 139, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

145. The system of claim 139, wherein said recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

146. The system of claim 139, wherein said 5' and 3' flanking sequences are greater than 50 bases in length.

147. The system of claim 139, wherein said 5' and 3' flanking sequences are greater than 100 bases in length.

148. The system of claim 139, wherein said 5' and 3' flanking sequences are greater than 200 bases in length.

149. The system of claim 139, wherein said 5' and 3' flanking sequences are greater than 500 bases in length.

150. The system of claim 139, wherein said 5' and 3' flanking sequences are greater than 1000 bases in length.

151. The system of claim 139, wherein said 5' and 3' flanking sequences are from about 20 to about 1000 bases in length.

152. The system of claim 139, wherein said 5' and 3' flanking sequences are from about 100 to about 1000 bases in length.

153. The system of claim 139, wherein said single stranded nucleic acid is single stranded DNA.

154. The system of claim 139, wherein the expression of two or more gene products is altered.

155. The system of claim 139, further comprising reagents for introducing a break in said targeted sequence in a gene encoding said gene product in said genome of said cell.

156. The system of claim 155, wherein said break is a double stranded break or a single stranded DNA break.

157. The system of claim 155, wherein said reagent is an enzyme selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases, and a Type I restriction endonuclease, Type II restriction endonuclease, Type III, restriction endonuclease, Type IV restriction endonuclease or nickase.

158. The system of claim 139, further comprising a CRISPR-Cas system guide RNA that hybridizes with the target sequence.

159. The system of claim 158, wherein the CRISPR-Cas system comprises a trans-activating cr (tracr) sequence.

160. The system of claim 159, wherein the guide RNAs comprise a guide sequence fused to a tracr sequence.

161. The system of claim 139, wherein said filament is introduced into said cell by electroporation.

162. The system of claim 139, wherein said cell is a eukaryotic cell.

163. The system of claim 162, wherein said eukaryotic cell is a mammalian cell.

164. The system of claim 163, wherein said mammalian cell is a human cell.

165. The system of claim 139, wherein the expression of one or more gene products is increased.

166. The system of claim 139, wherein the expression of one or more gene products is decreased.

167. A cell comprising the system of any one of claims 139 to 166.

168. Use of the system of any one of claims 139 to 166 to treat a disease by altering expression of gene in a target cell or editing the genome of a target cell.

169. A method of altering expression of at least one gene product and/or genome editing comprising:

introducing into a cell a) a nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at a break in said genome by homologous recombination; and b) a recombinase;

wherein when components (a) and (b) are introduced or expressed in said cell, said nucleic acid molecule encoding a nucleic acid sequence of interest is inserted by homologous recombination at said break to effect altered expression of at said at least one gene product and/or genome editing of said gene encoding said gene product of interest and wherein said nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said break and said recombinase do not naturally occur together.

170. The method of claim 169, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

171. The method of claim 169, wherein said recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

172. The method of claim 169, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are greater than 50 bases in length.

173. The method of claim 169, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences greater than 100 bases in length.

174. The method of claim 169, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are greater than 200 bases in length.

175. The method of claim 169, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are greater than 500 bases in length.

176. The method of claim 169, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are greater than 1000 bases in length.

177. The method of claim 169, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences from about 20 to about 1000 bases in length.

178. The method of claim 169, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are from about 100 to about 1000 bases in length.

179. The method of claim 169, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest is single stranded DNA.

180. The method of claim 169, wherein the expression of two or more gene products is altered.

181. The method of claim 169, wherein said break is a double stranded break or a single stranded break.

182. The method of claim 181, wherein said break is introduced by an enzyme selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases.

183. The method of claim 181, wherein said break is introduced by a Type I, II, III or IV restriction endonuclease or nickase.

184. The method of claim 169, further comprising introducing into said cell a CRISPR - Cas system guide RNA that hybridizes with the target sequence.

185. The method of claim 184, wherein the CRISPR-Cas system comprises a trans-activating cr (tracr) sequence.

186. The method of claim 185, wherein the guide RNAs comprise a guide sequence fused to a tracr sequence.

187. The method of claim 169, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest is introduced into said cell by electroporation.

188. The method of claim 169, wherein said cell is a eukaryotic cell.

189. The method of claim 188, wherein said eukaryotic cell is a mammalian cell.

190. The method of claim 189, wherein said mammalian cell is a human cell.

191. The method of claim 169, wherein the expression of one or more gene products is increased.

192. The method of claim 169, wherein the expression of one or more gene products is decreased.

193. An engineered, non-naturally occurring system for altering expression of a gene product and/or genome editing comprising:

a) a nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at a break in said genome by homologous recombination; and b) a recombinase;

wherein when components (a) and (b) are introduced or expressed in said cell, said nucleic acid molecule encoding a nucleic acid sequence of interest is inserted by homologous recombination at said break to effect altered expression of at said at least one gene product and/or genome editing of said gene encoding said gene product of interest and wherein said nucleic acid molecule encoding a nucleic acid sequence of interest to be inserted into said genome at said break and said recombinase do not naturally occur together.

194. The system of claim 193, wherein said recombinase is selected from the group consisting of a bacterial recombinase, a viral recombinase and a mammalian recombinase.

195. The system of claim 193, wherein said recombinase is selected from the group consisting of Rad51 recombinase, RecA recombinase and UvsX recombinase.

196. The system of claim 193, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are greater than 50 bases in length.

197. The system of claim 193, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are greater than 100 bases in length.

198. The system of claim 193, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are greater than 200 bases in length.

199. The system of claim 193, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are greater than 500 bases in length.

200. The system of claim 193, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are greater than 1000 bases in length.

201. The system of claim 193, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is

different from said genomic target sequence, wherein said 5' and 3' flanking sequences are from about 20 to about 1000 bases in length.

202. The system of claim 193, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest comprises 5' and 3' sequences flanking an insert sequence that is different from said genomic target sequence, wherein said 5' and 3' flanking sequences are from about 100 to about 1000 bases in length.

203. The system of claim 193, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest is single stranded DNA.

204. The system of claim 193, wherein the expression of two or more gene products is altered.

205. The system of claim 193, further comprising reagents for introducing a break in said targeted sequence in a gene encoding said gene product in said genome of said cell.

206. The system of claim 193, wherein said break is a double stranded break or a single stranded DNA break.

207. The system of claim 206, wherein said reagent is an enzyme selected from the group consisting of Cas9 nuclease, meganucleases, Zinc finger (ZNF)-nucleases, and transcription activator-like effector (TALE)-nucleases.

208. The system of claim 206, wherein said break is introduced by a Type I, II, III or IV restriction endonuclease or nickase.

209. The system of claim 193, further comprising a CRISPR-Cas system guide RNA that hybridizes with the target sequence.

210. The system of claim 209, wherein the CRISPR-Cas system comprises a trans-activating cr (tracr) sequence.

211. The system of claim 159, wherein the guide RNAs comprise a guide sequence fused to a tracr sequence.

212. The system of claim 193, wherein said nucleic acid molecule encoding a nucleic acid sequence of interest is introduced into said cell by electroporation.

213. The system of claim 193, wherein said cell is a eukaryotic cell.

214. The system of claim 213, wherein said eukaryotic cell is a mammalian cell.

215. The system of claim 214, wherein said mammalian cell is a human cell.

216. The system of claim 193, wherein the expression of one or more gene products is increased.

217. The system of claim 193, wherein the expression of one or more gene products is decreased.

218. A cell comprising the system of any of claims 193 to 217.

219. Use of the system of any one of claims 193 to 217 to treat a disease by altering expression of gene in a target cell or editing the genome of a target cell.

FIG. 1

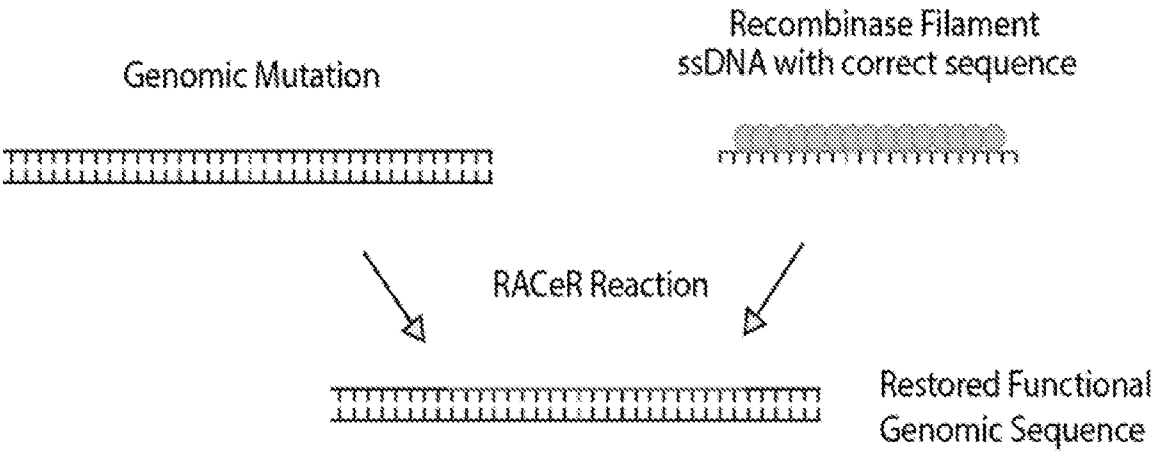
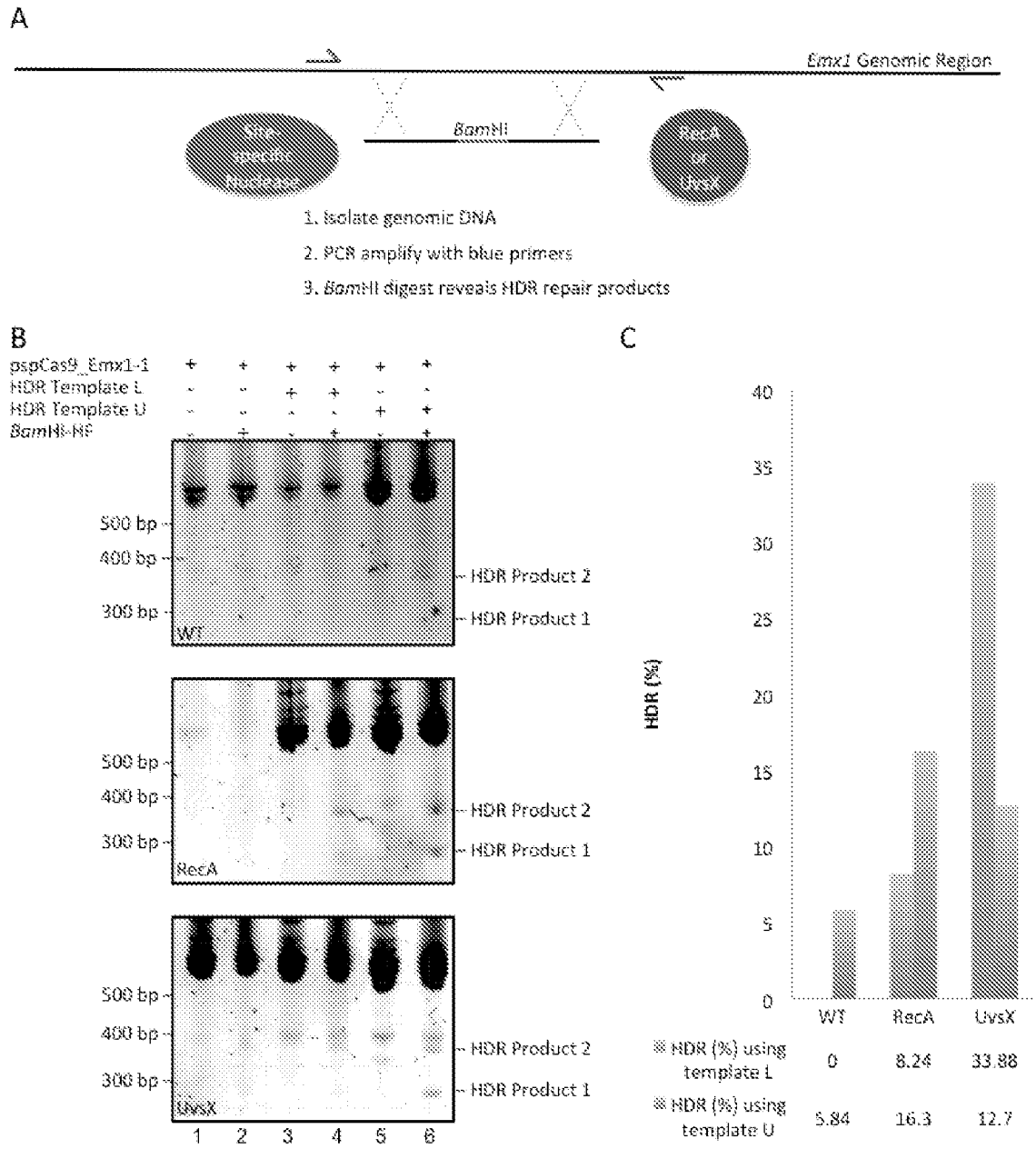

Genomic Mutation
Recombinase Filament
ssDNA with correct sequence
RACeR Reaction
Restored Functional
Genomic Sequence

FIG. 2


A
Emx1 Genomic Region
BamNI
RecA or UvsX
1. Isolate genomic DNA
2. PCR amplify with blue primers
3. BamNI digest reveals HDR repair products
B
pspCas9_Emx1-1
HDR Template L
HDR Template U
BamNI-NF
500 bp
400 bp
300 bp
HDR Product 2
HDR Product 1
WT
RecA
UvsX
1 2 3 4 5 6
C
HDR (%)
40
35
30
25
20
15
10
5
0
WT
RecA
UvsX
※ HDR (%) using template L 0 8.24 33.88
※ HDR (%) using template U 5.84 16.3 12.7

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2017/000967

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/90 C12N15/10 C12N15/63
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FAYU YANG ET AL: "CRISPR/Cas9- loxP -Mediated Gene Editing as a Novel Site-Specific Genetic Manipulation Tool", MOLECULAR THERAPY - NUCLEIC ACIDS, vol. 7, 25 April 2017 (2017-04-25), pages 378-386, XP055442108, GB ISSN: 2162-2531, DOI: 10.1016/j.omtn.2017.04.018 the whole document ----- -/--	1,2, 4-24, 26-45

[X] Further documents are listed in the continuation of Box C.

[X] 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
23 January 2018	26/03/2018

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Rutz, Berthold

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	XUN LIANG ET AL: "A CRISPR/Cas9 and Cre/Lox system-based express vaccine development strategy against re-emerging Pseudorabies virus", SCIENTIFIC REPORTS, vol. 6, no. 1, 18 January 2016 (2016-01-18), XP055442118, DOI: 10.1038/srep19176 the whole document	1,2, 4-14, 17-20, 22-24, 27-34, 37-45
X	MICHAEL E. PYNE ET AL: "Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in Escherichia coli", APPLIED AND ENVIRONMENTAL MICROBIOLOGY, vol. 81, no. 15, 22 May 2015 (2015-05-22), pages 5103-5114, XP055336375, ISSN: 0099-2240, DOI: 10.1128/AEM.01248-15 the whole document	1,2,4, 6-8,10, 12,14, 21,23, 24,27, 28,30, 32-34, 41-45
X	MARCELO C. BASSALO ET AL: "Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli", ACS SYNTHETIC BIOLOGY, vol. 5, no. 7, 12 April 2016 (2016-04-12), pages 561-568, XP055442201, Washington, DC,USA ISSN: 2161-5063, DOI: 10.1021/acssynbio.5b00187 the whole document	1,2,4, 6-8,10, 12,14, 21,23, 24,27, 28,30, 32-34, 41-45
X	LIN LIN ET AL: "Fusion of SpCas9 toE. coliRec A protein enhances CRISPR-Cas9 mediated gene knockout in mammalian cells", JOURNAL OF BIOTECHNOLOGY, ELSEVIER, AMSTERDAM, NL, vol. 247, 1 March 2017 (2017-03-01), pages 42-49, XP029951070, ISSN: 0168-1656, DOI: 10.1016/J.JBIOTEC.2017.02.024 the whole document	1-45
X	WO 2016/054326 A1 (GEN HOSPITAL CORP [US]) 7 April 2016 (2016-04-07) claims 16-21	1-4
X	WO 2014/204725 A1 (BROAD INST INC [US]; MASSACHUSETTS INST TECHNOLOGY [US]; HARVARD COLLE) 24 December 2014 (2014-12-24) paragraph [0151]	1-4

-/--

2

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2016/281111 A1 (COTTA-RAMUSINO CECILIA [US] ET AL) 29 September 2016 (2016-09-29) paragraphs [0971], [0973]	1-4
T	EIRIK ADIM MOREB ET AL: "Managing the SOS Response for Enhanced CRISPR-Cas-Based Recombineering in E. coli through Transient Inhibition of Host RecA Activity", ACS SYNTHETIC BIOLOGY, vol. 6, no. 12, 15 September 2017 (2017-09-15), pages 2209-2218, XP055442204, Washington, DC,USA ISSN: 2161-5063, DOI: 10.1021/acssynbio.7b00174	
A	JIAN-PING ZHANG ET AL: "Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage", GENOME BIOLOGY, vol. 18, no. 1, 20 February 2017 (2017-02-20), XP055399694, DOI: 10.1186/s13059-017-1164-8	1-45
A	WO 2016/073990 A2 (EDITAS MEDICINE INC [US]) 12 May 2016 (2016-05-12)	1-45

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2017/000967

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

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

   a. [X] forming part of the international application as filed:

      [X] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:

      [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2017/000967

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

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

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

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

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

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

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

see additional sheet

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

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

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

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

1-45

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2017/000967

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2016054326 A1	07-04-2016	AU 2015324935 A1	20-04-2017
		CA 2963080 A1	07-04-2016
		EP 3201340 A1	09-08-2017
		KR 20170061697 A	05-06-2017
		WO 2016054326 A1	07-04-2016
WO 2014204725 A1	24-12-2014	AU 2014281027 A1	28-01-2016
		CA 2915837 A1	24-12-2014
		CN 105492611 A	13-04-2016
		EP 3011030 A1	27-04-2016
		HK 1223645 A1	04-08-2017
		JP 2016521994 A	28-07-2016
		KR 20160034901 A	30-03-2016
		US 2016153006 A1	02-06-2016
		WO 2014204725 A1	24-12-2014
US 2016281111 A1	29-09-2016	EP 3274453 A2	31-01-2018
		US 2016281111 A1	29-09-2016
		WO 2016154579 A2	29-09-2016
WO 2016073990 A2	12-05-2016	AU 2015342749 A1	11-05-2017
		CA 2963820 A1	12-05-2016
		EP 3215617 A2	13-09-2017
		WO 2016073990 A2	12-05-2016

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-45

   method for altering expression of gene product or genome editing, comprising (i) DSB inducing enzmye, (ii) donor nucleic acid and (iii) recombinase; system comprising (i) DSB inducing enzmye, (ii) donor nucleic acid and (iii) recombinase; cell comprising the system; use of the system

   ---

2. claims: 46-109

   method for altering expression of gene product or genome editing, comprising (i) guide RNA, (ii) Cas9 protein and (iii) recombinase; system comprising (i) guide RNA, (ii) Cas9 protein and (iii) recombinase; cell comprising the system; use of the system

   ---

3. claims: 110-168

   method for altering expression of gene product or genome editing, comprising (i) donor nucleic acid bound to (ii) multimeric recombinase complex; system comprising(i) donor nucleic acid bound to (ii) multimeric recombinase complex; cell comprising the system; use of the system

   ---

4. claims: 169-219

   method for altering expression of gene product or genome editing, comprising (i) donor nucleic acid and (ii) recombinase; system comprising (i) donor nucleic acid and (ii) recombinase; cell comprising the system; use of the system

   ---