



(43) International Publication Date
31 August 2017 (31.08.2017)

- (51) **International Patent Classification:**
C12N 15/10 (2006.01) *C12N 15/90* (2006.01)
- (21) **International Application Number:**
PCT/US2017/018679
- (22) **International Filing Date:**
21 February 2017 (21.02.2017)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/298,129 22 February 2016 (22.02.2016) US
62/344,876 2 June 2016 (02.06.2016) US
- (71) **Applicant:** **CARIBOU BIOSCIENCES, INC.** [US/US];
2929 7th Street, Suite 105, Berkeley, CA 94710 (US).
- (72) **Inventors:** **CAPURSO, Daniel**; c/o Caribou Biosciences, Inc., 2929 7th Street, Suite 105, Berkeley, CA 94710 (US).
MAY, Andrew P.; c/o Caribou Biosciences, Inc., 2929 7th Street, Suite 105, Berkeley, CA 94710 (US). **VAN OVERBEEK, Megan**; c/o Caribou Biosciences, Inc., 2929 7th Street, Suite 105, Berkeley, CA 94710 (US).
- (74) **Agent:** **MCCLUNG, Barbara G.**; Chief Legal Officer, Caribou Biosciences, Inc., 2929 7th Street, Suite 105, Berkeley, CA 94710 (US).

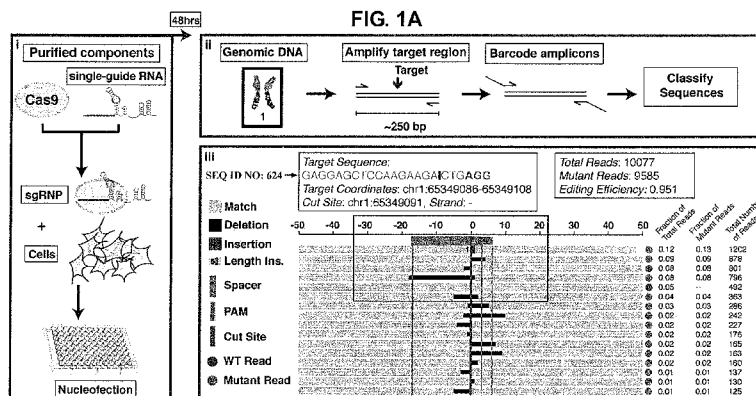
(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,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



(S7) Abstract: Methods for gene editing and predicting non-random editing events are described. The methods use Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems to characterize and manipulate DNA repair outcomes at Cas-initiated double-strand breaks (DSBs) to anticipate functional outcomes.

METHODS FOR MODULATING DNA REPAIR OUTCOMES

TECHNICAL FIELD

[0001] The present invention relates to DNA repair methods for modulating DNA repair outcomes, such as in gene editing applications.

BACKGROUND OF THE INVENTION

[0002] The CRISPR system has been widely adopted as a versatile tool for genome engineering in human cells (*see, e.g.,* Jinek *et al.*, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” *Science* (2012) 337:816–821; and Jinek *et al.*, “RNA-programmed genome editing in human cells” *Elife* (2013) 2:e00471). CRISPR-associated proteins, such as Cas9, are programmable and can be targeted to create chromosomal double-strand breaks (DSBs) at sites in genomic DNA by a single-guide RNA (sgRNA) molecule designed to complement the sequence of interest (Jinek *et al.*, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” *Science* (2012) 337:816–821). Following DSB generation, DNA repair machinery is recruited to promote DNA transactions at the break site and the outcome of this process dictates the functionality of the repair. In the absence of a donor template, this process has been thought to result in random DNA repair outcomes. Despite the critical role that DNA repair outcomes play in defining the nature of a genome edit, the specific insertion/deletion (indel) mutations that result from Cas-initiated DSB repair have not been well characterized to date. There remains an outstanding need, therefore, for detailed characterization of DNA repair outcomes at Cas-initiated DSBs to anticipate functional outcomes.

[0003] Gene inactivation can be achieved through repair of Cas/sgRNA-dependent DSBs at target sites by mutation-prone end-joining pathways (*e.g.*, classical non-homologous end-joining (c-NHEJ), alternative end-joining (alt-EJ)/microhomology mediated end-joining (MMEJ)) that produce mainly indels at the break site (Mandal *et al.*, “Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9” *Cell Stem Cell* (2014) 15:643–652; Hou *et al.*, “Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection” *Sci Rep* (2015) 5:15577; Canver *et al.*, “BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis” *Nature* (2015) 527:192–197). Repair through the alt-EJ/MMEJ pathway frequently results in large deletions between microhomology sites (Deriano

and Roth, “Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage” *Annual Review of Genetics* (2013) 47:433–455).

[0004] Biochemical and cell-based studies monitoring re-ligation of linearized plasmids suggest that c-NHEJ is an accurate process for blunt and cohesive ends (*see, e.g.*, Boulton and Jackson, “*Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways” (1996) *EMBO J.* 15:5093–5103; Feldmann *et al.*, “DNA double-strand break repair in cell-free extracts from Ku80-deficient cells: implications for Ku serving as an alignment factor in non-homologous DNA end joining” *Nucleic Acids Res* (2000) 28:2585–2596; Smith *et al.*, “Impact of DNA ligase IV on the fidelity of end joining in human cells” *Nucleic Acids Res* (2003) 31:2157–2167).

[0005] However, the repair of lymphocyte-specific programmed DSBs and DSBs generated from extrinsic agents such as radiomimetic drugs and ionizing radiation by c-NHEJ has been thought to result in random outcomes (Li *et al.*, “Multiple end joining mechanisms repair a chromosomal DNA break in fission yeast” *DNA Repair (Amst.)* (2012) 11:120–130; Dolan *et al.*, “Integrated Stochastic Model of DNA Damage Repair by Non-homologous End Joining and p53/p21-Mediated Early Senescence Signalling” *PLoS Comput. Biol.* (2015) 11:e1004246; Boboila *et al.*, “Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks” *Adv. Immunol.* (2012) 116:1–49). A study performed in yeast comparing the repair of DSBs with “ragged” ends induced by IR and “clean” ends induced by endonuclease cleavage revealed different processing requirements for break repair (Barlow *et al.*, “Differential regulation of the cellular response to DNA double-strand breaks in G1” (2008) *Mol. Cell* 30:73–85).

[0006] Repair outcomes resulting from end-joining pathways have been analyzed in detail for certain specialized nucleases, including RAG-induced V(D)J recombination (Purugganan *et al.*, “Ku80 is required for addition of N nucleotides to V(D)J recombination junctions by terminal deoxynucleotidyl transferase” *Nucleic Acids Res* (2001) 29:1638–1646; Corneo *et al.*, “Rag mutations reveal robust alternative end joining” *Nature* (2007) 449: 483–486), repair at I-SceI meganuclease sites (Rouet *et al.*, “Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease” *Mol. Cell. Biol.* (1994) 14:8096–8106; Lloyd *et al.*, “Single molecule PCR reveals similar patterns of non-homologous DSB repair in tobacco and Arabidopsis” *PloS One* (2012) 7:e32255.2012; Bindra *et al.*, “Development

of an assay to measure mutagenic non-homologous end-joining repair activity in mammalian cells” (2013) *Nucleic Acids Res* 41: e115–e115; Soong *et al.*, “Development of a novel method to create double-strand break repair fingerprints using next-generation sequencing” *DNA Repair* (2015) 26:44–53), and repair of DSBs induced by HO endonuclease (Moore and Haber, “Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*” *Mol. Cell. Biol.* (1996) 16:2164–2173; Li *et al.*, “Multiple end joining mechanisms repair a chromosomal DNA break in fission yeast” *DNA Repair (Amst.)* (2012) 11:120–130). The Moore and Haber study describes a characteristic “CA” insertion after repair of an HO endonuclease-induced DSB in *S. cerevisiae*, whereas Li *et al.* describe a characteristic insertion of an “A” nucleotide after repair of an HO endonuclease-induced DSB in *S. pombe*. Both studies identify genetic dependencies of these signature repair events. The range of sequences for which these DSB repair outcomes has been measured is, however, limited.

[0007] In contrast, Cas9, which can be easily targeted to any site in a genome that lies adjacent to a protospacer-adjacent motif (PAM) (e.g. NGG for *Streptococcus pyogenes* Cas9), enables large-scale analysis of DNA repair outcomes throughout the genome. Previous studies analyzing DNA repair outcomes following Cas9 activity (Tan *et al.*, “Off-target assessment of CRISPR-Cas9 guiding RNAs in human iPS and mouse ES cells” *Genesis* (2015) 53:225–236) identified a small number of signature deletions in clonally isolated lines that were attributed to MMEJ activity (Koike-Yusa *et al.*, “Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library” *Nat. Biotechnol.* (2014) 32:267–273).

[0008] There are several different CRISPR-Cas systems and the nomenclature and classification of these have changed as the systems have been characterized. In particular, CRISPR-Cas systems have recently been reclassified into two classes, containing five types and sixteen subtypes. Makarova *et al.* (*Nature Reviews Microbiology* (2015) 13:1-15). This classification is based upon identifying all *cas* genes in a CRISPR-Cas locus and then determining the signature genes in each CRISPR-Cas locus, thereby determining whether the CRISPR-Cas systems should be placed in either Class 1 or Class 2 based upon the genes encoding the effector module, *i.e.*, the proteins involved in the interference stage. These CRISPR-Cas systems are described in greater detail below.

[0009] The use of such CRISPR-Cas systems for predicting the outcome of genome editing would be highly advantageous.

SUMMARY

- 5 [0010] The DNA repair patterns at over 200 sites in the human genome are characterized and described in detail herein. The DNA repair profiles of these sites demonstrate that the distribution of indels resulting from DNA repair at Cas9-mediated DSBs is nonrandom and consistent across experimental replicates, cell lines and reagent delivery methods. Furthermore, these patterns are dependent on the sequence of the target, i.e., they are determined by the
10 protospacer sequence rather than the genomic context, and can be modulated by inhibiting specific components of the DNA repair machinery. Additionally, it has been found that the PAM-distal nucleotide in the protospacer sequence, nucleotide 17 at Cas9 targets (Cas9 cleaves between positions 17 and 18), is highly predictive of which nucleotide is most frequently inserted.
- 15 [0011] Thus, DNA repair profiling in cell lines can be used to anticipate the repair outcomes at a given target site in cells. The invention described herein provides methods for using and modulating end-joining DNA-repair machinery to generate precise edits, as well as methods for predictably inserting a single base pair at particular positions in protospacer regions.
- 20 [0012] In one embodiment, a method is provided for restoring genetic function. The method comprises: (1) identifying a genetic mutation that reduces the function of a gene product; (2) engineering one or more site-directed nucleases to cut within the vicinity of the identified genetic mutation; (3) introducing the one or more nucleases into a cell; (4) cleaving DNA within the cell with the one or more nucleases; (5) profiling DNA repair outcomes of nuclease-dependent breaks; (6) selecting one or more nucleases that produce the desired outcomes; and (7)
25 engineering a cell with the selected one or more nucleases, to restore genetic function.
- 30 [0013] In an additional embodiment, a method for altering genetic function is provided. The method comprises: (1) identifying a genetic region of interest where alteration changes the function of a gene product; (2) engineering one or more site-directed nucleases to cut within the vicinity of the identified genetic region; (3) introducing the one or more nucleases into a cell; (4) cleaving DNA within the cell with the one or more nucleases; (5) profiling DNA repair outcomes of nuclease-dependent breaks; (6) selecting one or more nucleases that produce the desired

outcomes; (7) engineering a cell with the selected one or more nucleases, to alter genetic function.

[0014] In both of the embodiments above, the site-directed nuclease can be a catalytically active Cas protein that is complexed to a guide polynucleotide. In certain embodiments, the Cas protein is a Cas9 protein, such as a Cas9 protein from *Streptococcus pyogenes* or an orthologous Cas9 protein.

[0015] In another embodiment, a method of modulating DNA repair outcomes is provided. The method comprises: (1) contacting a selected target nucleic acid comprising a region to be modified with (a) an agent that suppresses non-homologous end-joining (NHEJ), thereby favoring alternative end-joining (alt-EJ)/microhomology mediated end-joining (MMEJ); and (b) a DNA binding molecule that targets the selected target nucleic acid; and (2) producing one or more double-strand breaks in the targeted region using a programmable endonuclease, thereby triggering DNA repair pathways to repair the breaks, whereby repair of the cleaved target nucleic acid proceeds substantially by MMEJ and is done in the absence of a donor polynucleotide, thereby modulating the DNA repair outcome.

[0016] In additional embodiments, the invention is directed to a method of modulating DNA repair outcomes. The method comprises contacting a selected target nucleic acid comprising a region to be modified with (a) an agent that suppresses non-homologous end-joining (NHEJ), thereby favoring alternative end-joining (alt-EJ)/microhomology mediated end-joining (MMEJ); and (b) one or more complexes comprising: (i) a catalytically active Cas protein and (ii) a first guide polynucleotide that comprises a spacer adapted to bind to and cleave the selected target nucleic acid at a site adjacent the region to be modified, whereby repair of the cleaved target nucleic acid proceeds by MMEJ without a donor polynucleotide, thereby modulating the DNA repair outcome.

[0017] In another embodiment, the invention is directed to a method of modulating DNA repair outcomes comprising: (1) contacting a selected target nucleic acid comprising a region to be modified with (a) an agent that suppresses alternative end-joining (alt-EJ)/microhomology mediated end-joining (MMEJ), thereby favoring non-homologous end-joining (NHEJ); and (b) a DNA binding molecule that targets the selected target nucleic acid; and (2) producing one or more double-strand breaks in the targeted region using a programmable endonuclease, thereby triggering DNA repair pathways to repair the breaks, whereby repair of the cleaved target nucleic

acid proceeds substantially by NHEJ and is done in the absence of a donor polynucleotide, thereby modulating the DNA repair outcome.

[0018] In additional embodiments, the invention is directed to a method of modulating DNA repair outcomes. The method comprises contacting a selected target nucleic acid comprising a region to be modified with (a) an agent that suppresses alternative end-joining (alt-EJ)/microhomology mediated end-joining (MMEJ), thereby favoring non-homologous end-joining (NHEJ); and (b) one or more complexes comprising: (i) a catalytically active Cas protein and (ii) a first guide polynucleotide that comprises a spacer adapted to bind to and cleave the selected target nucleic acid at a site adjacent the region to be modified, whereby repair of the cleaved target nucleic acid proceeds by c-NHEJ without a donor polynucleotide, thereby modulating the DNA repair outcome.

[0019] In another embodiment, the invention is directed to a method of modulating DNA repair outcomes. The method comprises: (1) contacting a selected target nucleic acid comprising a region to be modified with (a) a first agent that suppresses alternative end-joining (alt-EJ)/microhomology mediated end-joining (MMEJ), and a second agent that suppresses non-homologous end-joining (NHEJ), thereby favoring homology directed repair (HDR), and (b) a DNA binding molecule that targets the selected target nucleic acid; (2) producing one or more double-strand breaks in the targeted region using a programmable endonuclease; and (3) inserting at least a portion of a donor polynucleotide into said target nucleic acid at the double-strand break by an HDR DNA repair pathway, thereby modulating the DNA repair outcome.

[0020] In certain embodiments of the methods above, the agent(s) in (1), such as an inhibitor of c-NHEJ and/or MMEJ, including without limitation, peptide inhibitors, small molecules, compounds, etc., prevent activity of certain components of the NHEJ and/or MMEJ pathways, and do not interact directly with the region to be modified. Alternatively, the agent can be delivered directly to the region to be modified.

[0021] In additional embodiments, the invention is directed to a method to discover homology directed repair (HDR) “hotspots,” *i.e.*, regions that demonstrate a propensity to incorporate nucleotide information from a donor template. The method comprises: (1) identifying a region of interest to either reduce or restore the function of a gene product or a region that is a “safe harbor” to insert novel genetic information without disrupting neighboring gene function; (2) engineering one or more site-directed nucleases to cut within the vicinity of the identified genetic

mutation; (3) introducing the one or more nucleases into a cell; (4) cleaving DNA within the cell with the one or more nucleases; (5) profiling DNA repair outcomes of nuclease-dependent breaks; (6) determining the relative contribution of c-NHEJ and MMEJ to the repair pattern; and (7) predicting HDR hotspots based on the contributions in step (6).

5 [0022] In yet a further embodiment, the invention is directed to a method to enhance HDR efficiency through modulating end-joining pathways. The method comprises: (1) identifying a region of interest to either reduce or restore the function of a gene product or a region that is a “safe harbor” to insert novel genetic information without disrupting neighboring gene function. This region has a particular ratio of c-NHEJ versus MMEJ events previously discovered through
10 DNA repair profiling; (2) engineering one or more site-directed nucleases to cut within the vicinity of the identified genetic region; (3) introducing the one or more nucleases into a cell with donor templates; (4) cleaving DNA within the cell with the one or more nucleases to initiate the incorporation of nucleotide information from a donor template (HDR); (5) promoting HDR activity with an agent that suppresses either or both MMEJ and NHEJ, thereby favoring HDR;
15 (6) profiling DNA repair outcomes of nuclease-dependent breaks to confirm HDR events; (7) selecting one or more nucleases that produce the desired outcomes; and (8) engineering a cell with the selected one or more nucleases, to alter existing genetic function or to introduce new genetic function.

[0023] In certain embodiments, the methods are performed in a cell, such as a eukaryotic cell.

20 [0024] In additional embodiments, the methods can be performed using cells with certain genetic mutations as a screen for sensitivity to certain compounds that modulate choice between DNA repair pathways.

[0025] In additional embodiments, the methods can be performed using cells to reveal certain genetic liabilities in a screen for sensitivity to certain compounds that modulate choice between
25 DNA repair pathways.

[0026] In additional embodiments, the invention is directed to a method for predictably inserting a particular single nucleotide at a target site following Cas-mediated cleavage. The method comprises: (1) selecting a gene comprising a target region to be modified; (2) designing a guide polynucleotide, such as sgRNA, to target a selected protospacer in the target region; (3)
30 producing a double-strand break in the target region using a programmable endonuclease,

wherein the protospacer in the target region is cleaved at nucleotide position 17; and (4) inserting a particular nucleotide at the cleavage site.

[0027] In additional embodiments of all of the above-described methods, the Cas protein is a Cas9 protein, such as a Cas9 protein from *Streptococcus pyogenes* or an orthologous Cas9 protein.

[0028] In any of the embodiments described above, the guide polynucleotide can be a sgRNA, or a guide designed for use with Cas proteins other than Cas9. In other embodiments, a dual-guide system can be used. Such guides are discussed in detail below.

[0029] These aspects and other embodiments of the methods described herein will readily occur to those of ordinary skill in the art in view of the disclosure herein.

INCORPORATION BY REFERENCE

[0030] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

SEQUENCE LISTING

[0031] The sequences referred to herein are listed in the **Sequence Listing** submitted as an ASCII text file entitled "1140-0002.40_ST25.txt" – 117 KB and was created on February 17, 2017. The Sequence Listing entitled "1140-0002.40_ST25.txt" is incorporated herein by reference in its entirety.

BRIEF DESCRIPTION OF THE DRAWINGS

[0032] **FIGS. 1A-1C** depict the profiling of DNA repair outcomes after Cas9 cleavage.

[0033] **FIGS. 2A-2C** show that DNA repair profiles are unique to each spacer sequence.

[0034] **FIGS. 3A-3F** show the results of an experiment using multiple target single spacers (MTSS) sequences and demonstrate that DNA repair outcomes at Cas9-mediated DSBs are sequence-dependent.

[0035] **FIGS. 4A-4I** show that the distribution of DNA repair outcomes after Cas9 cleavage changes over time.

[0036] **FIGS. 5A-5D** show that chemical perturbation of c-NHEJ promotes a subset of DNA repair outcomes after Cas9 cleavage.

[0037] FIGS. 6A-6E show high frequency in-frame mutation after Cas9 cleavage of a target in *CD34*.

[0038] FIGS. 7A-7F show that DNA repair outcome profiling in cell lines is predictive for human primary cells.

5 [0039] FIGS. 8A-8E show DNA repair outcomes are similar comparing across replicates and reagent delivery methods at a *JAK1* target.

[0040] FIGS. 9A-9F show that DNA repair profiles are unique to each target.

[0041] FIGS. 10A and 10B show visualizations of DNA repair outcomes at seven sites in the human genome with the same spacer sequence.

10 [0042] FIGS. 11A-11D show that DNA repair outcomes are more similar within spacer groups than between spacer groups.

[0043] FIGS. 12A-12D show that similarity of DNA repair outcomes within spacer groups is not fully explained by microhomology.

15 [0044] FIGS. 13A-13B show sgRNP time course data from K562 and HCT116 cells for 96 target sites.

FIGS. 14A-14L show that the distribution of DNA repair outcomes after Cas9 cleavage changes over time in a cell type-dependent manner.

FIG. 15 displays the frequencies of nucleotide insertions as a function of the position-17 nucleotide in the protospacer.

20

DETAILED DESCRIPTION OF THE INVENTION

[0045] It is to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, the singular forms “a,” “an” and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “a sgRNA/Cas9 complex” includes one or more such complexes, reference to “a mutation” includes one or more mutations, and the like. It is also to be understood that when reference is made to an embodiment using a sgRNA to target Cas9 or dCas9 to a target site, one skilled in the art can use an alternative embodiment of the invention based on the use of a dual-guide RNA (e.g. crRNA/tracrRNA) in place of the sgRNA.

25

30

[0046] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although other methods and materials similar, or equivalent, to those described herein can be used in the practice of the present invention, preferred materials and methods are described herein.

[0047] In view of the teachings of the present specification, one of ordinary skill in the art can apply conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics, and recombinant polynucleotides, as taught, for example, by the following standard texts: Antibodies: A Laboratory Manual, Second edition, E. A.

Greenfield, 2014, Cold Spring Harbor Laboratory Press, ISBN 978-1-936113-81-1; Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition, R. I. Freshney, 2010, Wiley-Blackwell, ISBN 978-0-470-52812-9; Transgenic Animal Technology, Third Edition: A Laboratory Handbook, 2014, C. A. Pinkert, Elsevier, ISBN 978-0124104907; The Laboratory Mouse, Second Edition, 2012, H. Hedrich, Academic Press, ISBN 978-0123820082; Manipulating the Mouse Embryo: A Laboratory Manual, 2013, R. Behringer, *et al.*, Cold Spring Harbor Laboratory Press, ISBN 978-1936113019; PCR 2: A Practical Approach, 1995, M. J. McPherson, *et al.*, IRL Press, ISBN 978-0199634248; Methods in Molecular Biology (Series), J.M. Walker, ISSN 1064-3745, Humana Press; RNA: A Laboratory Manual, 2010, D. C. Rio, *et al.*, Cold Spring Harbor Laboratory Press, ISBN 978-0879698911; Methods in Enzymology (Series), Academic Press; Molecular Cloning: A Laboratory Manual (Fourth Edition), 2012, M. R. Green, *et al.*, Cold Spring Harbor Laboratory Press, ISBN 978-1605500560; Bioconjugate Techniques, Third Edition, 2013, G. T. Hermanson, Academic Press, ISBN 978-0123822390; Methods in Plant Biochemistry and Molecular Biology, 1997, W. V. Dashek, CRC Press, ISBN 978-0849394805; Plant Cell Culture Protocols (Methods in Molecular Biology), 2012, V. M. Loyola-Vargas, *et al.*, Humana Press, ISBN 978-1617798177; Plant Transformation Technologies, 2011, C. N. Stewart, *et al.*, Wiley-Blackwell, ISBN 978-0813821955; Recombinant Proteins from Plants (Methods in Biotechnology), 2010, C. Cunningham, *et al.*, Humana Press, ISBN 978-1617370212; Plant Genomics: Methods and Protocols (Methods in Molecular Biology), 2009, D. J. Somers, *et al.*, Humana Press, ISBN 978-1588299970; Plant Biotechnology: Methods in Tissue Culture and Gene Transfer, 2008, R. Keshavachandran, *et al.*, Orient Blackswan, ISBN 978-8173716164.

[0048] Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated proteins (Cas) are found in prokaryotic immune systems. These systems provide resistance against exogenous genetic elements, such as viruses and plasmids, by targeting their nucleic acids for degradation, in a sequence-specific manner. There are three main stages in CRISPR-Cas immune system: (1) acquisition, (2) expression, and (3) interference. Acquisition involves cleaving the genome of invading viruses and plasmids and integrating segments (termed protospacers) of this genomic DNA into the CRISPR locus of the host organism. The segments that are integrated into the host genome are known as spacers, which mediate protection from subsequent attack by the same (or sufficiently related) virus or plasmid. Expression involves transcription of the CRISPR locus and subsequent enzymatic processing to produce short mature CRISPR RNAs, each containing a single spacer sequence. Interference is induced after the CRISPR RNAs associate with Cas proteins to form effector complexes, which are then targeted to complementary protospacers in foreign genetic elements to induce nucleic acid degradation.

[0049] A CRISPR locus includes a number of short repeating sequences referred to as “repeats.”

Repeats can form hairpin structures and/or repeats can be unstructured single-stranded sequences. The repeats occur in clusters. Repeats frequently diverge between species. Repeats are regularly interspaced with unique intervening sequences, referred to as “spacers,” resulting in a repeat-spacer-repeat locus architecture. Spacers are identical to or are homologous with known foreign invader sequences. A spacer-repeat unit encodes a crisprRNA (crRNA). A crRNA refers to the mature form of the spacer-repeat unit. A crRNA contains a spacer sequence that is involved in targeting a target nucleic acid (*e.g.*, possibly as a surveillance mechanism against foreign nucleic acid). A spacer sequence is typically located towards the 5’ end of a crRNA (*e.g.* in a Type I (*e.g.* CASCADE) system; for a description of the CASCADE complex *see, e.g.*, Jore, M. M. *et al.*, “Structural basis for CRISPR RNA-guided DNA recognition by CASCADE,” *Nature Structural & Molecular Biology* (2011) 18:529–536) or at the 3’ end of the spacer of a crRNA in a Type II system (*e.g.*, in a Type II CRISPR system, described more fully below), directly adjacent to the first stem.

[0050] Thus, crRNA has a region of complementarity to a potential DNA target sequence and a second region that forms base-pair hydrogen bonds with the tracrRNA to form a secondary structure, typically to form at least a stem structure. The tracrRNA and a crRNA interact through a number of base-pair hydrogen bonds to form secondary RNA structures. Complex formation

between tracrRNA/crRNA and a Cas9 protein (described more fully below) results in conformational change of the Cas protein that facilitates binding to DNA, endonuclease activities of the Cas9 protein, and crRNA-guided site-specific DNA cleavage by the endonuclease. For a Cas9 protein/tracrRNA/crRNA complex to cleave a DNA target sequence, the DNA target sequence is adjacent to a cognate protospacer adjacent motif (PAM).

[0051] A CRISPR locus comprises polynucleotide sequences encoding for CRISPR associated genes (*cas*) genes. *Cas* genes are involved in the biogenesis and/or the interference stages of crRNA function. *Cas* genes display extreme sequence (*e.g.*, primary sequence) divergence between species and homologues. For example, *casI* homologues can comprise less than 10% primary sequence identity between homologues. Some *cas* genes comprise homologous secondary and/or tertiary structures. For example, despite extreme sequence divergence, many members of the Cas6 family of CRISPR proteins comprise an N-terminal ferredoxin-like fold. *Cas* genes are named according to the organism from which they are derived. For example, *cas* genes in *Staphylococcus epidermidis* can be referred to as Csm-type, *cas* genes in *Streptococcus thermophilus* can be referred to as Csn-type, and *cas* genes in *Pyrococcus furiosus* can be referred to as Cmr-type.

[0052] The integration stage of a CRISPR system refers to the ability of the CRISPR locus to integrate new spacers into the crRNA array upon being infected by a foreign invader. Acquisition of the foreign invader spacers can help confer immunity to subsequent attacks by the same foreign invader. Integration typically occurs at the leader end of the CRISPR locus. *Cas* proteins (*e.g.*, Cas1 and Cas2) are involved in integration of new spacer sequences. Integration proceeds similarly for some types of CRISPR systems (*e.g.*, Type I-III).

[0053] Mature crRNAs are processed from a longer polycistronic CRISPR locus transcript (*i.e.*, pre-crRNA array). A pre-crRNA array comprises a plurality of crRNAs. The repeats in the pre-crRNA array are recognized by *cas* genes. *Cas* genes bind to the repeats and cleave the repeats. This action can liberate the plurality of crRNAs. crRNAs can be subjected to further events to produce the mature crRNA form such as trimming (*e.g.*, with an exonuclease). A crRNA may comprise all, some, or none of the CRISPR repeat sequence.

[0054] Interference refers to the stage in the CRISPR system that is functionally responsible for combating infection by a foreign invader. CRISPR interference follows a similar mechanism to RNA interference (RNAi: *e.g.*, wherein a target RNA is targeted (*e.g.*, hybridized) by a short

interfering RNA (siRNA)), which results in target RNA degradation and/or destabilization. CRISPR systems perform interference of a target nucleic acid by coupling crRNAs and *Cas* genes, thereby forming CRISPR ribonucleoproteins (crRNPs). crRNA of the crRNP guides the crRNP to foreign invader nucleic acid, (*e.g.*, by recognizing the foreign invader nucleic acid through hybridization). Hybridized target foreign invader nucleic acid-crRNA units are subjected to cleavage by Cas proteins. Target nucleic acid interference typically requires a protospacer adjacent motif (PAM) in a target nucleic acid.

[0055] By a “CRISPR-Cas system” as used herein, is meant any of the various CRISPR-Cas classes, types and subtypes. CRISPR systems are currently classified into two classes, Class 1 or Class 2, based upon the genes encoding the effector module, *i.e.*, the proteins involved in the interference stage, and include five types (Types I-V) and sixteen subtypes (Makarova *et al.*, Nature Reviews Microbiology (2015) 13:1-15).

[0056] Class 1 systems have a multi-subunit crRNA-effector complex, whereas Class 2 systems have a single protein, such as Cas9, Cpf1, C2c1, C2c2, C2c3, or a crRNA-effector complex.

Class 1 systems comprise Type I, Type III and Type IV systems. Class 2 systems comprise Type II and Type V systems.

[0057] Type I systems have a Cas3 protein that has helicase activity and cleavage activity. Type I systems are further divided into seven subtypes (I-A to I-F and I-U). Each type I subtype has a defined combination of signature genes and distinct features of operon organization. For example, subtypes I-A and I-B have the *cas* genes organized in two or more operons, whereas subtypes I-C through I-F appear to have the *cas* genes encoded by a single operon. Type I systems have a multiprotein crRNA-effector complex that is involved in the processing and interference stages of the CRISPR-Cas immune system. In *Escherichia coli*, this multiprotein complex is known as CRISPR-associated complex for antiviral defense (CASCADE). Subtype I-A comprises *csa5* which encodes a small subunit protein and a *cas8* gene that is split into two, encoding degraded large and small subunits and also has a split *cas3* gene. An example of an organism with a subtype I-A CRISPR-Cas system is *Archaeoglobus fulgidus*.

[0058] Subtype I-B has a *cas1-cas2-cas3-cas4-cas5-cas6-cas7-cas8* gene arrangement and lacks a *csa5* gene. An example of an organism with subtype I-B is *Clostridium kluyveri*.

Subtype I-C does not have a *cas6* gene. An example of an organism with subtype I-C is *Bacillus halodurans*. Subtype I-D has a Cas10d instead of a Cas8. An example of an organism with

subtype I-D is *Cyanothece* sp. Subtype I-E does not have a *cas4*. An example of an organism with subtype I-E is *Escherichia coli*. Subtype I-F does not have a *cas4* but has a *cas2* fused to a *cas3* gene. An example of an organism with subtype I-F is *Yersinia pseudotuberculosis*. An example of an organism with subtype I-U is *Geobacter sulfurreducens*.

5 [0059] All type III systems possess a *cas10* gene, which encodes a multidomain protein containing a Palm domain (a variant of the RNA recognition motif (RRM)) that is homologous to the core domain of numerous nucleic acid polymerases and cyclases and that is the largest subunit of type III crRNA-effector complexes. All type III loci also encode the small subunit protein, one Cas5 protein and typically several Cas7 proteins. Type III can be further divided
10 into four subtypes, III-A through III-D. Sub-type III-A has a *csm2* gene encoding a small subunit and also has *cas1*, *cas2* and *cas6* genes. An example of an organism with subtype III-A is *Staphylococcus epidermidis*. Subtype III-B has a *cmr5* gene encoding a small subunit and also typically lacks *cas1*, *cas2* and *cas6* genes. An example of an organism with subtype III-B is *Pyrococcus furiosus*. Subtype III-C has a Cas10 protein with an inactive cyclase-like domain
15 and lacks a *cas1* and *cas2* gene. An example of an organism with subtype III-C is *Methanothermobacter thermautotrophicus*. Subtype III-D has a Cas10 protein that lacks the HD domain and a *cas1* and *cas2* gene, and has a *cas5*-like gene known as *csx10*. An example of an organism with subtype III-D is *Roseiflexus* sp.

[0060] Type IV systems encode a minimal multisubunit crRNA-effector complex comprising a
20 partially degraded large subunit, Csf1, Cas5, Cas7, and in some cases, a putative small subunit. Type IV systems lack *cas1* and *cas2* genes. Type IV systems do not have subtypes, but there are two distinct variants. One Type IV variant has a DinG family helicase, whereas a second type IV variant lacks a DinG family helicase, but has a gene encoding a small α -helical protein. An example of an organism with a Type IV system is *Acidithiobacillus ferrooxidans*.

25 [0061] Type II systems include *cas1*, *cas2* and *cas9* genes. There are two strands of RNA in Type II systems, a CRISPR RNA (crRNA) and a transactivating CRISPR RNA (tracrRNA). The tracrRNA hybridizes to a complementary region of pre-crRNA causing maturation of the pre-crRNA to crRNA. The duplex formed by the tracrRNA and crRNA is recognized by, and associates with a multidomain protein, Cas9, encoded by the *cas9* gene, which combines the
30 functions of the crRNA-effector complex with target DNA cleavage. Cas9 is directed to a target

nucleic acid by a sequence of the crRNA that is complementary to, and hybridizes with, a sequence in the target nucleic acid.

[0062] It has been demonstrated that these minimal components of the RNA-based immune system can be reprogrammed to target DNA in a site-specific manner by using a single protein and two RNA guide sequences or a single RNA molecule. Type II systems are further divided into three subtypes, subtypes II-A, II-B and II-C. Subtype II-A contains an additional gene, *csn2*. An example of an organism with a subtype II-A system is *Streptococcus thermophilus*. Subtype II-B lacks *csn2*, but has *cas4*. An example of an organism with a subtype II-B system is *Legionella pneumophila*. Subtype II-C is the most common Type II system found in bacteria and has only three proteins, Cas1, Cas2 and Cas9. An example of an organism with a subtype II-C system is *Neisseria lactamica*.

[0063] As explained above, crRNA biogenesis in a Type II CRISPR system comprises a tracrRNA. The tracrRNA is typically modified by endogenous RNaseIII. The tracrRNA hybridizes to a crRNA repeat in the pre-crRNA array. Endogenous RNaseIII is recruited to cleave the pre-crRNA. Cleaved crRNAs are subjected to exoribonuclease trimming to produce the mature crRNA form (e.g., 5' trimming). The tracrRNA typically remains hybridized to the crRNA. The tracrRNA and the crRNA associate with a site-directed polypeptide (e.g., Cas9). The crRNA of the crRNA-tracrRNA-Cas9 complex can guide the complex to a target nucleic acid to which the crRNA can hybridize. Hybridization of the crRNA to the target nucleic acid activates a wild-type, cognate Cas9 for target nucleic acid cleavage. Target nucleic acid in a Type II CRISPR system comprises a PAM. In some embodiments, a PAM is essential to facilitate binding of a site-directed polypeptide (e.g., Cas9) to a target nucleic acid.

[0064] Cas9 is an exemplary Type II CRISPR Cas protein and serves as an endonuclease. The mature crRNA that is base-paired to trans-activating crRNA (tracrRNA) forms a two-part RNA structure, also called a "dual-guide," which directs the Cas9 to introduce double-strand breaks (DSBs) in target DNA. Cas9 can be programmed by the tracrRNA/crRNA to cleave, site-specifically, target DNA using two distinct endonuclease domains (HNH and RuvC/RNase H-like domains) (see U.S. Published Patent Application No. 2014/0068797 to Doudna *et al.*, published 6 March 2014 and incorporated herein by reference in its entirety; see also Jinek *et al.*, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," Science (2012) 337:816–821), one for each strand of the DNA's double helix. RuvC and HNH

together produce double-strand breaks, and separately can produce single-stranded breaks. At sites complementary to the crRNA-guide (spacer) sequence, the Cas9 HNH nuclease domain cleaves the complementary strand and the Cas9 RuvC-like domain cleaves the non-complementary strand. Dual-crRNA/tracrRNA molecules have been engineered into single-chain crRNA/tracrRNA molecules. These single-chain crRNA/tracrRNA direct target sequence-specific Cas9 double-strand DNA cleavage.

[0065] Typically, each CRISPR-Cas9 system comprises a tracrRNA and a crRNA. However, this requirement can be bypassed by using an engineered sgRNA, described more fully below, containing a designed hairpin that mimics the tracrRNA-crRNA complex (Jinek *et al.*, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” Science (2012) 337:816–821). The part of the sgRNA sequence that is complementary to the target sequence is known as a protospacer.

[0066] Base-pairing between the sgRNA and target DNA causes double-strand breaks (DSBs) due to the endonuclease activity of Cas9. Binding specificity is determined by both sgRNA-DNA base pairing and a short DNA motif (protospacer adjacent motif (PAM) sequence: NGG) juxtaposed to the DNA complementary region (Jinek *et al.*, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” Science (2012) 337:816–821). Thus, a Type II CRISPR system only requires a minimal set of two molecules -- the Cas9 protein and the sgRNA.

[0067] A large number of Cas9 orthologs are known in the art as well as their associated tracrRNA and crRNA components (*see, e.g.*, “Supplementary Table S2. List of bacterial strains with identified Cas9 orthologs,” Fonfara, Ines, *et al.*, “Phylogeny of Cas9 Determines Functional Exchangeability of Dual-RNA and Cas9 among Orthologous Type II CRISPR/Cas Systems,” Nucleic Acids Research (2014) 42:2577–2590, including all Supplemental Data; Chylinski K., *et al.*, “Classification and evolution of type II CRISPR-Cas systems,” Nucleic Acids Research (2014) 42:6091-6105, including all Supplemental Data.); Esvelt, K. M., *et al.*, “Orthogonal Cas9 proteins for RNA-guided gene regulation and editing,” Nature Methods (2013) 10:1116–1121). A number of orthogonal Cas9 proteins have been identified including Cas9 proteins from *Neisseria meningitidis*, *Streptococcus thermophilus* and *Staphylococcus aureus*.

[0068] As used herein, “a Cas protein” such as “a Cas9 protein,” “a Cas3 protein,” “a Cpf1 protein,” etc. refers to a Cas protein derived from any species, subspecies or strain of bacteria

that encodes the Cas protein of interest, as well as variants and orthologs of the particular Cas protein in question. The Cas proteins can either be directly isolated and purified from bacteria, or synthetically or recombinantly produced, or can be delivered using a construct encoding the protein, including without limitation, naked DNA, plasmid DNA, a viral vector and mRNA for Cas expression. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Cpf1, 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, C2C1, C2C2, C2C3, homologues thereof, or modified versions thereof. These enzymes are known; for example, the amino acid sequence of *Streptococcus pyogenes* Cas9 protein may be found in the SwissProt database (available at the website uniprot.org) under accession number Q99ZW2. In some embodiments, the CRISPR protein is codon-optimized for expression in a cell of interest. In some embodiments, the CRISPR protein directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR protein lacks DNA strand cleavage activity, or acts as a nickcase. The choice of Cas protein will depend upon the particular conditions of the methods used as described herein.

[0069] Variants and modifications of Cas9 proteins are known in the art. U.S. Patent Publication 2014/0273226 to Wu, published Sep 18, 2014, incorporated herein by reference in its entirety, discusses the *Streptococcus pyogenes* Cas9 gene, Cas9 protein, and variants of the Cas9 protein including host-specific codon-optimized Cas9 coding sequences (e.g., ¶¶ 0129-0137 therein) and Cas9 fusion proteins (e.g., ¶¶ 233-240 therein). U.S. Patent Publication 2014/0315985 to May *et al.*, published Oct. 23, 2014, incorporated herein by reference in its entirety, teaches a large number of exemplary wild-type Cas9 polypeptides (e.g., SEQ ID NO: 1-256, SEQ ID NOS: 795-1346, therein) including the sequence of Cas9 from *Streptococcus pyogenes* (SEQ ID NO: 8, therein). Modifications and variants of Cas9 proteins are also discussed (e.g., ¶¶ 504-608, therein). Non-limiting examples of Cas9 proteins include Cas9 proteins from *Streptococcus pyogenes* (GI:15675041); *Listeria innocua* Clip 11262 (GI:16801805); *Streptococcus mutans* UA159 (GI:24379809); *Streptococcus thermophilus* LMD-9 (*S. thermophilus* A, GI:11662823; *S. thermophilus* B, GI:116627542); *Lactobacillus buchneri* NRRL B-30929 (GI:331702228); *Treponema denticola* ATCC 35405 (GI:42525843); *Francisella novicida* U112 (GI:118497352); *Campylobacter jejuni* subsp. *Jejuni* NCTC 11168 (GI:218563121); *Pasteurella multocida* subsp.

multocida str. Pm70 (GI:218767588); *Neisseria meningitidis* Zs491 (GI:15602992) and *Actinomyces naeslundii* (GI:489880078).

[0070] The term “Cas9 protein” as used herein refers to Type II CRISPR-Cas9 proteins (as described, *e.g.*, in Chylinski, K., (2013) “The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems,” RNA Biol. 2013 10(5):726–737), including, but not limited to Cas9, Cas9-like, proteins encoded by Cas9 orthologs, Cas9-like synthetic proteins, and variants and modifications thereof. The term, as used herein, refers to Cas9 wild-type proteins derived from Type II CRISPR-Cas9 systems, modifications of Cas9 proteins, variants of Cas9 proteins, Cas9 orthologs, and combinations thereof. Cas9 proteins can be derived from any of various bacterial species which genomes encode such proteins. Cas proteins for use in the present methods are described further below.

[0071] Cpf1, another CRISPR-Cas protein found in Type V systems, prefers a “TTN” PAM motif that is located 5’ to its protospacer target, not 3’, like Cas9, which recognizes a “NGG” PAM motif. Thus, Cpf1 recognizes a PAM that is not G-rich and is on the opposite side of the protospacer. Cpf1 binds a crRNA that carries the protospacer sequence for base-pairing the target. Unlike Cas9, Cpf1 does not require a separate tracrRNA and is devoid of a tracrRNA gene at the Cpf1-CRISPR locus, which means that Cpf1 only requires a crRNA that is about 43 bases long. 24 nt represents the protospacer and 19 nt the constitutive direct repeat sequence. Cpf1 appears to be directly responsible for cleaving the 43 base crRNAs apart from the primary transcript (Fonfara *et al.*, (2016) “The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA,” Nature 532:517-521).

[0072] Aspects of the present invention can be practiced by one of ordinary skill in the art following the guidance of the specification to use CRISPR-Cas proteins, such as CRISPR-Cas9, Cas3, Cpf1 proteins and Cas-protein encoding polynucleotides, including, but not limited to proteins encoded by the native sequences and proteins encoded by Cas orthologs, Cas-like synthetic proteins, and variants and modifications thereof. The cognate RNA components of these Cas proteins can be manipulated and modified for use in the practice of the present invention by one of ordinary skill in the art following the guidance of the present specification.

[0073] The term “sgRNA” typically refers to a single-guide RNA (*i.e.*, a single, contiguous polynucleotide sequence) that essentially comprises a crRNA connected at its 3’ end to the 5’ end of a tracrRNA through a “loop” sequence (*see, e.g.*, U.S. Published Patent Application No.

20140068797 to Doudna *et al.*, published 6 March 2014, incorporated herein by reference in its entirety). sgRNA interacts with a cognate Cas protein essentially as described for tracrRNA/crRNA polynucleotides, as discussed above. Similar to crRNA, sgRNA has a spacer, a region of complementarity to a potential DNA target sequence, adjacent a second region that forms base-pair hydrogen bonds that form a secondary structure, typically a stem structure. The term includes truncated single-guide RNAs (tru-sgRNAs) of approximately 17-18 nt (*see, e.g.*, Fu, Y. *et al.*, "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs," Nat Biotechnol. (2014) 32:279-284). The term also encompasses functional miniature sgRNAs with expendable features removed, but that retain an essential and conserved module termed the "nexus" located in the portion of sgRNA that corresponds to tracrRNA (not crRNA). *See, e.g.*, U.S. Published Patent Application No. 2014/0315985 to May *et al.*, published 23 October 2014, incorporated herein by reference in its entirety; Briner *et al.*, "Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality," Molecular Cell (2014) 56:333-339. The nexus is located immediately downstream of (*i.e.*, located in the 3' direction from) the lower stem in Type II CRISPR-Cas9 systems. The nexus confers the binding of a sgRNA or a tracrRNA to its cognate Cas9 protein and confers an apoenzyme to haloenzyme conformational transition. [0074] With reference to a crRNA or sgRNA, a "spacer" or "spacer element," as used herein refers to the polynucleotide sequence that can specifically hybridize to a target nucleic acid sequence. The spacer element interacts with the target nucleic acid sequence through hydrogen bonding between complementary base pairs (*i.e.*, paired bases). A spacer element binds to a selected DNA target sequence. Accordingly, the spacer element is a DNA target-binding sequence. The spacer element determines the location of Cas protein's site-specific binding and endonucleolytic cleavage. Spacer elements range from ~17- to ~84 nucleotides in length, depending on the Cas protein with which they are associated, and have an average length of 36 nucleotides (Marraffini, *et al.*, "CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea," Nature reviews Genetics (2010) 11:181-190). For example, for SpyCas9, the functional length for a spacer to direct specific cleavage is typically about 12-25 nucleotides. Variability of the functional length for a spacer element is known in the art (*e.g.*, U.S. Published Patent Application No. 2014/0315985 to May *et al.*, published 23 October 2014, incorporated herein by reference in its entirety).

[0075] U.S. Patent Publication No. 2014/0315985 to May *et al.*, published 23 October 2014, incorporated herein by reference in its entirety; and Briner *et al.*, “Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality,” *Molecular Cell* (2014) 56:333-339, disclose consensus sequences and secondary structures of predicted sgRNAs for several sgRNA/Cas9 families.

[0076] Ran *et al.*, “*In vivo* genome editing using *Staphylococcus aureus* Cas9,” *Nature* (2015) 520:186-191, including all extended data) present the crRNA/tracrRNA sequences and secondary structures of eight Type II CRISPR-Cas9 systems (*see* Extended Data Figure 1 of Ran, *et al.*). Furthermore, Fonfara, *et al.*, (“Phylogeny of Cas9 Determines Functional Exchangeability of Dual-RNA and Cas9 among Orthologous Type II CRISPR/Cas Systems,” *Nucleic Acids Research* (2014) 42:2577–2590, including all Supplemental Data, in particular Supplemental Figure S11) present the crRNA/tracrRNA sequences and secondary structures of eight Type II CRISPR-Cas9 systems.

[0077] By “guide polynucleotide” such as a guide RNA, is meant any polynucleotide that site-specifically guides a Cas endonuclease (e.g., Cas9, Cas3, etc.), or a deactivated Cas endonuclease (e.g., dCas9) to a target nucleic acid. Many such guide polynucleotides are known, including, but not limited to, sgRNA (including miniature and truncated sgRNAs), crRNA, dual-guide RNA, including but not limited to, crRNA/tracrRNA molecules, as described above, and the like.

[0078] The terms “wild-type,” “naturally-occurring” and “unmodified” are used herein to mean the typical (or most common) form, appearance, phenotype, or strain existing in nature; for example, the typical form of cells, organisms, characteristics, polynucleotides, proteins, macromolecular complexes, genes, RNAs, DNAs, or genomes as they occur in and can be isolated from a source in nature. The wild-type form, appearance, phenotype, or strain serve as the original parent before an intentional modification. Thus, mutant, variant, engineered, recombinant, and modified forms are not wild-type forms.

[0079] As used herein, the terms “engineered,” “genetically engineered,” “recombinant,” “modified,” and “non-naturally occurring” are interchangeable and indicate intentional human manipulation.

[0080] As used herein, the terms “nucleic acid,” “nucleotide sequence,” “oligonucleotide,” and “polynucleotide” are interchangeable. All refer to a polymeric form of nucleotides. The

nucleotides may be deoxyribonucleotides (DNA) or ribonucleotides (RNA), or analogs thereof, and they may be of any length. Polynucleotides may perform any function and may have any secondary structure and three-dimensional structure. The terms encompass known analogs of natural nucleotides and nucleotides that are modified in the base, sugar and/or phosphate
5 moieties. Analogs of a particular nucleotide have the same base-pairing specificity (*e.g.*, an analog of A base pairs with T). A polynucleotide may comprise one modified nucleotide or multiple modified nucleotides. Examples of modified nucleotides include methylated nucleotides and nucleotide analogs. Nucleotide structure may be modified before or after a polymer is assembled. Following polymerization, polynucleotides may be additionally modified via, for
10 example, conjugation with a labeling component or target-binding component. A nucleotide sequence may incorporate non-nucleotide components. The terms also encompass nucleic acids comprising modified backbone residues or linkages that (i) are synthetic, naturally occurring, and non-naturally occurring, and (ii) have similar binding properties as a reference polynucleotide (*e.g.*, DNA or RNA). Examples of such analogs include, but are not limited to,
15 phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs), and morpholino structures.

[0081] Polynucleotide sequences are displayed herein in the conventional 5' to 3' orientation unless explicitly otherwise noted.

[0082] As used herein, the term "complementarity" refers to the ability of a nucleic acid
20 sequence to form hydrogen bond(s) with another nucleic acid sequence (*e.g.*, through traditional Watson-Crick base pairing). A percent complementarity indicates the percentage of residues in a nucleic acid molecule that can form hydrogen bonds with a second nucleic acid sequence. When two polynucleotide sequences have 100% complementarity, the two sequences are perfectly complementary, *i.e.*, all of a first polynucleotide's contiguous residues hydrogen bond with the
25 same number of contiguous residues in a second polynucleotide.

[0083] As used herein, the term "sequence identity" generally refers to the percent identity of bases or amino acids determined by comparing a first polynucleotide or polypeptide to a second polynucleotide or polypeptide using algorithms having various weighting parameters. Sequence identity between two polypeptides or two polynucleotides can be determined using sequence
30 alignment by various methods and computer programs (*e.g.*, BLAST, CS-BLAST, FASTA, HMMER, L-ALIGN, etc.), available through the worldwide web at sites including GENBANK

(ncbi.nlm.nih.gov/genbank/) and EMBL-EBI (ebi.ac.uk.). Sequence identity between two polynucleotides or two polypeptide sequences is generally calculated using the standard default parameters of the various methods or computer programs. Generally, the Cas proteins for use herein will have at least about 75% or more sequence identity to the wild-type or naturally occurring sequence of the Cas protein of interest, such as about 80%, such as about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or complete identity.

[0084] As used herein a “stem-loop structure” or “stem-loop element” refers to a polynucleotide having a secondary structure that includes a region of nucleotides that are known or predicted to form a double-stranded region (the “stem element”) that is linked on one side by a region of predominantly single-stranded nucleotides (the “loop element”). The term “hairpin” element is also used herein to refer to stem-loop structures. Such structures are well known in the art. The base pairing may be exact. However, as is known in the art, a stem element does not require exact base pairing. Thus, the stem element may include one or more base mismatches or non-paired bases.

[0085] As used herein, the term “recombination” refers to a process of exchange of genetic information between two polynucleotides.

[0086] As used herein, the term “homology-directed repair” or “HDR” refers to DNA repair that takes place in cells, for example, during repair of double-strand and single-strand breaks in DNA. HDR requires nucleotide sequence homology and uses a “donor template” (donor template DNA, polynucleotide donor, or oligonucleotide (used interchangeably herein) to repair the sequence where the double-strand break occurred (*e.g.*, DNA target sequence). This results in the transfer of genetic information from, for example, the donor template DNA to the DNA target sequence. HDR may result in alteration of the DNA target sequence (*e.g.*, insertion, deletion, mutation) if the donor template DNA sequence or oligonucleotide sequence differs from the DNA target sequence and part or all of the donor template DNA polynucleotide or oligonucleotide is incorporated into the DNA target sequence. In some embodiments, an entire donor template DNA polynucleotide, a portion of the donor template DNA polynucleotide, or a copy of the donor polynucleotide is copied or integrated at the site of the DNA target sequence.

[0087] As used herein the terms “classical non-homologous end joining” or “c-NHEJ” refer to the repair of double-strand breaks in DNA by direct ligation of one end of the break to the other end of the break without a requirement for a donor template DNA. NHEJ in the absence of a

donor template DNA often results in small insertions or deletions of nucleotides at the site of the double-strand break, also referred to as “indels.” This DNA repair pathway is genetically defined and requires the activity of Ligase IV, DNA-PKcs, Pol μ , Pol λ , and the Ku70/80 heterodimer, among other proteins (*see, e.g.*, Sfeir and Symington, “Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?” Trends Biochem Sci (2015) 40:701–714).

[0088] “Microhomology-mediated end joining (MMEJ),” a form of alternative nonhomologous end-joining (alt-NHEJ), is another pathway for repairing double-strand breaks in DNA. MMEJ is associated with deletions flanking a DSB and involves alignment of microhomologous sequences internal to the broken ends before joining. The proposed mechanism entails 5’- 3’ resection of the DNA ends at a DSB, annealing of the microhomologies (1-16 nucleotides of homology), removal of heterologous flaps, gap filling DNA synthesis, and ligation. MMEJ is genetically defined and requires the activity of CtIP, PARP1, Pol θ , Lig1 and Lig3, among other proteins (*see, e.g.*, Sfeir and Symington, “Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?” Trends Biochem Sci (2015) 40:701–714).

[0089] Alternative mechanisms of DNA insertion that do not require sequence homology between the donor and the target sequence can also be used for nucleic acid insertion. These mechanisms involve various components of the cellular DNA repair machinery and it is to be understood that the scope of the invention is not bound by the use of any particular mechanism for insertion of nucleic acid after target nucleic acid is cut or nicked by a site-specific polynucleotide.

[0090] By “donor polynucleotide” is meant a polynucleotide that can be directed to, and inserted into a target site of interest to modify the target nucleic acid. All or a portion of the donor polynucleotide can be inserted into the target nucleic acid. The donor polynucleotide is used for repair of the break in the target DNA sequence resulting in the transfer of genetic information (*i.e.*, polynucleotide sequences) from the donor at the site or in close proximity to the break in the DNA. Accordingly, new genetic information (*i.e.*, polynucleotide sequences) may be inserted or copied at a target DNA site. The donor polynucleotide can be double- or single-strand DNA, RNA, a vector, plasmid, or the like. Non-symmetrical polynucleotide donors can also be used that are composed of two DNA oligonucleotides. They are partially complementary, and each can include a flanking region of homology. The donor polynucleotide

can be used to insert or replace polynucleotide sequences in a target sequence, for example, to introduce a polynucleotide that encodes a protein or functional RNA (*e.g.*, siRNA), to introduce a protein tag, to modify a regulatory sequence of a gene, or to introduce a regulatory sequence to a gene (*e.g.* a promoter, an enhancer, an internal ribosome entry sequence, a start codon, a stop codon, a localization signal, or polyadenylation signal), to modify a nucleic acid sequence (*e.g.*, introduce a mutation), and the like.

[0091] Targeted DNA modifications using donor polynucleotides for large changes (*e.g.*, more than 100 bp insertions or deletions) traditionally use plasmid-based donor templates that contain homology arms flanking the site of alteration. Each arm can vary in length, but is typically longer than about 100 bp, such as 100-1500 bp, *e.g.*, 100...200...300...400...500...600...700...800...900...1000...1500 bp or any integer between these values. However, these numbers can vary, depending on the size of the donor polynucleotide and the target polynucleotide. This method can be used to generate large modifications, including insertion of reporter genes such as fluorescent proteins or antibiotic resistance markers. For transfection in cells, such as HEK cells, approximately 100-1000 ng, *e.g.*, 100...200...300...400...500...600... 700...800...900...1000 ng or any integer between these values, of a typical size donor plasmid (*e.g.*, approximately 5kb) containing a sgRNA/Cas9 vector, can be used for one well in 24-well plate (*see, e.g.*, Yang *et al.*, "One Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering" Cell (2013) 154:1370-1379).

[0092] Single-stranded and partially double-stranded oligonucleotides, such as DNA oligonucleotides, have been used in place of targeting plasmids for short modifications (*e.g.*, less than 50 bp) within a defined locus without cloning. To achieve high HDR efficiencies, single-stranded oligonucleotides containing flanking sequences on each side that are homologous to the target region can be used, and can be oriented in either the sense or antisense direction relative to the target locus. The length of each arm can vary in length, but the length of at least one arm is typically longer than about 10 bases, such as from 10-150 bases, *e.g.*, 10...20...30...40...50...60...70...80...90...100...110...120...130...140...150, or any integer within these ranges. However, these numbers can vary, depending on the size of the donor polynucleotide and the target polynucleotide. In a preferred embodiment, the length of at least one arm is 10 bases or more. In other embodiments, the length of at least one arm is 20 bases or

more. In yet other embodiments, the length of at least one arm is 30 bases or more. In some embodiments, the length of at least one arm is less than 100 bases. In further embodiments, the length of at least one arm is greater than 100 bases. In some embodiments, the length of at least one arm is zero bases. For single-stranded DNA oligonucleotide design, typically an

5 oligonucleotide with around 100-150 bp total homology is used. The mutation is introduced in the middle, giving 50-75 bp homology arms for a donor designed to be symmetrical about the target site. In other cases, no homology arms are required, and the donor polynucleotide is inserted using non-homologous DNA repair mechanisms.

[0093] As used herein, the term “modulate” refers to a change in the quantity, degree or amount of a function. For example, the methods disclosed herein may modulate, e.g., enhance, decrease or inhibit, DNA repair function, such as HDR, NHEJ, MMEJ, and the like. Moreover, the methods can be used to modulate the function of a gene product, such as a protein, to reduce or restore activity of the protein.

10 [0094] Modulation can be assayed by determining any characteristic directly or indirectly affected by the expression of the target gene. Such characteristics include, e.g., changes in targeting efficiency, RNA or protein levels, protein activity, product levels, associated gene expression, or activity level of reporter genes. Thus, “modulation” of gene expression includes both gene activation and gene repression.

[0095] The terms “vector” and “plasmid” are used interchangeably and, as used herein, refer to a polynucleotide vehicle to introduce genetic material into a cell. Vectors can be linear or circular. Vectors can integrate into a target genome of a host cell or replicate independently in a host cell. Vectors can comprise, for example, an origin of replication, a multicloning site, and/or a selectable marker. An expression vector typically comprises an expression cassette. Vectors and plasmids include, but are not limited to, integrating vectors, prokaryotic plasmids, eukaryotic plasmids, plant synthetic chromosomes, episomes, viral vectors, cosmids, and artificial chromosomes. As used herein, the term “expression cassette” is a polynucleotide construct, generated recombinantly or synthetically, comprising regulatory sequences operably linked to a selected polynucleotide to facilitate expression of the selected polynucleotide in a host cell. For example, the regulatory sequences can facilitate transcription of the selected polynucleotide in a host cell, or transcription and translation of the selected polynucleotide in a host cell. An

20
25
30

expression cassette can, for example, be integrated in the genome of a host cell or be present in an expression vector.

[0096] As used herein, the term “expression cassette” is a polynucleotide construct, generated recombinantly or synthetically, comprising regulatory sequences operably linked to a selected polynucleotide to facilitate expression of the selected polynucleotide in a host cell. For example, the regulatory sequences can facilitate transcription of the selected polynucleotide in a host cell, or transcription and translation of the selected polynucleotide in a host cell. An expression cassette can, for example, be integrated in the genome of a host cell or be present in an expression vector.

[0097] As used herein, the terms “regulatory sequences,” “regulatory elements,” and “control elements” are interchangeable and refer to polynucleotide sequences that are upstream (5’ non-coding sequences), within, or downstream (3’ non-translated sequences) of a polynucleotide target to be expressed. Regulatory sequences influence, for example, the timing of transcription, amount or level of transcription, RNA processing or stability, and/or translation of the related structural nucleotide sequence. Regulatory sequences may include activator binding sequences, enhancers, introns, polyadenylation recognition sequences, promoters, repressor binding sequences, stem-loop structures, translational initiation sequences, translation leader sequences, transcription termination sequences, translation termination sequences, primer binding sites, and the like.

[0098] As used herein, the term “operably linked” refers to polynucleotide sequences or amino acid sequences placed into a functional relationship with one another. For instance, a promoter or enhancer is operably linked to a coding sequence if it regulates, or contributes to the modulation of, the transcription of the coding sequence. Operably linked DNA sequences encoding regulatory sequences are typically contiguous to the coding sequence. However, enhancers can function when separated from a promoter by up to several kilobases or more. Accordingly, some polynucleotide elements may be operably linked but not contiguous.

[0099] As used herein, the term “expression” refers to transcription of a polynucleotide from a DNA template, resulting in, for example, an mRNA or other RNA transcript (*e.g.*, non-coding, such as structural or scaffolding RNAs). The term further refers to the process through which transcribed mRNA is translated into peptides, polypeptides, or proteins. Transcripts and encoded

polypeptides may be referred to collectively as “gene product.” Expression may include splicing the mRNA in a eukaryotic cell, if the polynucleotide is derived from genomic DNA.

[00100] As used herein, the term “amino acid” refers to natural and synthetic (unnatural) amino acids, including amino acid analogs, modified amino acids, peptidomimetics, glycine, and D or L optical isomers.

[00101] As used herein, the terms “peptide,” “polypeptide,” and “protein” are interchangeable and refer to polymers of amino acids. A polypeptide may be of any length. It may be branched or linear, it may be interrupted by non-amino acids, and it may comprise modified amino acids. The terms may be used to refer to an amino acid polymer that has been modified through, for example, acetylation, disulfide bond formation, glycosylation, lipidation, phosphorylation, cross-linking, and/or conjugation (*e.g.*, with a labeling component or ligand). Polypeptide sequences are displayed herein in the conventional N-terminal to C-terminal orientation.

[00102] Polypeptides and polynucleotides can be made using routine techniques in the field of molecular biology (*see, e.g.*, standard texts discussed above). Furthermore, essentially any polypeptide or polynucleotide can be custom ordered from commercial sources.

[00103] The term “binding,” as used herein, includes a non-covalent interaction between macromolecules (*e.g.*, between a protein and a polynucleotide, between a polynucleotide and a polynucleotide, and between a protein and a protein). Such non-covalent interaction is also referred to as “associating” or “interacting” (*e.g.*, when a first macromolecule interacts with a second macromolecule, the first macromolecule binds to second macromolecule in a non-covalent manner). Some portions of a binding interaction may be sequence-specific; however, all components of a binding interaction do not need to be sequence-specific, such as a protein's contacts with phosphate residues in a DNA backbone. Binding interactions can be characterized by a dissociation constant (K_d). “Affinity” refers to the strength of binding. An increased binding affinity is correlated with a lower K_d . An example of non-covalent binding is hydrogen bond formation between base pairs.

[00104] As used herein, the term “isolated” can refer to a nucleic acid or polypeptide that, by the hand of a human, exists apart from its native environment and is therefore not a product of nature. Isolated means substantially pure. An isolated nucleic acid or polypeptide can exist in a purified form and/or can exist in a non-native environment such as, for example, in a recombinant cell.

[00105] As used herein, a “host cell” generally refers to a biological cell. A cell can be the basic structural, functional and/or biological unit of a living organism. A cell can originate from any organism having one or more cells. Examples of host cells include, but are not limited to: a prokaryotic cell, eukaryotic cell, a bacterial cell, an archaeal cell, a cell of a single-cell

5 eukaryotic organism, a protozoa cell, a cell from a plant (e.g. cells from plant crops, fruits, vegetables, grains, soy bean, corn, maize, wheat, seeds, tomatoes, rice, cassava, sugarcane, sunflower, sorghum, millet, alfalfa, oil-producing *Brassica* (for example, but not limited to, oilseed rape/canola), pumpkin, hay, potatoes, cotton, cannabis, tobacco, flowering plants, conifers, gymnosperms, ferns, clubmosses, hornworts, liverworts, mosses), an algal cell, (e.g.,
 10 *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens* C. Agardh, and the like), seaweeds (e.g. kelp), a fungal cell (e.g., a yeast cell, a cell from a mushroom), an animal cell, a cell from an invertebrate animal (e.g. fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal (e.g., a pig, a cow, a goat, a sheep, a rodent, a rat, a mouse, a
 15 non-human primate, a human, etc.). Furthermore, a cell can be a stem cell or progenitor cell.

[00106] As used herein, the term “biological sample” refers to a sample of tissue or fluid, cultured cells isolated from a subject, or an environmental sample. Typical samples include but are not limited to, samples derived from blood, plasma, serum, fecal matter, urine, bone marrow, bile, spinal fluid, lymph fluid, samples of the skin, secretions of the skin, respiratory, intestinal,
 20 and genitourinary tracts, tears, saliva, milk, organs, tumors, biopsies, cells desired for gene editing, cell lines, and also samples of *in vitro* cell culture constituents including but not limited to conditioned media resulting from the growth of cells and tissues in culture medium, e.g., recombinant cells, and cell components. Environmental samples can be from water, dirt, rocks, and the like.

25 [00107] The present methods can be used in a number of genome-editing applications. In the absence of a donor template, DNA repair was previously thought to result in random outcomes. Through sequencing and detailed classification of indel classes, it is demonstrated herein that the distributions of DNA repair outcomes at Cas9-mediated DSBs are, in fact, nonrandom and dependent on the target site sequence. This observation has been unrecognized to date for the
 30 small indels formed by c-NHEJ-dependent repair. As shown in the examples, the frequency with which different microhomology opportunities are used by MMEJ were quantified. At the target

sites profiled, a large range of repair outcomes was observed. At some sites, one or two dominant outcomes comprised a large fraction of total edits (*e.g.*, FIGS. 6A-6E) whereas, at other sites, a collection of many reproducible, yet lower-frequency repair outcomes occurred (*e.g.*, FIGS. 1A-1C). As described herein, each target has a unique and highly reproducible DNA repair pattern that is not explained by microhomology alone (FIGS. 2A-2C) (Bae *et al.* “Microhomology-based choice of Cas9 nuclease target sites” (2014) Nat. Methods 11:705–706). [00108] There are numerous settings where using non-random DNA repair patterns as a predictive tool for genome editing experiments is advantageous. DNA repair profiling can be used to improve the ability to generate a gene knockout by facilitating the avoidance of in-frame mutations (FIGS. 6A-6E). Alternatively, an in-frame mutation can be used to dissect out a single amino acid in a functional domain of a protein. Furthermore, more tractable cell lines can be used rationally to model c-NHEJ/MMEJ indels intended for ultimate application in primary cells. As shown in the examples, this was conducted in peripheral blood-derived hematopoietic stem cells (HSCs) and DNA repair profiling revealed DNA repair outcomes with strong similarity to those observed in research cell lines (FIGS. 7A-7F).

[00109] Although editing of genomic locations to correct or introduce specific variants can be approached using DNA donor templates and HDR repair pathways, the efficiency of DNA insertion remains low, particularly in primary cells. In the methods described herein, specific, high frequency DNA editing outcomes can be induced at certain targets without the use of donor DNA, thus using pathways other than those utilized in gene-editing using HDR pathways. [00110] For example, c-NHEJ and MMEJ pathways may be employed to restore the reading frame of mutant *BRCA2* alleles with a single nucleotide insertion or specific four base deletions (FIGS. 7A-7F).

[00111] Additionally, these outcomes can be further exploited with the use of agents that suppress NHEJ, thereby favoring MMEJ pathways, so that DNA repair proceeds substantially by MMEJ pathways. By “substantially” is meant that more than 50% of the repairs are conducted using MMEJ pathways, such as more than 55%... 60%... 65%... 70%... 75%... 80%... 85%... 90%... 95%... 99% or more. For example, the repair outcome can be substantially MMEJ with other repair pathways occurring that represent a minor percentage of overall repair.

Additionally, all NHEJ repair activity can be suppressed if desired in order to produce a cleaner profile. Such agents include chemical inhibitors (FIGS. 5A-5D). NHEJ is initiated when free

DNA ends are bound by Ku70 and Ku80, which recruit the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs). The resulting complex, known as the DNA-dependent protein kinase (DNA-PK) complex, phosphorylates downstream targets leading to activation of the DNA damage response and initiation of NHEJ. Thus, suppression of the NHEJ key enzymes Ku70, Ku80, or DNA Ligase IV inhibit DNA-PK and can be used in the present methods to modulate DNA repair outcomes by inhibiting NHEJ. Such inhibitors include without limitation, NU7441 (Leahy *et al.*, "Identification of a highly potent and selective DNA-dependent protein kinase (DNA-PK) inhibitor (NU7441) by screening of chromenone libraries" *Bioorg. Med. Chem. Lett.* (2004) 14:6083–6087); KU-0060648 (Robert *et al.*, "Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing" *Genome Med* (2015) 7:93); DNA Ligase IV inhibitor, Scr7 (Maruyama *et al.*, "Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining" *Nat. Biotechnol.* (2015) 33:538–542); NU7026 (Willmore *et al.*, "A novel DNA-dependent protein kinase inhibitor, NU7026, potentiates the cytotoxicity of topoisomerase II poisons used in the treatment of leukemia" *Blood* (2004) 103); anti-EGFR-antibody C225 (Cetuximab) (Dittmann *et al.*, "Inhibition of radiation-induced EGFR nuclear import by C225 (Cetuximab) suppresses DNA-PK activity" *Radiother and Oncol* (2005) 76:157), and the like.

[00112] Similarly, agents that suppress MMEJ pathways can be used to favor NHEJ, such that DNA repair proceeds substantially using NHEJ pathways. By "substantially" is meant that more than 50% of the repairs are conducted using NHEJ pathways, such as more than 55%... 60%... 65%... 70%... 75%... 80%... 85%... 90%... 95%... 99% or more. For example, the repair outcome can be substantially NHEJ with other repair pathways occurring that represent a minor percentage of overall repair. Additionally, all MMEJ repair activity can be suppressed if desired in order to produce a cleaner profile. Such agents include chemical inhibitors and the like. In this regard, several factors are required for MMEJ including FEN1, Ligase III, MRE11, NBS1, PARP1 and XRCC1 (Sharma *et al.*, *Cell Death Dis.* (2015) 6:e1697). Thus, inhibitors of these factors or genes encoding therefore, will find use in suppressing MMEJ-directed repair. For example, Mirin and derivatives thereof have been shown to inhibit MRE11 (Shibata *et al.*, *Molec. Cell* (2014) 53:7-18) and have little effect on NHEJ and can therefore be used to suppress MMEJ and favor NHEJ-directed repair. Additionally, DNA Polymerase theta (Polθ; encoded by *PolQ*) is a critical MMEJ factor in mammalian cells. *PolQ* loss results in increased rates of

HDR. Thus, inhibitors of *PolQ* can be used to suppress MMEJ. CtIP has also been shown to play a critical role in MMEJ. Thus, inhibitors of CtIP will also find use herein. *See*, Sfeir and Symington, "Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?" *Trends Biochem Sci* (2015) 40:701–714), for a review of MMEJ and in particular, Table 1 therein for a list of genetic determinants of MMEJ, suppression of which could be used to favor NHEJ pathways.

[00113] In other contexts, agents that suppress both NHEJ and MMEJ pathways, such as one or more agents as described above, can be used in the presence of donor polynucleotides in order to substantially favor HDR. A first agent can be used to suppress NHEJ pathways and a second agent can be used to suppress MMEJ pathways. In some embodiments, the same agent can be used to suppress both pathways. If two agents are used, they can be provided concurrently, or one before the other.

[00114] Methods of determining which pathways are substantially favored are well known in the art and described in detail herein in the examples. *See*, also, Truong *et al.*, "Microhomology-mediated end joining and homologous recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells" *Proc. Natl. Acad. Sci. U.S.A.* (2013) 110:7720–7725; Bennardo *et al.*, "Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair" *PLoS Genet.* (2008) 4:e1000110; Simsek and Jasin, "Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation" *Nat. Struct. Mol. Biol.* (2010) 17:410–416; Ira *et al.*, "DNA end resection, homologous recombination and DNA damage check point activation require CDK1" *Nature* (2004) 431:1011–1017; Haber *et al.*, "In vivo biochemistry: physical monitoring of recombination induced by site-specific endonucleases" *Bioessays* (1995) 17:609–620; Rouet *et al.*, "Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease" *Mol. Cell. Biol.* (1994) 14:8096–8106; and Frank *et al.*, "Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV" *Nature* (1998) 396:173–177.

[00115] Moreover, the methods described herein can be used for repair profiling to screen patient cells for sensitivity to certain chemotherapeutics. For example, BRCA1-deficient cells can be screened for certain PARP inhibitors (inhibitors of the enzyme poly ADP ribose

polymerase) where Cas9 is simply a double-strand break initiator. Alternatively, the methods can be used to discover new genetic liabilities in a screen setting.

[00116] It is readily apparent that the ability to predictably modulate DNA repair outcomes by favoring NHEJ, MMEJ in the absence of a donor, as well as HDR using a donor polynucleotide, can be used in a large number of contexts.

[00117] In another embodiment, a method for designing a guide polynucleotide for predictably inserting a single nucleotide at a target site following Cas-mediated cleavage is provided.

Following CRISPR-Cas cleavage of a target site, one of the more common DNA repair outcomes is a single base pair insertion. Thus, the ability to predict which nucleotide is most frequently inserted at Cas cleavage sites, such as Cas9 cleavage sites, from the underlying DNA sequence, is highly valuable for the design of therapeutics and other precision applications. The ability to predict which nucleotide is most frequently inserted at Cas9 targets can be used to repair with specificity single base pair deletions, such as deletions that play a role in producing genetic defects. Additionally, the discovery that protospacers with A or T at position 17 have more frequent single base pair insertions can be used to select guides that are likely to have a higher fraction of frameshift mutations for inactivating a gene.

[00118] As explained above, a guide polynucleotide, such as sgRNA, is a short synthetic RNA sequence composed of a “scaffold” sequence necessary for Cas-binding, such as Cas9 binding, and a user-defined approximately 17-86 nucleotide “spacer” or “targeting” sequence which defines the genomic target to be modified. Thus, one can change the genomic target of Cas9 by changing the targeting sequence present in the sgRNA.

[00119] As shown in the examples, a data set of the repair outcomes of over 200 Cas9 targets in a human cell line (K562) was used as input for predictive analyses. A classifier accurately predicted which nucleotide was most frequently inserted at Cas9 targets based on the PAM-distal, cut-site nucleotide (protospacer position 17) and further analyses revealed that the position-17 nucleotide explained substantial portions of the variance in the frequencies of insertions of A or T. Thus, guide polynucleotides, such as sgRNAs, can be designed to target selected protospacers where it is desirable to insert a particular nucleotide, such as an A or a T, at base 17 of the protospacer targeted.

[00120] Methods of designing particular guide polynucleotides, such as sgRNAs, are known and described herein. See the examples herein, as well as Briner *et al.*, “Guide RNA Functional

Modules Direct Cas9 Activity and Orthogonality,” *Molecular Cell* (2014) 56:333-339. To do so, the genomic sequence for the gene to be targeted is first identified. The exact region of the selected gene to target will depend on the specific application. For example, in order to activate or repress a target gene using, for example, Cas activators or repressors, guide polynucleotides can be targeted to the promoter driving expression of the gene of interest. For genetic knock-outs, guide polynucleotides are commonly designed to target 5’ constitutively expressed exons which reduces the chances or removal of the targeted region from mRNA due to alternative splicing. Exons near the N-terminus can be targeted because frameshift mutations here will increase the likelihood of the production of a nonfunctional protein product. Alternatively, guide polynucleotides can be designed to target exons that code for known essential protein domains. In this regard, non-frameshift mutations such as insertions or deletions are more likely to alter protein function when they occur in protein domains that are essential for protein function. For gene editing using HDR, the target sequence should be close to the location of the desired edit. In this case, the location where the edit is desired is identified and a target sequence is selected nearby.

[00121] Using the methods described herein, any desired nucleic acid sequence for modification can be targeted, including without limitation, protein coding sequences in order to reduce or restore the function of the gene product; regions that have a propensity to incorporate nucleotide sequences from a donor template, termed “HDR hotspots” herein; safe harbor regions, i.e., regions where nucleotide sequences can be inserted without disrupting neighboring gene function; non-coding regulatory regions in nucleic acid sequences; and the like.

[00122] For example, protein coding sequences for targeting by the methods described herein include, but are not limited to, mammalian antibodies (ABs) (IgG, IgA, IgM, IgE), antibody fragments such as Fc regions, antibody Fab regions, antibody heavy chains, antibody light chains, antibody CDRs, nanobodies, chimeric antibodies and other IgG domains; T cell receptors (TCR); endonucleases and exonucleases, such as TALENS, CRISPR nucleases such as Cas9, Cas3, Cpf1, ZnFN, meganucleases, nuclease domains such as HNH domain, RuvC domain; recombinases such as Cre, Tre, Brec1, Flp, γ -integrase, IntI4 integrase, XerD recombinase, HP1 integrase; DNA topoisomerases; transposons such as the Tc1/mariner family, Tol2, piggyBac, Sleeping beauty; RAG proteins; retrotransposons such as LTR-retrotransposons and non-LTR retrotransposons (Alu, SINE, LINE); enzymes including but not limited to arginases,

glycosydases, proteases, kinases, and glycosylation enzymes such as glycosyltransferase; anticoagulants such as protein C, Protein S and antithrombin; coagulants such as thrombin; nucleases such as DNases, RNases, helicases, GTPases; DNA or RNA binding proteins; reporter molecules, such as Green Fluorescent Protein (GFP); cell penetrating peptides and their fusions with cargo proteins; membrane proteins such as GPCRs, pain receptors such as TRP channels and ion channels; cell surface receptors including but not limited to EGFR, FGFR, VEGFR, IGFR and ephrin receptor; cell adhesion molecules like integrins and cadherins; ion channels; rhodopsins; immunoreceptors such as CD28, CD80, PD-1, PD-L1, CTLA-4, CXCR4, CXCR5, B2M, TRACA, TRBC; proteins known to be involved with genetic defects; secreted proteins including but not limited to hormones, cytokines, growth factors; vaccine antigens such as viral proteins from human immunodeficiency virus (HIV), Dengue, cytomegalovirus (CMV), Ebola, Zika and oncolytic viruses; snake toxin proteins and peptides including but not limited to phospholipases and metalloproteases; ribosomal cyclic peptides.

[00123] Guide polynucleotides, such as sgRNAs, can be designed to target any DNA sequence containing the appropriate PAM necessary for each programmable nuclease, such as a Cas endonuclease, e.g., Cas9, Cpf1 and the like. Additional PAMs can also be created in the target DNA using a type of codon optimization, where silent mutations are introduced into amino acid codons in order to create new PAM sequences. For example, for strategies using Cas9, which recognizes an NGG PAM, a CGA serine codon can be changed to CGG, preserving the amino acid coding but adding a site where double-strand breaks can be introduced. The entire coding region or parts of the coding region can thus be optimized with suitable PAM sites on the coding and non-coding strand for DNA repair after nuclease cleavage. PAM optimized DNA sequences can be manufactured, e.g., commercially, and cloned into suitable vectors.

[00124] The sgRNAs can be delivered to a cell. If the cell constitutively expresses a Cas endonuclease, such as Cas9, Cpf1, or the like, the Cas endonuclease will then be recruited to the target site to cleave the DNA. If the cell does not express a Cas endonuclease, complexes of Cas proteins, such as Cas9 proteins, and guide RNAs, such as sgRNAs (sgRNA/Cas9 complexes) are delivered to the cells to make double-strand breaks, thereby triggering the DNA repair pathways in the cells. The cells are then screened using methods well known in the art, such as using high-throughput screening techniques including, but not limited to, fluorescence-activated cell sorting (FACS)-based screening platforms, microfluidics-based screening platforms, and the like. These

techniques are well known in the art and reviewed in *e.g.*, Wojcik *et al.*, *Int. J. Molec. Sci.* (2015) 16:24918-24945. The cells can then be expanded and re-transfected with additional guide complexes to introduce further diversity and this process can be repeated iteratively until a population with the desired properties is obtained. Single cell clones are sorted from the population, expanded and sequenced to recover the mutations that resulted in the desired function.

[00125] As discussed above, the methods described herein make use of programmable endonucleases derived from the CRISPR-Cas system. For each of the above-described embodiments, when Cas9 proteins are used, any of various Cas9-derived proteins can be used, as well as other CRISPR-Cas proteins as detailed above.

[00126] A number of catalytically active Cas9 proteins are known in the art and, as explained above, a Cas9 protein for use herein can be derived from any bacterial species, subspecies or strain that encodes the same. Although in certain embodiments herein, the methods are exemplified using *Streptococcus pyogenes* Cas9, orthologs from other bacterial species will find use herein. The specificity of these Cas9 orthologs is well known. Also useful are proteins encoded by Cas9-like synthetic proteins, and variants and modifications thereof. As explained above, the sequences for hundreds of Cas9 proteins are known and any of these proteins will find use with the present methods.

[00127] Additionally, it is to be understood that other Cas nucleases, in place of or in addition to Cas9, may be used, including any of the Cas proteins described in detail above, such as derived from any of the various CRISPR-Cas classes, types and subtypes.

[00128] Moreover, in the embodiments described herein, sgRNA is used as an exemplary guide polynucleotide, however, it will be recognized by one of skill in the art that other guide polynucleotides that site-specifically guide endonucleases, such as CRISPR-Cas proteins to a target nucleic acid can be used.

[00129] If CRISPR complexes are used, they can be produced using methods well known in the art. For example, guide RNA components of the complexes can be produced *in vitro* and Cas9 components can be recombinantly produced and then the two complexed together using methods known in the art. Additionally, cell lines such as but not limited to HEK293 cells, are commercially available that constitutively express *Streptococcus pyogenes* Cas9 as well as *Streptococcus pyogenes* Cas9-GFP fusions. In this instance, cells expressing Cas9 can be

transfected with the guide RNA components and complexes are purified from the cells using standard purification techniques, such as but not limited to affinity, ion exchange and size exclusion chromatography. *See, e.g.,* Jinek M., *et al.*, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” *Science* (2012) 337:816–821.

5 [00130] More than one set of complexes can be used, such as 2-30 or more, for example 5-20, 8-15, etc., or any number within these ranges.

[00131] The complexes, such as sgRNA/Cas9 complexes may be introduced to cells at differing concentrations. For example, sgRNA/Cas9 and sgRNA/dCas9 complexes can be introduced at a ratio of 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1, 3:1, or 2:1.

10 Additionally, all of these components, *i.e.*, sgRNA and Cas9, may be provided separately, *e.g.*, as separately assembled complexes, using separate DNA or RNA constructs, or together, in a single construct, or in any combination.

[00132] sgRNA/Cas9 complexes may be introduced at differing time points. For example, sgRNA/Cas9 complexes can be introduced at least 1 minute apart, 5 minutes apart, 10 minutes apart, 30 minutes apart, 1 hour apart, 5 hours apart, or 15 hours apart or more. sgRNA/Cas9 complexes can be introduced at most 1 minute apart, 5 minutes apart, 10 minutes apart, 30 minutes apart, 1 hour apart, 5 hours apart, or 15 hours apart or more. One set of complexes can be purified out before introducing another set of complexes. sgRNA/Cas9 complexes may be differentially regulated (*i.e.* differentially expressed or stabilized) via exogenously supplied agents (*e.g.* inducible DNA promoters or inducible Cas9 proteins).

20 [00133] Thus, in exemplary embodiments as described above, a sgRNA, complexed with Cas9 (sgRNA/Cas9 complex) is directed to a genomic locus of interest to induce double-strand breaks. The binding specificity is determined by both sgRNA-DNA base pairing and the PAM sequence juxtaposed to the DNA complementary region.

25 [00134] In all of the embodiments of the above-described methods, the various components can be produced by synthesis, or for example, using expression cassettes encoding a programmable endonuclease, such as a Cas protein, guide polynucleotide, etc. The various components can be provided to a cell or used *in vitro*. These components can be present on a single cassette or multiple cassettes, in the same or different constructs. Expression cassettes typically comprise regulatory sequences that are involved in one or more of the following: regulation of transcription, post-transcriptional regulation, and regulation of translation. Expression cassettes

30

can be introduced into a wide variety of organisms including bacterial cells, yeast cells, plant cells, and mammalian cells. Expression cassettes typically comprise functional regulatory sequences corresponding to the organism(s) into which they are being introduced.

[00135] In one aspect, all or a portion of the various components for use in the methods are produced in vectors, including expression vectors, comprising polynucleotides encoding therefor. Vectors useful for producing components for use in the present methods include plasmids, viruses (including phage), and Integra table DNA fragments (*i.e.*, fragments integratable into the host genome by homologous recombination). A vector replicates and functions independently of the host genome, or may, in some instances, integrate into the genome itself. Suitable replicating vectors will contain a replicon and control sequences derived from species compatible with the intended expression host cell. Transformed host cells are cells that have been transformed or transfected with the vectors constructed using recombinant DNA techniques

[00136] General methods for construction of expression vectors are known in the art. Expression vectors for most host cells are commercially available. There are several commercial software products designed to facilitate selection of appropriate vectors and construction thereof, such as insect cell vectors for insect cell transformation and gene expression in insect cells, bacterial plasmids for bacterial transformation and gene expression in bacterial cells, yeast plasmids for cell transformation and gene expression in yeast and other fungi, mammalian vectors for mammalian cell transformation and gene expression in mammalian cells or mammals, viral vectors (including retroviral, lentiviral, and adenoviral vectors) for cell transformation and gene expression and methods to easily enable cloning of such polynucleotides. SnapGene™ (GSL Biotech LLC, Chicago, Ill.; snapgene.com/resources/plasmid_files/your_time_is_valuable/), for example, provides an extensive list of vectors, individual vector sequences, and vector maps, as well as commercial sources for many of the vectors.

[00137] Expression cassettes typically comprise regulatory sequences that are involved in one or more of the following: regulation of transcription, post-transcriptional regulation, and regulation of translation. Expression cassettes can be introduced into a wide variety of organisms including bacterial cells, yeast cells, mammalian cells, and plant cells. Expression cassettes typically comprise functional regulatory sequences corresponding to the host cells or organism(s) into which they are being introduced. Expression vectors can also include

polynucleotides encoding protein tags (*e.g.*, poly-His tags, hemagglutinin tags, fluorescent protein tags, bioluminescent tags, nuclear localization tags). The coding sequences for such protein tags can be fused to the coding sequences or can be included in an expression cassette, for example, in a targeting vector.

5 [00138] In some embodiments, polynucleotides encoding one or more of the various components are operably linked to an inducible promoter, a repressible promoter, or a constitutive promoter.

[00139] Several expression vectors have been designed for expressing guide polynucleotides. *See, e.g.*, Shen, B. *et al.* "Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects" (2014) Mar 2. doi: 10.1038/nmeth.2857. 10.1038/nmeth.2857. Additionally, 10 vectors and expression systems are commercially available, such as from New England Biolabs (Ipswich, MA) and Clontech Laboratories (Mountain View, CA). Vectors can be designed to simultaneously express a target-specific sgRNA using a U2 or U6 promoter, a Cas9 and/or dCas9, and if desired, a marker protein, for monitoring transfection efficiency and/or for further 15 enriching/isolating transfected cells by flow cytometry.

[00140] Vectors can be designed for expression of various components of the described methods in prokaryotic or eukaryotic cells. Alternatively, transcription can be *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase. Other RNA polymerase and promoter sequences can be used.

20 [00141] Vectors can be introduced into and propagated in a prokaryote. Prokaryotic vectors are well known in the art. Typically a prokaryotic vector comprises an origin of replication suitable for the target host cell (*e.g.*, oriC derived from *E. coli*, pUC derived from pBR322, pSC101 derived from *Salmonella*), 15A origin (derived from p15A) and bacterial artificial chromosomes). Vectors can include a selectable marker (*e.g.*, genes encoding resistance for 25 ampicillin, chloramphenicol, gentamicin, and kanamycin). Zeocin™ (Life Technologies, Grand Island, NY) can be used as a selection in bacteria, fungi (including yeast), plants and mammalian cell lines. Accordingly, vectors can be designed that carry only one drug resistance gene for Zeocin for selection work in a number of organisms. Useful promoters are known for expression of proteins in prokaryotes, for example, T5, T7, Rhamnose (inducible), Arabinose (inducible), 30 and PhoA (inducible). Furthermore, T7 promoters are widely used in vectors that also encode the T7 RNA polymerase. Prokaryotic vectors can also include ribosome binding sites of varying

strength, and secretion signals (*e.g.*, *mal*, *sec*, *tat*, *ompC*, and *pelB*). In addition, vectors can comprise RNA polymerase promoters for the expression of sgRNAs. Prokaryotic RNA polymerase transcription termination sequences are also well known (*e.g.*, transcription termination sequences from *Streptococcus pyogenes*).

5 [00142] Integrating vectors for stable transformation of prokaryotes are also known in the art (*see, e.g.*, Heap, J. T., *et al.*, "Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker," *Nucleic Acids Res.* (2012) 40:e59).

[00143] Expression of proteins in prokaryotes is typically carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or
10 non-fusion proteins.

[00144] A wide variety of RNA polymerase promoters suitable for expression of the various components are available in prokaryotes (*see, e.g.*, Jiang, Y., *et al.*, "Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system," *Environ Microbiol.* (2015) 81:2506-2514); Estrem, S.T., *et al.*, (1999) "Bacterial promoter architecture: subsite structure of UP
15 elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit," *Genes Dev.* 15;13(16):2134-47).

[00145] In some embodiments, a vector is a yeast expression vector comprising one or more components of the above-described methods. Examples of vectors for expression in *Saccharomyces cerevisiae* include, but are not limited to, the following: pYepSec1, pMFa,
20 pJRY88, pYES2, and picZ. Methods for gene expression in yeast cells are known in the art (*see, e.g.*, *Methods in Enzymology*, Volume 194, "Guide to Yeast Genetics and Molecular and Cell Biology, Part A," (2004) Christine Guthrie and Gerald R. Fink (eds.), Elsevier Academic Press, San Diego, CA). Typically, expression of protein-encoding genes in yeast requires a promoter operably linked to a coding region of interest plus a transcriptional terminator. Various yeast
25 promoters can be used to construct expression cassettes for expression of genes in yeast.

Examples of promoters include, but are not limited to, promoters of genes encoding the following yeast proteins: alcohol dehydrogenase 1 (ADH1) or alcohol dehydrogenase 2 (ADH2), phosphoglycerate kinase (PGK), triose phosphate isomerase (TPI), glyceraldehyde-3-phosphate dehydrogenase (GAPDH; also known as TDH3, or triose phosphate dehydrogenase), galactose-
30 1-phosphate uridyl-transferase (GAL7), UDP-galactose epimerase (GAL10), cytochrome *ci* (CYC1), acid phosphatase (PHO5) and glycerol-3-phosphate dehydrogenase gene (GPD1).

Hybrid promoters, such as the ADH2/GAPDH, CYC1/GAL10 and the ADH2/GAPDH promoter (which is induced at low cellular-glucose concentrations, *e.g.*, about 0.1 percent to about 0.2 percent) also may be used. In *S. pombe*, suitable promoters include the thiamine-repressed nmt1 promoter and the constitutive cytomegalovirus promoter in pTL2M.

5 [00146] Yeast RNA polymerase III promoters (*e.g.*, promoters from 5S, U6 or RPR1 genes) as well as polymerase III termination sequences are known in the art (*see, e.g.*, www.yeastgenome.org; Harismendy, O., *et al.*, (2003) "Genome-wide location of yeast RNA polymerase III transcription machinery," The EMBO Journal. 22(18):4738-4747.)

[00147] In addition to a promoter, several upstream activation sequences (UASs), also called
10 enhancers, may be used to enhance polypeptide expression. Exemplary upstream activation sequences for expression in yeast include the UASs of genes encoding these proteins: CYC1, ADH2, GAL1, GAL7, GAL10, and ADH2. Exemplary transcription termination sequences for expression in yeast include the termination sequences of the α -factor, CYC1, GAPDH, and PGK genes. One or multiple termination sequences can be used.

15 [00148] Suitable promoters, terminators, and coding regions may be cloned into *E. coli*-yeast shuttle vectors and transformed into yeast cells. These vectors allow strain propagation in both yeast and *E. coli* strains. Typically, the vector contains a selectable marker and sequences enabling autonomous replication or chromosomal integration in each host. Examples of plasmids typically used in yeast are the shuttle vectors pRS423, pRS424, pRS425, and pRS426
20 (American Type Culture Collection, Manassas, VA). These plasmids contain a yeast 2 micron origin of replication, an *E. coli* replication origin (*e.g.*, pMB1), and a selectable marker.

[00149] The various components can also be expressed in insects or insect cells. Suitable expression control sequences for use in such cells are well known in the art. In some
25 embodiments, it is desirable that the expression control sequence comprises a constitutive promoter. Examples of suitable strong promoters include, but are not limited to, the following: the baculovirus promoters for the p10, polyhedrin (polh), p 6.9, capsid, UAS (contains a Gal4 binding site), Ac5, cathepsin-like genes, the *B. mori* actin gene promoter; *Drosophila melanogaster* hsp70, actin, α -1- tubulin or ubiquitin gene promoters, RSV or MMTV promoters, copia promoter, gypsy promoter, and the cytomegalovirus IE gene promoter. Examples of weak
30 promoters that can be used include, but are not limited to, the following: the baculovirus promoters for the iel, ie2, ieO, etl, 39K (aka pp31), and gp64 genes. If it is desired to increase

the amount of gene expression from a weak promoter, enhancer elements, such as the baculovirus enhancer element, hr5, may be used in conjunction with the promoter.

[00150] For the expression of some of the components of the present invention in insects, RNA polymerase III promoters are known in the art, for example, the U6 promoter. Conserved features of RNA polymerase III promoters in insects are also known (*see, e.g.*, Hernandez, G., (2007) "Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity," *Nucleic Acids Res.* 2007 Jan; 35(1):21–34).

[00151] In another aspect, the various components are incorporated into mammalian vectors for use in mammalian cells. A large number of mammalian vectors suitable for use with the systems of the present invention are commercially available (*e.g.*, from Life Technologies, Grand Island, NY; NeoBiolab, Cambridge, MA; Promega, Madison, WI; DNA2.0, Menlo Park, CA; Addgene, Cambridge, MA).

[00152] Vectors derived from mammalian viruses can also be used for expressing the various components of the present methods in mammalian cells. These include vectors derived from viruses such as adenovirus, papovirus, herpesvirus, polyomavirus, cytomegalovirus, lentivirus, retrovirus, vaccinia and Simian Virus 40 (SV40) (*see, e.g.*, Kaufman, R. J., (2000) "Overview of vector design for mammalian gene expression," *Molecular Biotechnology*, Volume 16, Issue 2, pp 151-160; Cooray S., *et al.*, (2012) "Retrovirus and lentivirus vector design and methods of cell conditioning," *Methods Enzymol.* 507:29-57). Regulatory sequences operably linked to the components can include activator binding sequences, enhancers, introns, polyadenylation recognition sequences, promoters, repressor binding sequences, stem-loop structures, translational initiation sequences, translation leader sequences, transcription termination sequences, translation termination sequences, primer binding sites, and the like. Commonly used promoters are constitutive mammalian promoters CMV, EF1a, SV40, PGK1 (mouse or human), Ubc, CAG, CaMKIIa, and beta-Act, and others known in the art (Khan, K. H. (2013) "Gene Expression in Mammalian Cells and its Applications," *Advanced Pharmaceutical Bulletin* 3(2), 257–263). Furthermore, mammalian RNA polymerase III promoters, including H1 and U6, can be used.

[00153] In some embodiments, a recombinant mammalian expression vector is capable of preferentially directing expression of the nucleic acid in a particular cell type (*e.g.*, using tissue-

specific regulatory elements to express a polynucleotide). Tissue-specific regulatory elements are known in the art and include, but are not limited to, the albumin promoter, lymphoid-specific promoters, neuron-specific promoters (*e.g.*, the neurofilament promoter), pancreas-specific promoters, mammary gland-specific promoters (*e.g.*, milk whey promoter), and in particular
5 promoters of T cell receptors and immunoglobulins. Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters and the alpha-fetoprotein promoter.

[00154] Numerous mammalian cell lines have been utilized for expression of gene products including HEK 293 (Human embryonic kidney) and CHO (Chinese hamster ovary). These cell lines can be transfected by standard methods (*e.g.*, using calcium phosphate or
10 polyethyleneimine (PEI), or electroporation). Other typical mammalian cell lines include, but are not limited to: HeLa, U2OS, 549, HT1080, CAD, P19, NIH 3T3, L929, N2a, Human embryonic kidney 293 cells, MCF-7, Y79, SO-Rb50, Hep G2, DUKX-X11, J558L, and Baby hamster kidney (BHK) cells.

[00155] Methods of introducing polynucleotides (*e.g.*, an expression vector) into host cells are
15 known in the art and are typically selected based on the kind of host cell. Such methods include, for example, viral or bacteriophage infection, transfection, conjugation, electroporation, calcium phosphate precipitation, polyethyleneimine-mediated transfection, DEAE-dextran mediated transfection, protoplast fusion, lipofection, liposome-mediated transfection, particle gun technology, direct microinjection, and nanoparticle-mediated delivery.

[00156] The present invention also includes methods of modulating *in vitro* or *in vivo*
20 transcription using the various components and complexes described herein. In one embodiment, a guide polynucleotide/Cas protein complex can repress gene expression by interfering with transcription when the guide polynucleotide directs DNA target binding of the complex to the promoter region of the gene. Use of the complexes to reduce transcription also
25 includes complexes wherein the Cas protein is fused to a known down regulator of a target gene (*e.g.*, a repressor polypeptide). For example, expression of a gene is under the control of regulatory sequences to which a repressor polypeptide can bind. A guide polynucleotide can direct DNA target binding of a repressor protein complex to the DNA sequences encoding the regulatory sequences or adjacent the regulatory sequences such that binding of the repressor
30 protein complex brings the repressor protein into operable contact with the regulatory sequences.

Similarly, Cas9 can be fused to an activator polypeptide to activate or increase expression of a gene under the control of regulatory sequences to which an activator polypeptide can bind.

[00157] The present invention also encompasses gene-therapy methods for preventing or treating diseases, disorders, and conditions using the various methods described herein. In one embodiment, a gene-therapy method uses the introduction of nucleic acid sequences into an organism or cells of an organism (*e.g.*, patient) to achieve expression of components of the present invention to provide modification of a target function. For example, cells from an organism may be engineered, *ex vivo*, by (i) introduction of vectors comprising expression cassettes expressing the various components, (ii) direct introduction of sgRNA and/or donor polynucleotides and Cas9 proteins, or (iii) introduction of combinations of these components. The engineered cells are provided to an organism (*e.g.*, patient) to be treated.

[00158] Examples of gene-therapy and delivery techniques for therapy are known in the art (see, *e.g.*, Kay, M.A., (2011) "State-of-the-art gene-based therapies: the road ahead," *Nature Reviews Genetics* 12, 316-328; Wang *et al.*, "State-of-the-art human gene therapy: part I. Gene delivery technologies," *Discov Med.* (2014) 18:67-77; Wang *et al.*, "State-of-the-art human gene therapy: part II. Gene therapy strategies and clinical applications," *Discov Med.* (2014) 18:151-61; "The Clinibook: Clinical Gene Transfer State of the Art," Odile Cohen-Haguenaer (Editor), EDP Sciences (October 31, 2012), ISBN-10: 2842541715).

[00159] In some aspects, components of the present invention are delivered using nanoscale delivery systems, such as nanoparticles. Additionally, liposomes and other particulate delivery systems can be used. For example, vectors including the components of the present methods can be packaged in liposomes prior to delivery to the subject or to cells derived therefrom, such as described in U.S. Patent Nos. 5,580,859; 5,264,618; 5,703,055, all incorporated herein by reference in their entireties. Lipid encapsulation is generally accomplished using liposomes that are able to stably bind or entrap and retain nucleic acid.

[00160] The methods described herein can also be used to generate non-human genetically modified organisms. Generally, in these methods expression cassettes comprising polynucleotide sequences of the various components, as well as a targeting vector are introduced into zygote cells to site-specifically introduce a selected polynucleotide sequence at a DNA target sequence in the genome to generate a modification of the genomic DNA. The selected polynucleotide sequence is present in the targeting vector. Modifications of the genomic DNA

typically include, insertion of a polynucleotide sequence, deletion of a polynucleotide sequence, or mutation of a polynucleotide sequence, for example, gene correction, gene replacement, gene tagging, transgene insertion, gene disruption, gene mutation, mutation of gene regulatory sequences, and so on. In one embodiment of methods to generate non-human genetically modified organisms, the organism is a mouse. Generating transgenic mice involves five basic steps (Cho A., *et al.*, "Generation of Transgenic Mice," Current protocols in cell biology, (2009); CHAPTER.Unit-19.11): (1) purifying a transgenic construct (*e.g.*, expression cassettes comprising the various components of the various methods described herein, as well as a targeting vector); (2) harvesting donor zygotes; (3) microinjecting the transgenic construct into the mouse zygote; (4) implanting the microinjected zygotes into pseudo-pregnant recipient mice; and (5) performing genotyping and analysis of the modification of the genomic DNA established in founder mice.

[00161] In another embodiment of methods to generate non-human genetically modified organisms, the organism is a plant. Thus, the components described herein are used to effect efficient, cost-effective gene editing and manipulation in plant cells. It is generally preferable to insert a functional recombinant DNA in a plant genome at a non-specific location. However, in certain instances, it may be useful to use site-specific integration to introduce a recombinant DNA construct into the genome. Recombinant vectors for use in plant are known in the art. The vectors can include, for example, scaffold attachment regions (SARs), origins of replication, and/or selectable markers.

[00162] Methods and compositions for transforming plants by introducing a recombinant DNA construct into a plant genome includes any of a number of methods known in the art. One method for constructing transformed plants is microprojectile bombardment. *Agrobacterium*-mediated transformation is another method for constructing transformed plants. Alternatively, other non-*Agrobacterium* species (*e.g.*, *Rhizobium*) and other prokaryotic cells that are able to infect plant cells and introduce heterologous nucleotide sequences into the infected plant cell's genome can be used. Other transformation methods include electroporation, liposomes, transformation using pollen or viruses, chemicals that increase free DNA uptake, or free DNA delivery by means of microprojectile bombardment. DNA constructs of the present invention may be introduced into the genome of a plant host using conventional transformation techniques

that are well known to those skilled in the art (*see, e.g.*, “Methods to Transfer Foreign Genes to Plants,” Y Narusaka, *et al.*, cdn.intechopen.com/pdfs-wm/30876.pdf).

[00163] Although preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. From the above description and the following Examples, one skilled in the art can ascertain essential characteristics of this invention, and without departing from the spirit and scope thereof, can make changes, substitutions, variations, and modifications of the invention to adapt it to various usages and conditions. Such changes, substitutions, variations, and modifications are also intended to fall within the scope of the present disclosure.

EXPERIMENTAL

[00164] Aspects of the present invention are further illustrated in the following Examples. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.*, amounts, concentrations, percent changes, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, temperature is in degrees Centigrade and pressure is at or near atmospheric. It should be understood that these Examples, although indicating some embodiments of the invention, are given by way of illustration only.

[00165] The following examples are not intended to limit the scope of what the inventors regard as various aspects of the present invention.

Materials and Methods

Cas9 and sgRNAs

[00166] Recombinant *Streptococcus pyogenes* (Spy) Cas9 and *in vitro* transcribed single-guide RNAs (sgRNAs) were generated as described by Briner *et al.*, “Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality,” *Molecular Cell* (2014) 56:333-339). Oligonucleotides used in the generation of sgRNA templates are listed in Table 1, Table 3 and Table 5.

Transcription reactions were digested with 2 Units RNase-free DNase I (NEB) at 37°C for 15 minutes; the reaction was stopped by adding EDTA to final concentration of 5 mM and incubating at 75°C for 10 minutes.

sgRNP Formation and Nucleofection

[00167] A single set of single-guide ribonucleoproteins (sgRNPs) was prepared for three independent nucleofection reactions. sgRNA was denatured by incubating at 98°C for 2 minutes and then allowed to cool at room temperature for 10 minutes. 2.9 µl of denatured sgRNA was placed in a 96-well plate (average concentration 4 µg/ml) and combined with 3.1 µl of Cas9/CCE mix (1.9 µL of Cas9 protein at 10 mg/mL + 1.2 µl 5 x CCE buffer (100 mM HEPES, pH 7.4, 500 mM KCl, 25 mM MgCl₂, 5 mM DTT, 25% Glycerol). The sgRNA and Cas9 mixture was gently vortexed and incubated at 37°C for 10 mins to form sgRNPs. 60 µl of cell suspension (10,000 cells/µl) was added to each well and mixed. 20 µl was transferred from the cell/sgRNP mixture into three 96-well nucleofection plates. Cells were nucleofected using the 96-well shuttle amaxa nucleofector II (Lonza) using the program recommended by the manufacturer for each cell line. HEK293 and K562 cell lines were nucleofected using the SF cell line kit (Lonza V4SC-2960). HCT116 cell line was nucleofected using the SE cell line kit (Lonza V4SC-1960). Hematopoietic stem cells (HSCs) were nucleofected using the P3 primary cell kit (Lonza V4SP-3960).

Cell Culture

[00168] HEK293 and HCT116 cell lines were cultured in DMEM Glutamax (Gibco, 10569-010) with 10% Fetal Bovine Serum (FBS; Gibco, 16000-044) and 1% Antibiotic–Antimycotic (Gibco, 15240-062). K562 cell line was cultured in IMDM Glutamax (Gibco, 31980-097) + 10% FBS (Gibco, 16000-044) + 1% Antibiotic–Antimycotic (Gibco, 15240-062). All cell lines were cultured at 37°C and 5% CO₂. HEK293 and HCT116 cell lines were split when they reached 90% confluence and trypsinized with TrypLE Express (Gibco, 12604-021). K562 cell line was split before reaching a density of 1×10^6 cells/ml. The three cell lines were plated 48 hours prior to nucleofection and were at 70-80% of confluence (HEK293 and HCT116) before harvesting. Cells were counted using the Countess II FL system (Life Technologies). All cell lines were cultured in 96-well tissue culture plates post-nucleofection (Santa Cruz Biotechnologies, sc-204447). Adherent cell lines (HEK293 and HCT116) were plated on collagen coated 96-well plates post-nucleofection (Corning, 354236). All cell lines were authenticated using Geneprint 10 from Promega and the ATCC STR database. For each cell

line, a wild type (WT) control plate was created by plating 200,000 cells/well. The cells were cultured for 48 hours and genomic lysate was prepared as described below.

[00169] Mobilized peripheral blood hematopoietic stem cells (HSCs) (AllCells # mPB018F) were thawed the day prior to nucleofection in DMEM/F12 Glutamax (Dulbecco Minimum Essential Medium, 10565-018, Gibco) supplemented with 100X MNEAA solution (Minimum Essential Medium Non Essential Amino Acids, 11140-050, Gibco) and 1000X 2-mercaptoethanol (21985-023, Gibco). HSCs were counted with a hemocytometer and plated at 1.5×10^5 /ml in the same media with 100X StemSpan™ CC100 (02690, StemCell technologies), 1000X human recombinant TPO (ThromboPOetin, 02522, StemCell Technologies) and 10,000X human recombinant GM-CSF (Granulocyte-Macrophage Colony-Stimulating factor, PHC6025, Gibco). 40 min after nucleofection, an equal volume of 2 x Antibiotic–Antimycotic (Gibco, 15240-062) containing media was added to cells. The HSCs were incubated at 37°C and 5% CO₂ and harvested 48 hours post nucleofection.

Compound Treatments

[00170] For compound treatments, sgRNP formation and nucleofection of 293T cells was performed as described above and cells were plated in 150 µl of culture medium. One hour after plating, 50 µl of media containing compound at 4x the final concentration were delivered to the cells. The DNA-PK inhibitor NU7441 (Leahy *et al.*, “Identification of a highly potent and selective DNA-dependent protein kinase (DNA-PK) inhibitor (NU7441) by screening of chromenone libraries” Bioorg. Med. Chem. Lett. (2004) 14:6083–6087) delivered in a five point dose response across a concentration range from 25 µM – 1.56 µM (1:2 dilutions). 48 hours after compound addition, cells were harvested and processed for genomic DNA preparation as described below.

Genomic Lysate

[00171] K562 cells and HSCs were transferred to a thermal cycler compatible plate and pelleted by centrifugation. After the media was gently aspirated, 50 µl EpiBio QuickExtract DNA extraction solution (Epicentre, QE09050) was added into each well and the plate was incubated at 37°C for 15 minutes. For HEK293 and HCT116 cells, media was gently aspirated from 96-well culture plates and cells were gently rinsed with 1 x Phosphate Buffered Saline (PBS). 50 µl

of EpiBio QuickExtract DNA extraction solution was added to each well and the plates were incubated at 37°C for 15 minutes. For all cell types, molecular biology grade water (Teknova) was added to each well to attain a concentration of 1000 genomic copies/ μ l for subsequent PCR reactions. After addition of water, adherent cells were then transferred to thermal cycler compatible plates. All cell types were lysed (75°C 10 min, 95°C 5 min) and the lysate was stored at -80°C.

Lentivirus Cloning

[00172] For Cas9 expression, a *Streptococcus pyogenes* Cas9 cDNA codon optimized for expression in human cells including an N-terminal 3 x FLAG tag as well as both an N- and C-terminal NLS signal as previously described (Smurnyy *et al.*, “DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells” Nat. Chem. Biol. (2014) 10:623–625), was synthesized as a gBlock and ligated into pLenti6 (Thermo Scientific). The Cas9 construct is expressed under a constitutive CMV promoter and expresses a blasticidin resistance marker co-transcriptionally with Cas9 using a T2A sequence. sgRNAs were cloned by synthesizing an oligo corresponding to the spacer sequence into a modified version of the pRSI16 lentiviral plasmid (Cellecta) containing a modified tracrRNA sequence modified as previously described including an RFP and puromycin expression cassette (Chen *et al.*, “Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system” Cell (2013) 155:1479–1491).

Lentivirus Cell Culture

[00173] Lentiviral particles were generated according to previously described methods (Hoffman *et al.*, “Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers” Proc. Natl. Acad. Sci. U.S.A. (2014) 111:3128–3133). 0.75×10^6 293T cells were plated in well of a 6-well plates 24 hours prior to transfection. For each well, cells were transfected using 1.5 μ l of TransIT reagent diluted 55.1 μ l of OPTI-MEM that was combined with 0.23 μ g of the lentiviral plasmid and 0.28 μ g of the Cellecta packaging mix (containing the psPAX2 and pMD2 plasmids that encode Gag/Pol and VSV-G respectively). Virus was harvested at 72 hours post transfection, aliquoted, and frozen at -80C for later use. Viral titer was measured using the LentiX qPCR kit and was typically in the

range of 5×10^6 TU/mL using this procedure. Stable Cas9 expression was established via lentiviral transduction at an MOI of approximately 1.0 followed by treatment with blasticidin for 96 hours to select for Cas9 expressing cells. Cas9 expression was confirmed by Western blot and by IF using a mouse monoclonal anti-Flag M2 antibody (Sigma). For sgRNA expression cells were transduced at an MOI of approximately 1 and treated with puromycin for 72 hours to select for sgRNA expressing cells.

Sequencing Library Generation for sgRNP Delivery

[00174] For sequencing sites of DNA repair following RNP delivery of sgRNA, PCR primers were designed to amplify uniquely a 180-220 base pair region of genomic DNA surrounding the target site using the Primer3 software package (Untergasser *et al.*, "Primer3 - new capabilities and interfaces" Nucleic Acids Res. (2012) 40:e115). Sequences were appended to the PCR primers such that the Illumina P5 adapter (CACTCTTCCCTACACGACGCTCTTCCGATCT; SEQ ID NO: 643) would be on the 5' end of the amplicon and the P7 adapter (GGAGTTCAGACGTGTGCTCTTCCGATCT; SEQ ID NO:644) would be on the 3' end of the amplicon (sequencing primers are listed in Table 6 and Table 7).

[00175] In the first PCR 8,000 copies of genomic template were used in a 25 μ l reaction with Q5 Master Mix (NEB, M0494L) and 500 nM final concentration of forward and reverse primers (each). The thermal cycler program for the first PCR is as follows: 1 cycle x 1 min at 98°C, 35 cycles x 10 sec at 98°C, 20 sec at 60°C, 30 sec at 72°C, 1 cycle x 2 min at 72°C.

[00176] A second PCR was performed to add index barcodes to each sample. For this PCR, the product of the first PCR was diluted 100x and 8 μ l of this dilution was used as template in a 25 μ l reaction with Q5 Master Mix and 500 nM final concentration of each of the forward (i5, see Table 6) and reverse (i7, see Table 7) primers. The thermal cycler program for PCR2 is as follows: 1 cycle x 1 min at 98°C, 12 cycles x 10 sec at 98°C, 20 sec at 60°C, 30 sec at 72°C, 1 cycle x 2 min at 72°C.

[00177] All wells were pooled after the second PCR and mixed by vortexing to form an Indexed Sample Pool (ISP). 450 μ l of SPRISelect beads were added (Beckman Coulter, B24965AA) to a new microcentrifuge tube. 500 μ l of the ISP were added to the beads. The mixture of beads and ISP was incubated for 10 minutes at room temperature. The tube was transferred to a magnetic separator and incubated for 4 minutes until the solution cleared. The beads were then rinsed

carefully (so as to not disturb the pellet) with 1 ml 85% ethanol, and incubated for 30 seconds before removing the ethanol. Beads were air-dried on the magnetic separator for 10 minutes. The library was then eluted from the beads by adding 500 µl low EDTA TE buffer (10 mM Tris, 0.5 mM EDTA, pH 8.0). Sequencing libraries were evaluated by high sensitivity gel electrophoresis (Fragment Analyzer (Advanced Analytical Technologies)). Sequencing libraries were then quantified using the KAPA Library Quantification Kit for Illumina Systems (Kapa Biosystems, KK4824) and were sequenced on a MiSeq sequencer (Illumina). MiSeq Reagent Kit v2 (300 cycles) (Illumina MS-102-2002) was used for paired-end sequencing (2 x 151), yielding 4,000-20,000 reads/sample depending on level of multiplexing.

Sequencing Library Generation for Lentiviral sgRNA Delivery

[00178] Protocols for library generation and sequencing of DNA repair structures following lentiviral delivery of the sgRNA were similar to those described above with the following minor modifications. For PCR1 primers were appended with partial 'Nextera' Illumina P5 adapter (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG; SEQ ID NO: 645) and P7 adapter (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG; SEQ ID NO: 646) sequences (sequencing primers are listed in Tables 6 and 7). For PCR2, standard Nextera v2 index and sequencing primers were used such that the adapter TCGTCGGCAGCGTC (SEQ ID NO: 647) would be on the 5' end and GTCTCGTGGGCTCGG (SEQ ID NO: 648) would be on the 3' end of the amplicon. Libraries were quantified using an in-house SYBR green qPCR protocol using primers to P5 and P7, and subsequently sequenced with 2 x 250b reads using MiSeq Reagent Kit v2 (500 cycles) (Illumina MS-102-2003).

Multiple Target Single Spacer

[00179] Approximately 16,000 target sequences were identified that exist with at least one exact copy of the target sequence in the hg38 reference genome yielding 192,955 sites with an average of 12 sites per target sequence. A filter was applied to this list to extract the target sequences where at least 50% of the sites could be uniquely amplified resulting in 51 targets covering 494 sites in the genome with an average of 10 sites per target sequence. From this group, 29 targets were tested experimentally covering 179 sites with an average of 6 sites that could be amplified uniquely per target and their DNA repair outcomes were profiled. A filter was placed on the

resulting data to exclude sites with a cleavage efficiency of less than 1% yielding a final group of 21 targets covering 127 sites in the genome.

Amplicon Sequencing Pipeline and Repair Classification

5 [00180] FASTQ read pairs generated by the Illumina MiSeq were first processed using cutAdapter to trim adapter sequences. Read pairs were then stitched into a single contig using FLASH (Magoč and Salzberg, "FLASH: fast length adjustment of short reads to improve genome assemblies" *Bioinformatics* (2011) 27:2957–2963).

[00181]). The contigs were aligned to the reference genome using the "mem" algorithm of
10 bwa. The resulting SAM files were subsequently filtered to only retain reads that overlap with the on-target amplicon location using a relational database of the known on-target amplicon location. The location of an indel within a given read was determined from the CIGAR string (Li *et al.*, "The Sequence Alignment/Map format and SAMtools" *Bioinformatics* (2009) 25:2078–2079).

15 [00182] When an aligner such as bwa attempts to align a read containing an indel to a reference genome and discovers multiple, equally good local alignments of the indel, the aligner arbitrarily selects one of the local indel alignments. To eliminate this ambiguity and attain a consistent indel justification, a "secondary alignment" of the indel in each mutant read was then performed. This secondary aligner is a custom-written Python function that re-discovers the multiple,
20 equally good local alignments for each indel and then selects a local indel alignment deterministically using the following algorithm. First, it identifies the start and end of each possible indel relative to the cut site. Second, for each possible local alignment, it calculates its "edge distance," *i.e.*, the minimum of the distance of the indel start to the cut site and the distance of the indel end to the cut site. Third, it selects the local indel alignment with the
25 smallest edge distance. If two local indel alignments have the same edge distance, it chooses the one with the smaller start site. Practically speaking, the secondary aligner chooses the alignment that is closest to the cut-site as is most biologically plausible.

[00183] Following secondary alignment, reads were then categorized as "mutant" if they contained an indel within ± 5 base pairs of the cut site; otherwise, they were categorized as "wild
30 type". Each mutant read was assigned to an "indel class" based on the indel type (insertion or deletion), start site, and length.

Heat Maps

[00184] Heat maps were generated from the processed sequencing data as follows. For a given target site, the frequencies of insertions and deletions were tallied by length (for 1 nucleotide insertions, the four nucleotides were tallied separately). Each column in the heat map represents a different target site. The color intensities in the cells correspond to frequency as a fraction of mutant reads, with the color scale ranging from 0 to 0.2 (saturation). The editing efficiency of each target site is indicated in the orange bar plot above the heat map.

MH Mask

[00185] A stringent mask for microhomology deletions was applied (FIG. 12D). For deletions with multiple possible local alignments (see above), a predicted microhomology score was first computed for each of the local alignments and then the maximum (most conservative) across all of these was retained. Deletions with a maximum microhomology score > 3 (i.e., deletions with more than one basepair of microhomology are masked) were then masked (Bae *et al.* “Microhomology-based choice of Cas9 nuclease target sites” (2014) Nat. Methods 11:705–706).

Clustering

[00186] Cluster analysis was applied to the repair outcomes at different genomic loci as follows. First, a binary vector, or bit string, was used as per Willett *et al.*, “Implementation of Nearest-Neighbor Searching in an Online Chemical Structure Search System” Journal of Chemical Information and Computer Sciences (1986) 26:36-41, to encode the indels observed at each locus. All possible insertions (based on start site and length) that start within ± 5 bp of the cut site and with length ≤ 10 were considered (there were 110 possibilities); all possible deletions (based on start site and length) fully enclosed in ± 25 bp of the cut site and with at least 1 bp present in ± 5 bp of the cut site were considered (there were 906 possibilities). For a given genomic locus, this vector of possible indels was scored as follows: any indel among the 10 most frequent by fraction of mutant reads for that locus was assigned a value 1, otherwise 0. Genomic loci that had less than 500 aligned read contigs or that had less than 10 observed indel classes were filtered out to avoid biasing downstream clustering performance metrics.

[00187] Next, given these bit strings for the genomic loci, a similarity matrix was computed using the Jaccard/Tanimoto similarity coefficient (Jaccard, P, “The distribution of flora in the alpine zone” *New Phytologist* (1912) 11:37-50); Rogers and Tanimoto, “A computer program for classifying plants” *Science* (1960) 132:1115-1118), which is the ratio of the number of elements in the intersection of two bit strings to the number of elements in their union:

$$J = M_{11} / (M_{10} + M_{01} + M_{11})$$

where M_{11} is the number of elements for which both bit strings have a value of 1, M_{10} is the number of elements for which the first bit string (but not the second) has a value of 1, and M_{01} is the number of elements for which the second bit string (but not the first) has a value of 1. This similarity matrix was used as the input to Affinity Propagation (AP) clustering (Jaccard, P, “The distribution of flora in the alpine zone” *New Phytologist* (1912) 11:37-50), an algorithm for which the number of clusters is optimized internally, rather than being specified as a parameter. The AP algorithm identifies “exemplars” — data points that are representative of other data points — by iteratively updating one of two matrices based on values in the other matrix: a “responsibility” matrix, wherein $r(i, k)$ indicates how representative a potential exemplar k is of data point i compared to other potential exemplars, and an “availability” matrix, wherein $a(i, k)$ indicates how “appropriate” it would be for data point i to choose potential exemplar k based on the evidence that k could also be an exemplar for other data points (Frey and Dueck, “Clustering by passing messages between data points” *Science* (2007) 315:972-976). The number of exemplars identified then becomes the number of clusters, and each non-exemplar data point is assigned to an exemplar.

[00188] To determine if clustering based on repair outcomes led to partitions that were similar to those corresponding to spacer sequence labels, the Adjusted Rand Index (ARI) was employed (Hubert and Arabie, “Comparing partitions” *Journal of Classification* (1985) 2:193–218.). Given a “ground truth” partitioning U of a set of labeled data points and an observed partitioning V , the Rand Index (RI) (Rand, “Objective criteria for the evaluation of clustering methods” *Journal of the American Statistical Association* 66:846–850) is the fraction of pairs of data points that are concordant across U and V (i.e., partitioned together in U and partitioned together in V , or in different partitions in U and in different partitions in V). The ARI is a version of the RI that is corrected for the expectation under randomness, such that an ARI of 0 indicates that the partition membership of V is random and an ARI of 1 indicates that the partition membership of V

perfectly matches that of *U*. For a detailed explanation and demonstration, see the Supplementary Information of (Yeung and Ruzzo, “Principal component analysis for clustering gene expression data” *Bioinformatics* (2001) 17:763-774).

5 [00189] The following figures display data and outcomes from the examples herein:

[00190] **FIGS. 1A-1C** depict the profiling of DNA repair outcomes after Cas9 cleavage. **FIG. 1A:** (i), Cell editing workflow. (ii) Sequencing library steps. (iii) Indel class visualization following Cas9 cleavage at a target (SEQ ID NO: 624) in *JAK1* (Spacer 54, Table 1). The position of each insertion class and deletion class is plotted relative to the cut site. Throughout the text, indel visualizations are annotated as follows: (right) each class is annotated with frequency (fraction of total reads and fraction of mutant reads) and the number of reads observed. Classes are ranked by frequency (classes with frequency < 0.01 (fraction of mutant reads) are not displayed). **FIG. 1B:** Visualization of the five most frequent indel classes and wild-type (WT) at the same target in three cell lines (the first replicate (R1) of each is displayed). **FIG. 1C:** An indel frequency heat map by length for each cell line (three replicates and WT control shown). Throughout the text, heat maps are annotated as follows: insertions of 1 to 8 nucleotides are displayed. Single base insertions are separated by nucleotide (A, T, C, G). Deletion lengths of 1 to 50 nucleotides are displayed. The color intensity scales with frequency as a fraction of mutant reads up to 0.2. The bar graph on the right displays the mean frequency of each indel l. The bar graph above displays editing efficiency (“Edit. Eff.”) as a fraction of total reads.

[00191] **FIGS. 2A-2C** show that DNA repair profiles are unique to each spacer sequence.

FIGS. 2A and 2B: A matrix of the (Jaccard/Tanimoto) similarity of the top ten indel classes across pairs of 69 target sites in HEK293, K562, and HCT116 cell lines comparing sgRNP-only delivery of reagents (**FIG. 2A**) and sgRNP and constitutive delivery of reagents (**FIG. 2B**).

Targets with the same spacer label (within the ticks) are different experimental replicates of each cell type targeted by the same sgRNA. A similarity score of 1 represents complete overlap of the top ten indel classes between two sites, whereas 0 represents no overlap of the top ten indel classes between two sites. **FIG. 2C:** Adjusted Rand Index (ARI) values from cluster analysis (see Examples) of **FIG. 2A** and **FIG. 2B**.

[00192] FIGS. 3A-3F show the results of an experiment using multiple target single spacers (MTSS) sequences and display that DNA repair outcomes at Cas9-mediated DSBs are sequence-dependent. **FIG. 3A:** Genomic coordinates (hg38) that contain exact copies of the spacer 15 sequence (SEQ ID NO: 649). **FIG. 3B:** A heat map of the frequencies of indels by length for the seven targets listed in FIG. 3A from replicate experiments plus wild-type (WT) controls 48 hours after nucleofection into HEK293 cells. **FIG. 3C:** A heat map of the frequencies of indels by length in HEK293 cells for 22 different spacer groups (outlined in grey boxes). Each target sequence occurs at 2-14 times in the genome. For each target site within each spacer group, three experimental replicates and a WT control are displayed (within the minor ticks). **FIG. 3D:** A heat map of the frequencies of indels by length in K562 cells for the same 22 spacer groups (outlined in grey boxes) as described for FIG. 3C. **FIGS. 3E and 3F:** A matrix of the (Jaccard/Tanimoto) similarity of the top ten indel classes across pairs of target sites in HEK293 cells (**FIG. 3E**) and K562 cells (**FIG. 3F**). Targets with the same spacer label (within the ticks) are different genomic loci targeted by the same sgRNA. ARI values from subsequent cluster analysis are displayed.

[00193] FIGS. 4A-4I show that the distribution of DNA repair outcomes after Cas9 cleavage changes over time. **FIG. 4A:** Heat map of DNA repair outcomes in HEK293 cells for 96 different spacers, each of with three experimental replicates for multiple time points (4, 8, 16, 24 and 48 hours) and a wild type control (within the minor ticks). **FIGS. 4B and 4F:** Heat maps for the indicated spacer (**FIG. 4B:** spacer 13; **FIG. 4F:** spacer 54) showing single experiments at each time point for three cell lines. **FIGS. 4C and 4G:** Bar graphs of indel frequencies by length for the indicated spacer (**FIG. 4C:** spacer 13; **FIG. 4G:** spacer 54) displayed as a fraction of mutant reads (mean and standard deviation across triplicates) in three cell lines at the 48 hour time point. **FIGS. 4D and 4H:** A heat map of spacer 13 (**FIG. 4D**) or spacer 54 (**FIG. 4H**) showing single experiments at each time point for three cell lines after applying a stringent microhomology mask (MH_score >3, see FIG. 12D). **FIGS. 4E and 4I:** Bar graph of indel frequencies by length displayed as a fraction of mutant reads (mean and standard deviation across triplicates) at the 48 hour time point for spacer 13 (**FIG. 4E**) or spacer 54 (**FIG. 4I**) after applying a stringent microhomology mask (MH_score >3).

[00194] FIGS. 5A-5D show that chemical perturbation of c-NHEJ promotes a subset of DNA repair outcomes after Cas9 cleavage. **FIG. 5A:** Heat map representing DNA repair classes

present 48 hours after sgRNP introduction into HEK293T cells (Spacer 93, Table 1). Cells were treated with the DNA-PK inhibitor NU7441 (Leahy *et al.*, “Identification of a highly potent and selective DNA-dependent protein kinase (DNA-PK) inhibitor (NU7441) by screening of chromenone libraries” *Bioorg. Med. Chem. Lett.* (2004) 14:6083–6087) in a 2 fold dilution series ranging from 1.56 μ M to 25 μ M. Each concentration was performed in duplicate.

Untreated control replicates are shown on the right. Mean frequency plots are shown on the left comparing untreated samples with samples treated with DNA-PK inhibitor NU7441 (average mean frequency of entire dilution series 1.56 μ M-25 μ M displayed). Arrows indicate the repair classes that change frequency after treatment with NU7441 (down arrow indicates a decrease in mean frequency in the presence of inhibitor, up arrow indicates an increase in mean frequency in the presence of inhibitor). **FIG. 5B:** Visualization of the five most frequent indel classes and wild-type (WT) of the same target shown in FIG. 5A (the first replicate (R1) of each is displayed). NU7441 inhibitor concentrations 1.56, 3.13 and 6.25 μ M are shown. **FIG. 5C:** Same as in FIG. 5A for target spacer 54 (Table 1). **FIG. 5D:** Same as in FIG. 5B for target spacer 54.

[00195] FIGS. 6A-6E show high frequency in-frame mutation after Cas9 cleavage of a target in *CD34*. **FIG. 6A:** Genomic location (hg19 coordinates; SEQ ID NOS: 625 and 626) of a target in the *CD34* gene with the protospacer adjacent motif (PAM) and Cas9 cut site indicated by an arrow. The entire protospacer sequence is boxed (Spacer 16 in Table 1). Deletion of the three nucleotide sequence, TTG, shaded, is the most frequent indel after either sgRNP delivery or constitutive expression of sgRNA to direct Cas9 cleavage activity to this site. **FIG. 6B:** A visualization of a subset of the indel classes (the five most frequent) and wild-type (WT) at the *CD34* target in three different cell lines using sgRNP delivery and in two different cell lines for constitutive expression of Cas9/sgRNA as indicated on the left (a single replicate of each is displayed). **FIG. 6C:** A heat map of the frequencies of indels by length at the *CD34* target in HEK293, K562 and HCT116 cell lines (three experimental replicates of each and a WT control) 48 hours post-sgRNP delivery. **FIG. 6D:** A heat map of the frequencies of indels by length at the *CD34* target in HEK293 and HCT116 cell lines 11 and 14 days post constitutive expression of Cas9/sgRNA (single replicate of two different time points and a WT control (Cas9-only)). **FIG. 6E:** Bar graph of indel frequencies by length displayed as a fraction of mutant reads (mean and standard deviation across three experimental replicates) in three cell lines 48 hours post sgRNP delivery.

[00196] **FIGS. 7A-7F** show that DNA repair outcome profiling in cell lines is predictive for human primary cells. DNA repair outcome after cleavage of a *BRCA2* target results in two dominant products. **FIG. 7A:** Genomic location (hg19 coordinates) of a target in the *BRCA2* gene (SEQ ID NOS: 627 and 628); PAM, boxed in light grey; protospacer, grey box. Cas9 cut site is indicated by an arrow. The entire protospacer sequence is boxed in grey (Spacer 19 in Table 1). Deletion of the four-nucleotide sequence, shaded, and a single nucleotide insertion represent the most frequent indels after Cas9 cleavage. **FIG. 7B:** Visualization of a subset (the five most frequent) of the indel classes and wild-type (WT) at the *BRCA2* target in three different cell lines and hematopoietic stem cells (HSCs) for sgRNP and three different cell lines for constitutive expression of Cas9/sgRNA as indicated on the left (a single replicate of each is displayed). **FIGS. 7C and 7D:** Heat maps of the frequencies of indels by length at the *BRCA2* target in various cell types (three experimental replicates and a WT control) 48 hours post-sgRNP delivery (**FIG. 7C**) or 11 and 14 days post constitutive expression of Cas9/sgRNA (**FIG. 7D**). **FIG. 7E:** Bar graph of indel frequencies by length displayed as a fraction of mutant reads (mean and standard deviation across three experimental replicates) in three cell lines and HSCs 48 hours post sgRNP delivery. **FIG. 7F:** Use of DNA repair profiles to restore the reading frame of specific mutant alleles of *BRCA2*. 1. Wild type *BRCA2* locus (hg19 coordinates; (SEQ ID NOS: 627 and 628). 2. Sequence of *BRCA2* mutant allele (dbSNP ID: 80359550; (SEQ ID NOS: 629 and 630) missing a “C” nucleotide shifting the frame of *BRCA2* resulting in a premature stop codon (box with asterisk). 3. Genomic location of a target site on the *BRCA2* mutant allele shown in (2) with the protospacer adjacent motif (PAM) boxed in light grey and Cas9 cut site indicated by an arrow. The entire protospacer sequence is boxed in grey. 4. An insertion of a single nucleotide after Cas9-cleavage (SEQ ID NOS: 631 and 632) at the target site shown in (3) would restore the frame of the *BRCA2* mutant allele. Depending on the nucleotide that was inserted during the DNA repair reaction, the fifth amino acid from the left would be a Phenylalanine, Isoleucine, Valine, or Leucine (WT). 5. Sequence of a *BRCA2* mutant allele (dbSNP ID: 397507830; SEQ ID NOS: 633 and 634) containing a duplication of “CTTA” resulting in a frame shift and a premature stop codon (box with asterisk). 6. Genomic location of a target site on the *BRCA2* mutant allele shown in (5) with the protospacer adjacent motif (PAM) boxed in light grey and Cas9 cut site indicated by an arrow. The entire protospacer sequence is boxed in grey. 7. A four base deletion after Cas9-cleavage at the target site shown in (6) would

restore the frame of the *BRCA2* mutant allele (SEQ ID NOS: 635 and 636). The resulting allele would differ from the WT allele by a single amino acid (Leucine instead of a Threonine, fourth amino acid from the left).

[00197] **FIGS. 8A-8E** show DNA repair outcomes are similar comparing across replicates and reagent delivery methods at a *JAK1* target. **FIG. 8A:** Genomic location (hg19 coordinates; SEQ ID NOS: 637 and 638) of a target in the *JAK1* gene with the protospacer adjacent motif (PAM) boxed in light grey and Cas9 cut site indicated by an arrow. The entire protospacer sequence is boxed in grey. This is the same target shown in FIG. 1 (Spacer 54 in Table 1). **FIG. 8B:** A visualization of a subset of the indel classes (the five most frequent) and wild-type (WT) at the *JAK1* target comparing three experimental replicates in HEK293 cells, as indicated on the left, 48 hours after sgRNP delivery. **FIG. 8C:** A heat map of the frequencies of indels by length at the *JAK1* target of the same three experimental replicates and a WT control in HEK293. **FIG. 8D:** A visualization of a subset of the indel classes (the five most frequent) and wild-type (WT) at a *JAK1* target comparing HEK293, K562 and HCT116 cell lines, as indicated on the left, 14 days after constitutive expression of Cas9/sgRNA. **FIG. 8E:** A heat map of the frequencies of indels by length at the *JAK1* target (three replicates and a WT control) in HEK293, K562, and HCT116 cell lines 48 hours after sgRNP delivery compared with 11 and 14 days after constitutive expression of Cas9/sgRNA in the same parental cell lines (plus WT control (Cas9-only)).

[00198] **FIGS. 9A-9F** show that DNA repair profiles are unique to each target. **FIG. 9A:** Genomic location of a target in the *XRCC5* gene (SEQ ID NOS: 639 and 640) with the protospacer adjacent motif (PAM) boxed in light grey and Cas9 cut site indicated by an arrow (Spacer 51 in Table 1). **FIG. 9B:** Genomic location of a target in the *JAK3* gene (SEQ ID NOS: 641 and 642); (Spacer 66 in Table 1). **FIG. 9C:** Visualization of a subset of the indel classes (the five most frequent) and wild-type (WT) at the *XRCC5* target comparing HEK293, K562 and HCT116 cell lines for both sgRNP and 14 days after constitutive expression of Cas9/sgRNA, as indicated on the left (a single replicate of each is displayed). **FIG. 9D:** A heat map of the frequencies of indels by length at the *XRCC5* target (three experimental replicates and a WT control) in HEK293, K562, and HCT116 cell lines 48 hours after sgRNP delivery compared with 11 and 14 days after constitutive expression of Cas9/sgRNA in the same parental cell lines

(Lenti) plus WT control (Cas9-only)). **FIGS. 9E and 9F:** same as FIGS. 9C and 9D at the *JAK3* target.

[00199] FIGS. 10A and 10B show visualizations of DNA repair outcomes at seven sites in the human genome with the same spacer sequence. **FIG. 10A:** Visualization of a subset (the five

most frequent) of the indel and wild-type (WT) classes for seven sites in spacer group 15 in HEK293 (15a-15g), as indicated on the left (a single replicate of each is displayed). Genomic coordinates for each site are described in FIG. 3A. Corresponding heat map visualization of indel distribution for each site is shown in FIG. 3B. **FIG. 10B:** Visualization of the top five most frequent indel classes for spacer 15 WT controls (unedited) at sites *a* and *e* in HEK293

(related to FIGS. 3A and 3B). Each class expressed as a fraction of mutant reads. The number of reads attributed to each class is shown in the second column on the right. The small number of reads that contain indels in WT controls are attributed to sequencing errors.

[00200] FIGS. 11A-11D show that DNA repair outcomes are more similar within spacer groups than between spacer groups. **FIGS. 11A and 11B:** A matrix of the (Jaccard/Tanimoto) similarity

of indel classes with a frequency ≥ 0.01 (as a fraction of mutant reads) across pairs of target sites in HEK293 cells (**FIG. 11A**) and K562 cells (**FIG. 11B**). Spacer group labels are indicated on the horizontal and vertical axes. Targets with the same spacer label (within the minor ticks) indicate different genomic loci targeted by the same sgRNA (Multiple Target Single Spacer (MTSS)). A value of 1 represents complete overlap of the top ten indel classes between two

sites. A value of 0 represents no overlap of the top ten indel classes between two sites. **FIG. 11C:** Cluster analysis of the 127 target sites by DNA repair outcomes using the Jaccard similarity matrix as input to Affinity Propagation (AP) clustering. The “ground truth” cluster membership, where targets are simply arranged by spacer group label, is shown in the left column. The cluster membership resulting from clustering analysis of DNA repair outcomes for HEK293 and K562 cells is shown in the center and right columns, respectively. Each target site is displayed as its spacer label (1-22). The number of sites per spacer group is the number of points in each cell of the ground truth column. **FIG. 11D:** ARI scores calculated based on pairwise overlap and clustering of top 10 indels or ranked by frequency for all 127 spacer sites.

[00201] FIGS. 12A-12D show that similarity of DNA repair outcomes within spacer groups is not fully explained by microhomology. **FIG. 12A:** A matrix of the (Jaccard/Tanimoto) similarity of top ten indel classes across pairs of target sites in HEK293 cells. Spacer group

labels are indicated on the top and left axes. Targets with the same spacer label (within the minor ticks) indicate different genomic loci targeted by the same sgRNA (Multiple Target Single Spacer (MTSS)). A value of 1 represents complete overlap of the top ten indel classes between two sites. A value of 0 represents no overlap of the top ten indel classes between two sites.

Microhomology masks were applied with varying stringencies as indicated on the right of each matrix (see FIG. 12D) for details). **FIG. 12B:** same as in FIG. 12A for K562. **FIG. 12C:** ARI scores calculated based on pairwise overlap and clustering of top 10 indels for all 127 spacer sites with and without microhomology masking (see FIG. 12D). **FIG. 12D:** Definition of microhomology masks.

[00202] **FIGS. 13A-13B** show sgRNP time course data from K562 and HCT116 cells for 96 target sites. **FIG. 13A:** Heat map of DNA repair outcomes in K562 cells for 96 different spacers. For each spacer, five time points in each of three experimental replicates and a wild type control are displayed (4, 8, 16, 24 and 48 hours) for a total of 16 data points per spacer (within the minor ticks). Target sites are arranged based on the results of hierarchical clustering (Target sites shown for HEK293 in FIG. 4A are arranged in the same order). **FIG. 13B:** Same as in FIG. 13A for HCT116. Target sites are arranged in the same order as in FIGS. 13A and 4A. Zooms of individual spacers in this series are shown in FIGS. 4 and 14.

[00203] **FIGS. 14A-14L** show that the distribution of DNA repair outcomes after Cas9 cleavage changes over time in a cell type-dependent manner. **FIGS. 14A, 14E and 14I:** Heat map zooms of the indicated spacer (**FIG. 14A:** spacer 1; **FIG. 14E:** spacer 51; **FIG. 14I:** spacer 66) showing single replicates of each time point comparing three cell lines. (**FIGS. 14B, 14F and 14J**) Bar graphs of indel frequencies by length for the indicated spacer (**FIG. 14B:** spacer 1; **FIG. 14F:** spacer 51; **FIG. 14J:** spacer 66) displayed as a fraction of mutant reads (mean and standard deviation across triplicates) in three cell lines at the 48 hour time point. DNA repair outcomes for three spacers with a computational mask of microhomology deletions (**FIGS. 14C and 14D;** **FIGS. 14G and 14H;** **FIGS. 14K and 14L**) (see FIG. 12D). **FIGS. 14C, 14G and 14K:** Heat map zooms of the indicated spacer (**FIG. 14C:** spacer 1; **FIG. 14G:** spacer 51; **FIG. 14K:** spacer 66) showing single replicates of each time point comparing three cell types after applying a stringent computational mask of microhomology deletions (MH_score > 3). **FIGS. 14D, 14H and 14L:** Bar graphs of indel frequencies by length for the indicated spacer (**FIG. 14D:** spacer 1; **FIG. 14H:** spacer 51; **FIG. 14L:** spacer 66) displayed as a fraction of mutant reads (mean and

standard deviation across triplicates) in three cell lines at the 48 hour time point after applying a stringent computational mask of microhomology deletions (MH_score > 3).

[00204] FIG. 15 displays the frequencies of nucleotide insertions as a function of the position-17 nucleotide in the protospacer.

5

Example 1

DNA Repair Outcomes at Cas9 Breaks

[00205] The repair outcomes following Cas9 cleavage of double-stranded DNA, which result in blunt end products (Gasiunas *et al.*, “Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria” Proc. Natl. Acad. Sci. U.S.A. (2012) 109:E2579–86; Jinek *et al.*, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” Science (2012) 337:816–821), were profiled using computational tools developed to categorize indels from our cell-based assay. Results are shown in FIGS. 1A-1C.

[00206] To this end, HEK293 cells were transfected with pre-assembled complexes of Cas9 protein and sgRNA (single-guide RNA ribonucleoprotein complexes (sgRNPs)) (Lin *et al.*, “Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery” Elife (2014) 3:e04766.2014; Schumann *et al.*, “Generation of knock-in primary human T cells using Cas9 ribonucleoproteins” Proc. Natl. Acad. Sci. U.S.A.(2015) 112:10437–10442). DNA repair patterns following Cas9 cleavage were analyzed by amplicon sequencing (FIG. 1A). Sequencing reads were assigned to a specific indel class based on the indel type (insertion or deletion), start site, and length (or to the wild type class), and then the frequency of each class was calculated as a fraction of aligned reads or as a fraction of mutant reads (*see*, FIG. 1A). A window of 50 base pairs on either side of the Cas9 cut site was captured for all targets, which defines the range within which the distribution of DNA repair outcomes was monitored. Surprisingly, the identities and frequencies of specific indel classes were similar between independent experimental replicates, and this reproducibility extended to the specific start and end coordinates for each class (FIGS. 8B, 8C). Furthermore, the highest frequency indel classes were also preserved when the same site was cut in three different cell lines (HCT116, HEK293, K562; FIG. 1B). However, the relative frequencies of those classes were not identical between cell lines. HCT116 cells showed a higher frequency of insertion of a single adenine base and a lower frequency of a 19-base pair deletion compared to either HEK293

30

or K562 cell lines as a fraction of total edits. HCT116 is a mismatch repair-deficient cell line and this attribute might contribute to the observed pattern, potentially through lack of recruitment of Exo1 at DSBs (De las Alas *et al.*, “Loss of DNA mismatch repair: effects on the rate of mutation to drug resistance” J. Natl. Cancer Inst. (1997) 89:1537–1541; Goellner *et al.*,

5 “Exonuclease 1-dependent and independent mismatch repair” DNA Repair (2015) 32:24–32).

[00207] Next, DNA repair outcomes at the same spacer target sites were assessed by delivering sgRNAs via lentiviral transduction into cells that express Cas9 constitutively (Zhou *et al.*, “High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells”

Nature (2014) 509:487–491; Wang *et al.*, “Genetic screens in human cells using the CRISPR-

10 Cas9 system” Science (2014) 343:80–84). DNA repair outcomes were measured at a later time point in lentiviral delivery experiments than in sgRNP delivery experiments (11 and 14 days post-delivery compared to 48 hours post-delivery) to gauge what constituted a mature DNA

repair profile and when this profile was achieved in each delivery method. sgRNPs are transient and are metabolized by the cell; whereas constitutive expression of Cas9 and sgRNA provides an

15 environment where a target site can be re-cleaved until an editing outcome is achieved that prevents further cutting of the target.

[00208] The repair outcomes were quantified to determine whether those resulting from one

spacer target (across all replicates, cell types, and reagent delivery methods) were more similar to each other than to the repair outcomes of samples corresponding to a different spacer target from

20 a set of 69 spacers. First, a matrix of the similarity in DNA repair outcomes between all pairs of samples was generated, using Jaccard/Tanimoto similarity (FIGS. 2A-2C). The top 10 indels were scored for each of the 69 genomic loci (in a binary list of possible indels) and the overlap in this list was then compared pair-wise across all loci (FIGS. 2A-2C). The landscape of indel

classes was consistent at an individual site when compared across replicates, cell types and

25 reagent delivery methods (mean Jaccard/Tanimoto similarity score of 0.70 +/-0.06 (*see*, FIGS. 6-9). In contrast, when DNA repair outcomes were compared across different spacer target sites,

the landscapes of indel classes were not consistent (mean similarity score of 0.09 +/-0.03) (FIGS. 5, 6, 8 and 9). Samples corresponding to the same target sequence had higher similarity scores,

indicating more overlap in indel landscapes, than samples corresponding to different target

30 sequences (FIGS. 2A, 2B, along the diagonal). Additionally, cluster analysis (via Affinity

Propagation clustering) was applied to the indel similarity matrix to partition the samples by

repair outcome similarity (without referencing spacer labels) and the resulting cluster membership was then compared to the “ground truth” where samples were simply grouped by spacer label. An ARI (Adjusted Rand Index) value of 0.92 was attained for sgRNP-only delivery and an ARI value of 0.76 was attained when sgRNP and lentivirus delivery were combined (FIG. 2C) (An ARI value of 1 indicates that the observed clustering perfectly matches the ground truth, while an ARI value of 0 indicates a random distribution of labels in the observed clustering.) These statistics of similarity strongly corroborate the heat map visualizations and indicate that each target sequence has a characteristic DNA repair indel landscape.

[00209] Thus, DNA repair outcomes at Cas9 breaks are nonrandom and unique to each target.

Example 2

DNA Repair Signatures after Cas9 Cleavage

[00210] The characteristic patterns of DNA repair outcomes observed for different target sequences could be the consequence of either sequence-dependent or genomic context-dependent factors. To distinguish between sequence and context dependence, a series of sgRNAs were designed such that the spacer sequence of each sgRNA perfectly matched multiple locations in the human genome (multiple target single spacers; MTSS), whereas the sequences flanking the protospacer ensured that each of these sites could be amplified uniquely by PCR (Tables 3; 4A; and 4B). Twenty-two spacer sequences were selected that met these criteria; the targeted sequences each occur 2-14 times in the genome, providing a total of 127 sites for analysis (FIGS. 3A-3F). Consistent with the initial observations, distinct patterns of the highest frequency indel classes were conserved across experimental replicates at each of the individual sites (FIGS. 3B, 10A). Furthermore, the DNA repair patterns at each of the sites targeted by the same spacer sequence were strikingly similar to each other. For example, “spacer 15” targets seven sites in the genome that can be uniquely amplified; for each of these sites, two dominant repair outcomes, a single insertion of “A” and a single base deletion, comprise 53-67% of the total observed repair events. In addition, less frequent yet reproducible repair events resulting in 8- and 10-base pair deletions and a 2 base pair insertion were observed (FIGS. 3B, 10A). Furthermore, for all of the target sites for a given spacer sequence, the same pattern was also observed in different cell types (FIGS. 3C-3D). The relative editing efficiencies were different between sites of the same spacer group and between the same site comparing between cell types;

however, the overall pattern of DNA repair outcomes was the same. The distribution of the mean indel frequencies by length across all samples in the twenty-two spacer groups was nearly identical when compared across different cell types (FIGS. 3C-3D (right distributions)). The same methods described above were used to assess whether the repair outcomes of the genomic target sites of one spacer group were more similar to each other than to the repair outcomes of the genomic target sites of the other 21 spacer groups (FIGS. 3E, 3F, 11A, 11B). Genomic loci corresponding to the same spacer sequence had higher similarity scores, indicating more overlap in indel landscapes, than genomic target sites corresponding to different spacer sequences (FIGS. 3E-3F, along the diagonal).

[00211] Cluster analysis was applied to the indel similarity matrix to partition the 127 loci by repair outcome similarity, and an ARI value of 0.87 was attained for both HEK293 and K562 cell lines from this analysis indicating that the repair outcomes at sites with the same target sequence located in diverse regions of the genome are more similar to each other than to the repair outcomes resulting from different sgRNAs (FIG. 11D). Comparable results were obtained when an indel frequency-based threshold was used rather than a rank-based threshold for scoring indels (FIGS. 11A-11B, 11D). These computational analyses demonstrate that repair outcomes at target sites primarily cluster within spacer groups. There were a small number of cases where samples corresponding to the same spacer group were split into multiple clusters or were clustered with samples corresponding to other spacer groups (*e.g.*, HEK293 groups 12 and 14, and K562 groups 2 and 5 (FIG. 11C)). In the case of spacer group 2, there were sequence differences 5' of the target sequence such that a region of microhomology was present at site 1 but not site 2, coincident with a 12-base pair deletion at site 1 but not site 2. Taken together, these findings strongly evidence that the DNA repair patterns observed after sgRNP cleavage are nonrandom and are dependent on DNA sequence rather than functional genomic context.

[00212] It has been proposed that a subset of deletion outcomes after Cas9 cleavage are due to MMEJ where a reproducible deletion occurs between regions of microhomology (Nakade *et al.*, "Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9" *Nat Commun* (2014) 5:5560; Bae *et al.*, "Microhomology-based choice of Cas9 nuclease target sites" *Nat. Methods* (2014) 11:705–706). DNA repair patterns from the MTSS experiment were analyzed to determine whether microhomology could be a driver of the non-random patterns. A microhomology (MH) score

was predicted for each repair class at each target site (FIG. 12D). MH scores were used to filter potential MH-dependent sites from the repair patterns at various stringencies. After MH masking, ARI values dropped to a range of 0.66-0.74 depending on the MH-mask stringency, yet strong clustering of all 127 loci remained indicating that c-NHEJ repair outcomes are a substantial portion of each repair landscape (FIG. 12C). Furthermore, similarity matrices generated after MH-masking revealed less indel class overlap between loci with the same spacer sequence, but a strong pattern remained (FIGS. 12A-12B). Examination of the change in overall indel pattern after MH masking showed that large deletions were masked, supporting the hypothesis that large deletion formation may rely on MH (FIGS. 4, 14). The pattern that remains after the most stringent MH mask (score >3, FIG. 12D) contains all insertions and primarily small deletions that are specific to a particular spacer sequence and are most probably the result of DSB repair by the c-NHEJ machinery; whereas, larger deletions are more likely the result of microhomology and MMEJ-dependent repair. However, even the most stringent computational MH-mask, where only 1-2 nucleotides of homology are required, likely provides only an approximation of the MMEJ-dependent repair landscape. It should also be noted that, although larger deletions were filtered out using the MH mask, MH scores were not predictive for the highest frequency deletions observed for a given target site.

Example 3

Distribution of DNA Repair Outcomes after Cas9 Cleavage

[00213] Genome editing outcomes are determined by the DNA repair pathway that is engaged. These pathways are, in turn, influenced by cell cycle stage. MMEJ initiates with end resection at a DSB (Truong *et al.*, “Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells” Proc. Natl. Acad. Sci. U.S.A. (2013) 110:7720–7725), an activity that is dependent on CDK levels (Huertas *et al.*, “CDK targets Sae2 to control DNA-end resection and homologous recombination” Nature (2008) 455:689–692; Huertas and Jackson, “Human CtIP mediates cell cycle control of DNA end resection and double strand break repair” J. Biol. Chem. (2009) 284:9558–9565). c-NHEJ, however, can operate throughout the cell cycle (Aylon *et al.*, “The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle.” (2004) EMBO J. 23:4868–4875; Ira *et al.*, “DNA end resection, homologous

recombination and DNA damage checkpoint activation require CDK1” Nature (2004) 431:1011–1017). Given that heat map profiling and MH masking could distinguish between c-NHEJ-dependent small indels and MH-dependent deletions, the evolution of repair outcomes over a 48-hour period of time was investigated. Genomic DNA was harvested 4, 8, 16, 24 and 48 hours post nucleofection from HEK293, K562 and HCT116 cell lines edited with 96 different sgRNPs (FIGS. 4A, 13A, 13B). The overall distribution of repair classes was similar for HEK293 and K562 cell lines at each time point, whereas the frequencies of insertions and smaller deletions were relatively higher in HCT116 cells (FIGS. 4A, 13A, 13B, right distribution of mean frequencies). For example, when the distribution of indel frequencies by length for spacer 13 was compared between different cell lines harvested at 48 hours, HCT116 cells had a lower frequency of 3 and 8 base pair deletions and a higher frequency of single base pair insertions (FIGS. 4B-4C). Similarly, with spacer 54, HCT116 cells had a lower frequency of 4 and 19 base pair deletions compared to HEK293 and K562 cells; however, other deletion classes were at similar frequencies as observed in HEK293 and K562 cells (FIGS. 4F, 4G). When the MH-mask was applied to these sites, the deletion lengths with dissimilar frequency in HCT116 cells disappeared (FIGS. 4D, 4E, 4H, 4I).

[00214] Analysis of the full set of 96 spacers at each time point revealed a consistent pattern where single base pair insertions and small deletions (1-2 base pairs) were dominant at early time points (expressed as a fraction of the total classes) and larger deletions became more prominent at later time points (FIGS. 4, 13, 14). However, large deletions (> 5 base pairs) were generally less prominent in HCT116 cells throughout the time course as a fraction of total repair classes, consistent with the overall pattern (FIGS. 4, 13, 14). When a stringent MH mask was applied to the time course data, the larger deletions that appear at later time points were filtered out, indicating that at least a subset of deletions are likely dependent on MH and that the reduced level of large deletions in HCT116 is potentially due to a deficiency in MMEJ (FIGS. 4D, 4E, 4H, 4I, 14C, 14D, 14G, 14H, 14K, 14L).

[00215] Thus, the distribution of DNA repair outcomes after Cas9 cleavage changes over time.

Example 4

Effect of Chemical Perturbation of c-NHEJ after Cas9 Cleavage

[00216] To test the hypothesis that suppression of c-NHEJ would alter the DNA repair landscapes, favoring MMEJ repair outcomes, a chemical inhibitor of DNA-PK, NU7441, was added to HEK293T cells one hour post nucleofection in a five point dose response (Leahy *et al.*, “Identification of a highly potent and selective DNA-dependent protein kinase (DNA-PK) inhibitor (NU7441) by screening of chromenone libraries” *Bioorg. Med. Chem. Lett.* (2004) 14:6083–6087; Robert *et al.*, “Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing” *Genome Med* (2015) 7:93). Genomic lysate was harvested 48 hours post nucleofection and processed as in previous experiments (*see*, description of FIG. 5A). Chemical inhibitors of the c-NHEJ pathway have been used in conjunction with delivery of Cas9/sgRNA before (Maruyama *et al.*, “Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining” *Nat. Biotechnol.* (2015) 33:538–542; Chu *et al.*, “Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells” *Nat. Biotechnol.* (2015) 33:543–548; Robert *et al.*, “Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing” *Genome Med* (2015) 7:93); however, this was in the context of using an exogenous donor template and attempting to promote HDR by suppressing indel formation by c-NHEJ.

[00217] At the lowest concentration of inhibitor (1.56 μ M), a change of the DNA repair patterns at 12 different target sites compared with untreated samples was apparent (FIG. 5). As inhibitor concentration increased, the mean frequency of single base insertions and small deletions (< 3 base pairs) decreased, whereas the mean frequency of a subset of large deletions (> 3 base pairs) present in the DNA repair profiles increased (FIGS. 5A, 5C). This experiment evidences that the suppression of c-NHEJ enhances DNA repair by MMEJ pathways after DSB formation by Cas9. These data also indicate that we can effectively segregate DNA editing events produced by different DNA repair machineries by DNA repair profiling and that individual repair profiles can be modulated by suppressing or enhancing individual components of these pathways. The ability to promote a subset of editing outcomes by chemical perturbation, as has already been described for HDR, can be applied to correct a mutant allele in the absence of a donor template (e.g. FIG. 7E).

[00218] Thus, chemical perturbation of c-NHEJ alters DNA repair profiles after Cas9 cleavage.

Example 5

Prediction of the Most Frequently Inserted Nucleotide at Cas9 Target Sites

[00219] One of the more common DNA repair outcomes following Cas9 cleavage of targets is a single base pair insertion. Furthermore, the frequencies with which each nucleotide is inserted (A, T, C, or G) vary widely across targets. To test the hypothesis that the cut-site nucleotides of the protospacer sequence may be predictive of which nucleotide is most frequently inserted (Cas9 cleaves between positions 17 and 18 in the protospacer) the following experiment was conducted.

[00220] A data set of repair outcome data for 216 Cas9 target sites in K562 cells was generated. A multiclass classifier was performed using linear discriminant analysis (James *et al.*, An Introduction to Statistical Learning, First Edition, 2013, Springer, ISBN 978-1461471370) via the Python Scikit-learn library (Pedregosa *et al.*, “Sckit-learn: Machine Learning in Python” J. Mach. Learn. Res. (2011) 12:2825-2830). For each target site, the cut-site nucleotide of interest (spacer position 17 or 18) was represented in a feature vector with three indicator variables. Predictive performance was assessed using the stratified, five-fold cross-validated accuracy (percentage of predictions that were true positives or true negatives). The baseline accuracy was determined by permuting the outcome labels and assessing performance.

[00221] Using the above methods, it was determined that the PAM-proximal, cut-site nucleotide (protospacer position 18) was not predictive of which nucleotide was most frequently inserted at Cas9 targets. Cross-validated accuracy was $33\% \pm 6\%$ (compared to a baseline of $29\% \pm 7\%$ when labels were permuted). However, the PAM-distal, cut-site nucleotide (protospacer position 17) was highly predictive of which nucleotide was most frequently inserted. Cross-validated accuracy was $83\% \pm 4\%$. FIG. 15 displays the frequencies of nucleotide insertions as a function of the position-17 nucleotide in the protospacer.

[00222] Next, linear regression was used to evaluate the performance of using the position-17 nucleotide to predict the frequencies with which each nucleotide was inserted. Linear regression was performed via the Python Scikit-learn library (Pedregosa *et al.*, “Sckit-learn: Machine Learning in Python” J. Mach. Learn. Res. (2011) 12:2825-2830). Predictive performance was assessed using the stratified, five-fold cross-validated R^2 . The cross-validated R^2 values were

0.52 $\pm$ 0.12 for insertion of *A*; 0.54 $\pm$ 0.13 for insertion of *T*; 0.30 $\pm$ 0.10 for insertion of *C*; and not significantly different than 0 for insertion of *G*.

[00223] In summary, the protospacer position-17 nucleotide accurately predicts which nucleotide is most frequently inserted at Cas9 targets and also explains substantial portions of the variance in the frequencies of insertion of A and insertion of T.

[00224] Although preferred embodiments of the subject methods have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the methods as defined by the appended claims.

Table 1
Genomic coordinates and sgRNA primer sequences for Spacers 1-96

<u>Spacer</u>	<u>Genomic location of spacer</u>	<u>sgRNA primer</u>	<u>SEQ ID NO.:</u>
1	chr1:836293-836316	TAATACGACTCACTATAGGCTGATGTAGTCACTCTTGAGTTTTAGAGCTAGAAATAGC	1
2	chr4:145041692-145041714	TAATACGACTCACTATAGAGTGTGCATTGCCACCTCAGGTTTTAGAGCTAGAAATAGC	2
3	chr15:90348654-90348677	TAATACGACTCACTATAGAGATGCAGGCTGCAGATGCCGTTTTAGAGCTAGAAATAGC	3
4	chr17:40440261-40440283	TAATACGACTCACTATAGGGCCCTCGGGCGGGCTCCCGTTTTAGAGCTAGAAATAGC	4
5	chrX:138644010-138644032	TAATACGACTCACTATAGACAGAACATGTTGTTATAGAGTTTTAGAGCTAGAAATAGC	5
6	chrX:66765045-66765067	TAATACGACTCACTATAGACAGATTCTGGAAGCTCCTGTTTTAGAGCTAGAAATAGC	6
7	chrX:138643957-138643979	TAATACGACTCACTATAGGGTCAACAAGTGGAAGCTCTAGTTTTAGAGCTAGAAATAGC	7
8	chr8:39871134-39871156	TAATACGACTCACTATAGGGAGTATTTTCAGGGGCTCTTGTTTTAGAGCTAGAAATAGC	8
9	chr1:183218409-183218431	TAATACGACTCACTATAGGGGAGGGCTGTGCTGCTAGTGTGTTTTAGAGCTAGAAATAGC	9
10	chr7:50526196-50526218	TAATACGACTCACTATAGGCAACTTGAAATTATATCTGGTTTTAGAGCTAGAAATAGC	10
11	chr8:39845460-39845482	TAATACGACTCACTATAGGTAGAGAAAAGAGCAGTGCCGTTTTAGAGCTAGAAATAGC	11
12	chr1:208072333-208072356	TAATACGACTCACTATAGTAGGAGAAAGATGATGTATAGTTTTAGAGCTAGAAATAGC	12
13	chr10:89506437-89506459	TAATACGACTCACTATAGGCATACAGTGATTTGATGAAGTTTTAGAGCTAGAAATAGC	13
14	chr7:55240709-55240731	TAATACGACTCACTATAGGGGGGCCCTCCTCTTGCTGCGTTTTAGAGCTAGAAATAGC	14
15	chr1:183220487-183220509	TAATACGACTCACTATAGGAAAATCATCAGTTATCATCGTTTTAGAGCTAGAAATAGC	15
16	chr1:208073317-208073340	TAATACGACTCACTATAGTTTCATGAGTCTTGACAACAAGTTTTAGAGCTAGAAATAGC	16
17	chr11:836360-836383	TAATACGACTCACTATAGGTGGCGGGCACTGTCTGCTCAGTTTTAGAGCTAGAAATAGC	17

18	chr22:42032717-42032740	TAATACGACTCACTATAGATGACAGTGCCAAAGCCAGCGTTTTAGAGCTAGAAATAGC	18
19	chr13:32900719-32900742	TAATACGACTCACTATAGACAGTAGAACTAAGGGTGGGGTTTTAGAGCTAGAAATAGC	19
20	chrX:66765028-66765050	TAATACGACTCACTATAGCCGCCGTCCAAGACCTACCGGTTTTAGAGCTAGAAATAGC	20
21	chr15:90348584-90348607	TAATACGACTCACTATAGCATCACGCTTATCCACCCCAAGTTTTAGAGCTAGAAATAGC	21
22	chr2:216977771-216977794	TAATACGACTCACTATAGACCATGAGTAACTCCATTCGTTTTAGAGCTAGAAATAGC	22
23	chrX:138644048-138644070	TAATACGACTCACTATAGGGAGGTAGAGATTGTCAGTTTTAGAGCTAGAAATAGC	23
	chr13:108863525-		
24	108863547	TAATACGACTCACTATAGGAACGAATACAGAAAAGTAAGTTTTAGAGCTAGAAATAGC	24
25	chr22:42032577-42032600	TAATACGACTCACTATAGATCCGTGGCCCATCATGTCTGTTTTAGAGCTAGAAATAGC	25
26	chr2:216974068-216974091	TAATACGACTCACTATAGGGTGGACAAAGCGGCAGATAGGTTTTAGAGCTAGAAATAGC	26
27	chr4:156826243-156826265	TAATACGACTCACTATAGGAACTGCAAAAGTGAAACAAAAGTTTTAGAGCTAGAAATAGC	27
28	chr1:198668757-198668780	TAATACGACTCACTATAGACTGACACGCAGACATTGAGGTTTTAGAGCTAGAAATAGC	28
29	chr22:37461708-37461730	TAATACGACTCACTATAGCGAAGCTGGAATCTGCTCTCGTTTTAGAGCTAGAAATAGC	29
30	chr10:6104130-6104152	TAATACGACTCACTATAGGACCAGCCGGGGCAGTGAAGGTTTTAGAGCTAGAAATAGC	30
31	chr15:90358013-90358036	TAATACGACTCACTATAGGGGCGCTGGAACCTGGACCCGTTTTAGAGCTAGAAATAGC	31
32	chr17:37883173-37883195	TAATACGACTCACTATAGGGACAGAAAGCCCTGCTGGTTTTAGAGCTAGAAATAGC	32
33	chr3:169482616-169482638	TAATACGACTCACTATAGCTGGGCAGGCGACCCCGCGTTTTAGAGCTAGAAATAGC	33
34	chr17:37884086-37884108	TAATACGACTCACTATAGTTCTCCACGGCACCCCAAGTTTTAGAGCTAGAAATAGC	34
35	chr7:55259538-55259560	TAATACGACTCACTATAGGAGAAAGAAATACCATGCAGAGTTTTAGAGCTAGAAATAGC	35
36	chrX:133594274-133594296	TAATACGACTCACTATAGCTCCTCTGCTCGCCACGTTTTAGAGCTAGAAATAGC	36
37	chrX:133594290-133594312	TAATACGACTCACTATAGGCTCAGGAGGAGGAAGCCGGTTTTAGAGCTAGAAATAGC	37

38	chr1:65432129-65432151	TAATACGACTCACTATAGGCTGAGGAGGGGTGCGCGGGTTTTAGAGCTAGAAATAGC	38
39	chr16:29708698-29708720	TAATACGACTCACTATAGGGGCACTGGAGCCACCTCTTGTTTTAGAGCTAGAAATAGC	39
40	chr19:17942176-17942198	TAATACGACTCACTATAGGCACGCAGCGGGGAGCCCGTTTTAGAGCTAGAAATAGC	40
41	chr19:17945473-17945495	TAATACGACTCACTATAGGGCCCCCAAGTGACAGAGCGTTTTAGAGCTAGAAATAGC	41
42	chr13:32893320-32893343	TAATACGACTCACTATAGTGTGGAGTTTTAAATAGGTTGTTTTAGAGCTAGAAATAGC	42
43	chr2:191878760-191878782	TAATACGACTCACTATAGGGAAGGGCTAGGCGGGGCGGTTTTAGAGCTAGAAATAGC	43
44	chr12:72425547-72425569	TAATACGACTCACTATAGGCAAAATAACCTTCTGTGTCAGTTTTAGAGCTAGAAATAGC	44
45	chr8:39840236-39840258	TAATACGACTCACTATAGGCACCAAGTCTGAGTGGACCGTTTTAGAGCTAGAAATAGC	45
46	chr12:72416242-72416264	TAATACGACTCACTATAGGCAGGACTCCTTTCCTCCATGTTTTAGAGCTAGAAATAGC	46
47	chr13:32889667-32889690	TAATACGACTCACTATAGGCTGCGCCTCTGCTGCGCCTGTTTTAGAGCTAGAAATAGC	47
48	chr13:32893350-32893373	TAATACGACTCACTATAGGAAGCCAGCTGATTATAAGAGTTTTAGAGCTAGAAATAGC	48
49	chr13:32889776-32889799	TAATACGACTCACTATAGCCAAAAAAGAACTGCACCTCGTTTTAGAGCTAGAAATAGC	49
50	chr15:90348325-90348348	TAATACGACTCACTATAGGTTGACTACGTGGAGAAAGCGTTTTAGAGCTAGAAATAGC	50
51	chr2:216974143-216974166	TAATACGACTCACTATAGGCACCATGTTGCCGGTCCCTCGTTTTAGAGCTAGAAATAGC	51
52	chr11:836347-836370	TAATACGACTCACTATAGGCCCGCCACCACCAGGATGTGTTTTAGAGCTAGAAATAGC	52
53	chr2:216981432-216981455	TAATACGACTCACTATAGGGCACTGACAATCCCCCTTTCGTTTTAGAGCTAGAAATAGC	53
54	chr1:65349086-65349108	TAATACGACTCACTATAGGAGGAGCTCCAAGAAGACTGGTTTTAGAGCTAGAAATAGC	54
55	chr4:156825204-156825226	TAATACGACTCACTATAGGAAGAAGACAAAATCACAAACGTTTTAGAGCTAGAAATAGC	55
56	chr1:208084481-208084504	TAATACGACTCACTATAGGGGGCGGGAAGAGCGCGTCCGTTTTAGAGCTAGAAATAGC	56
57	chr16:29708389-29708411	TAATACGACTCACTATAGGCCCGTGCAGGCAGCTGAGGCGTTTTAGAGCTAGAAATAGC	57

58	chr1:198608241-198608264	TAATACGACTCACTATAGCTCGTCTGATAAGACAACAGGTTTTAGAGCTAGAAATAGC	58
59	chr4:156825228-156825250	TAATACGACTCACTATAGGTGAATAGAGCCAGCAAGGGTTTTAGAGCTAGAAATAGC	59
60	chr1:198668733-198668756	TAATACGACTCACTATAGGACTCGCAGACGCCCTCTGCGTTTTAGAGCTAGAAATAGC	60
61	chr22:42017307-42017330	TAATACGACTCACTATAGTTGTCGTCCTTCTGTCCAAGTGTCTTTAGAGCTAGAAATAGC	61
62	chr13:32889654-32889677	TAATACGACTCACTATAGGAGGCCGACAGTGCACAGGTTTTAGAGCTAGAAATAGC	62
63	chr17:40441426-40441448	TAATACGACTCACTATAGGGCCATGGCGGGCTGGATCCGTTTTAGAGCTAGAAATAGC	63
64	chr22:37539598-37539620	TAATACGACTCACTATAGGAGCCAAAGATGGGCTCTGCGTTTTAGAGCTAGAAATAGC	64
65	chr6:43737291-43737313	TAATACGACTCACTATAGGGGTGGGGGAGTTTGCTCCGTTTTAGAGCTAGAAATAGC	65
66	chr19:17952223-17952245	TAATACGACTCACTATAGCCCGAGGCTGCTGGAGGAAGTTTTAGAGCTAGAAATAGC	66
67	chrX:66765119-66765141	TAATACGACTCACTATAGGCCCGGAGGTGCTGCGCTCGGTTTTAGAGCTAGAAATAGC	67
68	chr1:198668793-198668816	TAATACGACTCACTATAGGCAAAACTCAACCCTACCCGTTTTAGAGCTAGAAATAGC	68
69	chr10:6104107-6104129	TAATACGACTCACTATAGGAATCCATCTTCTGACCCCTGTTTTAGAGCTAGAAATAGC	69
70	chr1:198608161-198608184	TAATACGACTCACTATAGCTAGGTGATGATGTCAGATTGTTTTAGAGCTAGAAATAGC	70
71	chr17:40481444-40481466	TAATACGACTCACTATAGGAGCAGAGATGTGGGAATGGGTTTTAGAGCTAGAAATAGC	71
72	chr17:40466486-40466508	TAATACGACTCACTATAGGGGACTGGGTCGGGAGGGTGTTTTAGAGCTAGAAATAGC	72
73	chr7:50605568-50605590	TAATACGACTCACTATAGGAGAAAGCTGGAGAAGGGGGTTTTAGAGCTAGAAATAGC	73
74	chr22:42017321-42017344	TAATACGACTCACTATAGCGCGCAGGGAACCGACCAACTGTTTTAGAGCTAGAAATAGC	74
75	chr17:40466373-40466395	TAATACGACTCACTATAGGAAGAGGGGGAGAGAGTTACGTTTTAGAGCTAGAAATAGC	75
76	chr1:65311293-65311315	TAATACGACTCACTATAGGTGGAAGAGTTTGTGGAAGGGTTTTAGAGCTAGAAATAGC	76
77	chr15:40987440-40987463	TAATACGACTCACTATAGATTCCCAGCTGCACGCCTCGTTTTAGAGCTAGAAATAGC	77

78	chr7:506333109-506333131	TAATACGACTCACTATAGGAAAGAGGGAAACGTGTGGCTGTTTTAGAGCTAGAAATAGC	78
79	chr17:37863274-37863296	TAATACGACTCACTATAGGGTGGGTCTCGGACTGGCAGTTTTAGAGCTAGAAATAGC	79
80	chr2:191835123-191835145	TAATACGACTCACTATAGGAATGAGGGTCCTTTGGGAAGTTTTAGAGCTAGAAATAGC	80
81	chr2:191878928-191878950	TAATACGACTCACTATAGGGTGGGGCGGAAGGGGGCCGTTTTAGAGCTAGAAATAGC	81
82	chr15:40987403-40987426	TAATACGACTCACTATAGCGGCCAGAGACCCGAGCCCTAGTTTTAGAGCTAGAAATAGC	82
83	chr7:55220290-55220312	TAATACGACTCACTATAGTTGTGGCAGCAGTCACCTGGGTTTTAGAGCTAGAAATAGC	83
84	chr17:40458414-40458436	TAATACGACTCACTATAGGGCCACTGTAGTCTCCAGGGTTTTAGAGCTAGAAATAGC	84
85	chr22:37532281-37532303	TAATACGACTCACTATAGGGAGCCAGCCCCCTGGCCTTCGTTTTAGAGCTAGAAATAGC	85
86	chr1:208084623-208084646	TAATACGACTCACTATAGTCCCAAAGGGCGGAGGGCGTTGTTTTAGAGCTAGAAATAGC	86
87	chr2:73160981-73161003	TAATACGACTCACTATAGTGAGTCCGAGCAGAGAAGAGTTTTAGAGCTAGAAATAGC	87
88	chr2:73160982-73161004	TAATACGACTCACTATAGGAGTCCGAGCAGAGAAGAAAGTTTTAGAGCTAGAAATAGC	88
89	chr1:65312386-65312408	TAATACGACTCACTATAGGGCAGCCAGCATGATGAGACGTTTTAGAGCTAGAAATAGC	89
90	chr2:216977802-216977825	TAATACGACTCACTATAGCTTCTTTGCTTGTTCAAATGGTTTTAGAGCTAGAAATAGC	90
91	chr19:17953906-17953928	TAATACGACTCACTATAGGCAGGGTGAGTGTCTCAGCCGTTTTAGAGCTAGAAATAGC	91
92	chr22:37523901-37523923	TAATACGACTCACTATAGGAACTGTGTGTGTTGCAGGGGTTTTAGAGCTAGAAATAGC	92
93	chr1:55529322-55529344	TAATACGACTCACTATAGGGGCTGAGAGAGGGACAAAGTGTTTTAGAGCTAGAAATAGC	93
94	chr22:42032568-42032591	TAATACGACTCACTATAGAAAACGTTTCCAAGACATGAGTTTTAGAGCTAGAAATAGC	94
95	chr15:40987560-40987583	TAATACGACTCACTATAGCCACGCCCGCGGGGTGAAGTGTTTTAGAGCTAGAAATAGC	95
96	chrX:133594214-133594236	TAATACGACTCACTATAGGAGCCCTCAGGCCGAACCTCTGTTTTAGAGCTAGAAATAGC	96

Table 2
Sequencing library primers for Spacers 1-96 (PCR1)

5	Spacer	<u>ILMN primer sequence</u>	<u>SEQ ID NO.:</u>
	1	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGATGCGATGACCTTTGTG	97
	1	GGAGTTCAGACGTGTGCTCTTCCGATCTAGTCACCATGACGACAGTGC	98
	2	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGTCCCCTAAAATGGGT	99
	2	GGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTATGGTCCGCTCAGTC	100
	3	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTCACCTTTGGGAAGCATGT	101
	3	GGAGTTCAGACGTGTGCTCTTCCGATCTAGCTCTGGCACACCCCTCTAA	102
	4	CACTCTTTCCCTACACGACGCTCTTCCGATCTcagGTTTGGGATTTCCAGAG	103
	4	GGAGTTCAGACGTGTGCTCTTCCGATCTCCTGCAAGTGCACAACAG	104
	5	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCCTCAAATTTGGATCTGGC	105
	5	GGAGTTCAGACGTGTGCTCTTCCGATCTCCCCCACTATCTCCTTGACA	106
	6	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTCTGAGCAAGAGAAGGGA	107
	6	GGAGTTCAGACGTGTGCTCTTCCGATCTctgcAGCAGCAGCAAAC	108
	7	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCCTCAAATTTGGATCTGGC	105
	7	GGAGTTCAGACGTGTGCTCTTCCGATCTCCCCCACTATCTCCTTGACA	106
	8	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGCACCCCTTTTCATCTCT	109
	8	GGAGTTCAGACGTGTGCTCTTCCGATCTCACCTCTCCTTCTCCTCCCT	110

9 CACTCTTTCCCTACACGACGCTCTTCCGATCTGAGACTTTCCCCCTTGTTC 111
9 GGAGTTCAGACGTGTGCTCTTCCGATCTACAGGCAGAAAGGAAAAACCCCT 112
10 CACTCTTTCCCTACACGACGCTCTTCCGATCTGCGGTTTAAAAACATCCAA 113
10 GGAGTTCAGACGTGTGCTCTTCCGATCTAAGTGGTAGGAAAGCCTCACTG 114
11 CACTCTTTCCCTACACGACGCTCTTCCGATCTAAGGAACCTGGAGACCATCA 115
11 GGAGTTCAGACGTGTGCTCTTCCGATCTGAAAGGCACCTGAGTGGGAAG 116
12 CACTCTTTCCCTACACGACGCTCTTCCGATCTGGAAGAAACAGCTTGCCTG 117
12 GGAGTTCAGACGTGTGCTCTTCCGATCTGAAGCCTAGCCTGTCACTG 118
13 CACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGTGAAGACCTGACTGG 119
13 GGAGTTCAGACGTGTGCTCTTCCGATCTCAATGACCACAGCAAAAGAGC 120
14 CACTCTTTCCCTACACGACGCTCTTCCGATCTATCTGTCAAGAACCTCACCC 121
14 GGAGTTCAGACGTGTGCTCTTCCGATCTACTGGCACTCACCTCCCTC 122
15 CACTCTTTCCCTACACGACGCTCTTCCGATCTAACAAAGACCAAGGCACCTGCT 123
15 GGAGTTCAGACGTGTGCTCTTCCGATCTCTCAACCCCTGGAGGTCTTTG 124
16 CACTCTTTCCCTACACGACGCTCTTCCGATCTCCATGTTGAGACACAGGGTG 125
16 GGAGTTCAGACGTGTGCTCTTCCGATCTTCAGGAAATTGCATCAGGTG 126
17 CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGATCGATGACCTTTGTG 97
17 GGAGTTCAGACGTGTGCTCTTCCGATCTGTTCCGACGCTCCTTGAA 127
18 CACTCTTTCCCTACACGACGCTCTTCCGATCTTCAGTGAAGTGTGTGGTGC 128
18 GGAGTTCAGACGTGTGCTCTTCCGATCTTGCCCAATTTAAGAGAACGGG 129

19 CACTCTTTCCCTACACGACGCTCTTCCGATCTGGGTGTCAGACACCAAAAC 130
19 GGAGTTCAGACGTGTGCTCTTCCGATCTCAACCTCATCTGCTCTTTCTTG 131
20 CACTCTTTCCCTACACGACGCTCTTCCGATCTTTCTGAGCAAGAGAAGGGGA 107
20 GGAGTTCAGACGTGTGCTCTTCCGATCTCCGGGTTCTGGATCACTTC 132
21 CACTCTTTCCCTACACGACGCTCTTCCGATCTAGGGAGCATGTGTGTGAG 133
21 GGAGTTCAGACGTGTGCTCTTCCGATCTGGAAGTCCTTCCCATGCTTC 134
22 CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCCTTTTACACACCTTTCC 135
22 GGAGTTCAGACGTGTGCTCTTCCGATCTCTGCGCTGTACAAACATGG 136
23 CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCACAAAGGAGATCAGC 137
23 GGAGTTCAGACGTGTGCTCTTCCGATCTTTGCCCTTTCATTGCACACTC 138
24 CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCATAGGCCATTCTCTCTC 139
24 GGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCTCACAACCTTCACA 140
25 CACTCTTTCCCTACACGACGCTCTTCCGATCTAGGCAGCCACTGACATTCTT 141
25 GGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTTGTCTTCATTGGTGA 142
26 CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCGCATGCTCAGAGTTC 143
26 GGAGTTCAGACGTGTGCTCTTCCGATCTCCAAGTCCCATGGCTTTCTTT 144
27 CACTCTTTCCCTACACGACGCTCTTCCGATCTTGAGTTTGGAGGAGCATTTG 145
27 GGAGTTCAGACGTGTGCTCTTCCGATCTCAATGAGAAATGCCTGTGGA 146
28 CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGGGAAGACTGATGT 147
28 GGAGTTCAGACGTGTGCTCTTCCGATCTCCGCAAAACCTGAGATAGCAT 148

29	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCAGGACTTCCCCACCTT	149
29	GGAGTTCAGACGTGTGCTCTTCCGATCTTGGTTCTACATCCCGAGGAG	150
30	CACTCTTTCCCTACACGACGCTCTTCCGATCTGACTCCCTCTGTTCTGTGG	151
30	GGAGTTCAGACGTGTGCTCTTCCGATCTGATGCCAAAAAGAGGCTGAC	152
31	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGAAATCCAGCTCCGCAC	153
31	GGAGTTCAGACGTGTGCTCTTCCGATCTAAAAAGACGGGAAAGGAGGA	154
32	CACTCTTTCCCTACACGACGCTCTTCCGATCTGACAGCACTTCTACCGCTC	155
32	GGAGTTCAGACGTGTGCTCTTCCGATCTGGTAAGCAGACAGCCACACA	156
33	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCTGACAGAGCCCAACTCTT	157
33	GGAGTTCAGACGTGTGCTCTTCCGATCTGCCCTTCCACCGTTCAATCTA	158
34	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTGGTGCCACTCTGGAAAG	159
34	GGAGTTCAGACGTGTGCTCTTCCGATCTGTTCTCTGCGGTAGGTGTCC	160
35	CACTCTTTCCCTACACGACGCTCTTCCGATCTCCTCACAGCAGGGTCTTCTC	161
35	GGAGTTCAGACGTGTGCTCTTCCGATCTCCTGGTGTGTCAGGAAAAATGCT	162
36	CACTCTTTCCCTACACGACGCTCTTCCGATCTcCTGCTTCTCCTCAGCTTCA	163
36	GGAGTTCAGACGTGTGCTCTTCCGATCTGAGCTGCTCACCACGACG	164
37	CACTCTTTCCCTACACGACGCTCTTCCGATCTcCTGCTTCTCCTCAGCTTCA	163
37	GGAGTTCAGACGTGTGCTCTTCCGATCTGAGCTGCTCACCACGACG	164
38	CACTCTTTCCCTACACGACGCTCTTCCGATCTACACTCACCACCTCCGTGTG	165
38	GGAGTTCAGACGTGTGCTCTTCCGATCTGCGGAGTATCCTGGAGCTG	166

39 CACTCTTTCCCTACACGACGCTCTTCCGATCTATCACCCCTGGACAACCTCC 167
39 GGAGTTCAGACGTGTGCTCTTCCGATCTAAGATGTGTACCCAGAGGG 168
40 CACTCTTTCCCTACACGACGCTCTTCCGATCTGCTTAGCTAGGCCGAAAGTCA 169
40 GGAGTTCAGACGTGTGCTCTTCCGATCTGCTCGGGGTAGGGTTATAG 170
41 CACTCTTTCCCTACACGACGCTCTTCCGATCTGCACCTGAAGAGATGAGGCT 171
41 GGAGTTCAGACGTGTGCTCTTCCGATCTGGAGATTGGGGTGGGTCTAT 172
42 CACTCTTTCCCTACACGACGCTCTTCCGATCTTTTTCAGAAAGCTCCACCC 173
42 GGAGTTCAGACGTGTGCTCTTCCGATCTTCAGCCCTTGCTCTTTGAAT 174
43 CACTCTTTCCCTACACGACGCTCTTCCGATCTCCCGAGGACTCTGTCCCT 175
43 GGAGTTCAGACGTGTGCTCTTCCGATCTCTTTTCTCCTGCGGGTAGT 176
44 CACTCTTTCCCTACACGACGCTCTTCCGATCTTATCTGGGGATTTGATGCCT 177
44 GGAGTTCAGACGTGTGCTCTTCCGATCTGAGTGGTTATCTGCCATTGGA 178
45 CACTCTTTCCCTACACGACGCTCTTCCGATCTGAAAGGTCTGCCAAGGAAT 179
45 GGAGTTCAGACGTGTGCTCTTCCGATCTGTGTTTCTCGGGGAAAGTT 180
46 CACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGTGTTTGGGTGCCATTT 181
46 GGAGTTCAGACGTGTGCTCTTCCGATCTTTTCCCTGACCTTGAACCCAG 182
47 CACTCTTTCCCTACACGACGCTCTTCCGATCTAATTTCTTTTCGCCACACTG 183
47 GGAGTTCAGACGTGTGCTCTTCCGATCTGGTCACAAATCTGTCCCCCTC 184
48 CACTCTTTCCCTACACGACGCTCTTCCGATCTTTTCTTTCAGAAAGCTCCACCC 173
48 GGAGTTCAGACGTGTGCTCTTCCGATCTACAGGAGATTGGTACAGCGG 185

49	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTTTGAAACTAGCGGGCAG	186
49	GGAGTTCAGACGTTGTGCTCTTCCGATCTGACTGGGACTGCGGAAGAC	187
50	CACTCTTTCCCTACACGACGCTCTTCCGATCTccctggcctAACAAATTCAGA	188
50	GGAGTTCAGACGTTGTGCTCTTCCGATCTGACCCCAACTGGAATGTCAC	189
51	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTTCAGGCCTAGCAGGAAAC	190
51	GGAGTTCAGACGTTGTGCTCTTCCGATCTCCCATCTTTGTCTTGACCG	191
52	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGATGCGATGACCTTTGTG	97
52	GGAGTTCAGACGTTGTGCTCTTCCGATCTGTTCCGACGCTCCTTGAA	127
53	CACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGTCTGGTTGTCCTGCTC	192
53	GGAGTTCAGACGTTGTGCTCTTCCGATCTCAATGTCTCCAGCAAATCA	193
54	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGTGTACTCTCCACTGCCCA	194
54	GGAGTTCAGACGTTGTGCTCTTCCGATCTTCAGAACACTCCCTTTTGCC	195
55	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCAATGAGAAATTTAATCACCC	196
55	GGAGTTCAGACGTTGTGCTCTTCCGATCTTGCCTAACAAATGGACACCAA	197
56	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGCAAACCTCAGCAAGCAAA	198
56	GGAGTTCAGACGTTGTGCTCTTCCGATCTGGAGGGGAGAGAGAGGAAA	199
57	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTCTCTCACCCCTTGTCTC	200
57	GGAGTTCAGACGTTGTGCTCTTCCGATCTTAAGAGCCCCACACAGATCC	201
58	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCTGTCAAAGAGGACCCCTTA	202
58	GGAGTTCAGACGTTGTGCTCTTCCGATCTGCGTCTCTTCTGAAAAAGCAAA	203

59	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCAATGAGAAATTTTAATCACCCC	196
59	GGAGTTCAGACGCTGTGCTCTTCCGATCTTGCCTAACAAATGGACACCAA	197
60	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGGGAAGACTGATGT	147
60	GGAGTTCAGACGCTGTGCTCTTCCGATCTAGGGTTGAGTTTTGCATTGG	204
61	CACTCTTTCCCTACACGACGCTCTTCCGATCTTAGAAGCTGGTTGGGGAGTG	205
61	GGAGTTCAGACGCTGTGCTCTTCCGATCTCTAGCTGGCGAACAAACACAA	206
62	CACTCTTTCCCTACACGACGCTCTTCCGATCTAATTTCCCTTCGCCACACTG	183
62	GGAGTTCAGACGCTGTGCTCTTCCGATCTGGTCACAAATCTGTCCCCTC	184
63	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGATAGGTAGGCATGGCAAG	207
63	GGAGTTCAGACGCTGTGCTCTTCCGATCTTACCATGGCTGGCTCTCAAT	208
64	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTCCTTGGAGTCCAGTGCAT	209
64	GGAGTTCAGACGCTGTGCTCTTCCGATCTACTACACCTCATCTTGCCCC	210
65	CACTCTTTCCCTACACGACGCTCTTCCGATCTCCAGATCCAGATGGCACATTGTCAGA	211
65	GGAGTTCAGACGCTGTGCTCTTCCGATCTCCTAGTGACTGCCGTCTGC	212
66	CACTCTTTCCCTACACGACGCTCTTCCGATCTGAAAACCTTGACCCCTGTCCA	213
66	GGAGTTCAGACGCTGTGCTCTTCCGATCTCTCGTGGACGGCTACTTCC	214
67	CACTCTTTCCCTACACGACGCTCTTCCGATCTACCGAGGAGCTTTCCAGAAAT	215
67	GGAGTTCAGACGCTGTGCTCTTCCGATCTTGGGGAGAACCATCCTCAC	216
68	CACTCTTTCCCTACACGACGCTCTTCCGATCTGTACAGACGCCCTCACCTTCC	217
68	GGAGTTCAGACGCTGTGCTCTTCCGATCTGCTGCACATTGAATAAGTGGTT	218

69 CACTCTTTCCCTACACGACGCTCTTCCGATCTGACTCCCTCTGTTCTGTGG 151
69 GGAGTTCAGACGTGTGCTCTTCCGATCTGATGCCAAAAAGAGGCTGAC 152
70 CACTCTTTCCCTACACGACGCTCTTCCGATCTAACATCTTCCTTGATGGGAAAA 219
70 GGAGTTCAGACGTGTGCTCTTCCGATCTCAATTTCCCTCCTCTGTACCC 220
71 CACTCTTTCCCTACACGACGCTCTTCCGATCTTCATGTCACTTTGGCCTGAA 221
71 GGAGTTCAGACGTGTGCTCTTCCGATCTCTAGGGAGAGCCTCACAGGA 222
72 CACTCTTTCCCTACACGACGCTCTTCCGATCTCTCCCCCTCTTCTTCCAT 223
72 GGAGTTCAGACGTGTGCTCTTCCGATCTGTTGTTTCTGTGGTGCCTT 224
73 CACTCTTTCCCTACACGACGCTCTTCCGATCTAAGCATGAGTGCCTCTTTCC 225
73 GGAGTTCAGACGTGTGCTCTTCCGATCTTCGGTTAATCCCTTCCCTTC 226
74 CACTCTTTCCCTACACGACGCTCTTCCGATCTCATGCGTGATGACGTAGAGG 227
74 GGAGTTCAGACGTGTGCTCTTCCGATCTTCGCACACACTTAAGGCTAACG 228
75 CACTCTTTCCCTACACGACGCTCTTCCGATCTAAGACGCCATTACAAGTGCC 229
75 GGAGTTCAGACGTGTGCTCTTCCGATCTGCGTGCTCTAAAGGTCCCTCA 230
76 CACTCTTTCCCTACACGACGCTCTTCCGATCTCTCACCACCAAGTAGCTCAGGGC 231
76 GGAGTTCAGACGTGTGCTCTTCCGATCTGGACTGTCGTAAGGGGATGA 232
77 CACTCTTTCCCTACACGACGCTCTTCCGATCTGAAATCTGAAAAGCCGCTGG 233
77 GGAGTTCAGACGTGTGCTCTTCCGATCTCGCTCCACTTCTCTACTCGC 234
78 CACTCTTTCCCTACACGACGCTCTTCCGATCTCCAACCTTTGGGACTGAAGA 235
78 GGAGTTCAGACGTGTGCTCTTCCGATCTGCTTCCAGGATTTGGAATGA 236

79	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCTTCTCCCTGTCTGAGGTG	237
79	GGAGTTCAGACGTTGTGCTCTTCCGATCTGCAGGTAGGTGAGTTCCAGG	238
80	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCCTACGTCAAGCAGTTCCC	239
80	GGAGTTCAGACGTTGTGCTCTTCCGATCTGGCATTCTCTGAAGAGTGGG	240
81	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTGCTCGGTCTGGGGTCT	241
81	GGAGTTCAGACGTTGTGCTCTTCCGATCTGAAGCCGGCGGAAATACC	242
82	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCTGTAAACTCGCGCAGG	243
82	GGAGTTCAGACGTTGTGCTCTTCCGATCTCAGATGAGTTGCAGTTCCCA	244
83	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCACAGGGAACCTTTGCTCT	245
83	GGAGTTCAGACGTTGTGCTCTTCCGATCTTACCAGGCAGTCGCTCTC	246
84	CACTCTTTCCCTACACGACGCTCTTCCGATCTACATGAAATTCAAGGCCGAA	247
84	GGAGTTCAGACGTTGTGCTCTTCCGATCTACCTGTCTGTGAGGTGGAGG	248
85	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGGGATGGAGCTGACTGCTA	249
85	GGAGTTCAGACGTTGTGCTCTTCCGATCTACCCCGACACACCCAGTATGA	250
86	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTTCTCTTCTCCCTCC	251
86	GGAGTTCAGACGTTGTGCTCTTCCGATCTCTGCCACAAAGGGTTAAAA	252
87	CACTCTTTCCCTACACGACGCTCTTCCGATCTGTTTCTCATCTGTGCCCTC	253
87	GGAGTTCAGACGTTGTGCTCTTCCGATCTGTTGCCACCCCTAGTCATTG	254
88	CACTCTTTCCCTACACGACGCTCTTCCGATCTGTTTCTCATCTGTGCCCTC	253
88	GGAGTTCAGACGTTGTGCTCTTCCGATCTGTTGCCACCCCTAGTCATTG	254

89	CACTCTTTCCCTACACGACGCTCTTCCGATCTaagaaaggCAAGAAAGCCTGG	255
89	GGAGTTCAGACG GTGCTCTTCCGATCTGCTGGCCTGAGACATTCCCTA	256
90	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTGGAACCTTTGTTCCAGGC	257
90	GGAGTTCAGACG GTGCTCTTCCGATCTGGCAACAAGCAGTTCAAACA	258
91	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGGCTCTCACCTGACAGTCTT	259
91	GGAGTTCAGACG GTGCTCTTCCGATCTACAACAGGGCTTGAAGTTGG	260
92	CACTCTTTCCCTACACGACGCTCTTCCGATCTagaggagcgatccttgAG	261
92	GGAGTTCAGACG GTGCTCTTCCGATCTACTTGGTCCATCCATTTCCA	262
93	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGAGCTCCAGTGACAGC	263
93	GGAGTTCAGACG GTGCTCTTCCGATCTGGCACCCAGAGTGAGTGAGT	264
94	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGGCAGCCACTGACATTCTT	141
94	GGAGTTCAGACG GTGCTCTTCCGATCTGGGGTTGCTTCA TTGGTGA	142
95	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGAGACCGAGCCCTAAGGAG	265
95	GGAGTTCAGACG GTGCTCTTCCGATCTCTCACACACTCACCTCGGTC	266
96	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGAAAATTCCCACGGCTACC	267
96	GGAGTTCAGACG GTGCTCTTCCGATCTGACTGCTCAGGAGGAGGAAG	268

Table 3
Multiple Target Single Spacer (MTSS) genomic coordinates and sgRNA primer sequences

5	Spacer	Genomic location (hg38)	sgRNA primer	SEQ ID NO.:
1a		chr1:58056377-58056400	TAATACGACTCACTATAGCCTTGTCCTTGGGCACGCATGTTT	269
1b		chr2:171375175-171375198	TAATACGACTCACTATAGCCTTGTCCTTGGGCACGCATGTTT	270
1c		chr4:113214290-113214313	TAATACGACTCACTATAGCCTTGTCCTTGGGCACGCATGTTT	270
2a		chr5:41951514-41951537	TAATACGACTCACTATAGGTGACCAAGGAGGAATTCAGTTT	271
2b		chr10:14723370-14723393	TAATACGACTCACTATAGGTGACCAAGGAGGAATTCAGTTT	271
3a		chr12:124927891-124927914	TAATACGACTCACTATAGctctcccaactgagctatttGTTTT	272
3b		chr12:19613825-19613848	TAATACGACTCACTATAGctctcccaactgagctatttGTTTT	272
3c		chr19:1383410-1383433	TAATACGACTCACTATAGctctcccaactgagctatttGTTTT	272
3d		chr13:94549698-94549721	TAATACGACTCACTATAGctctcccaactgagctatttGTTTT	272
3e		chr6:28790770-28790793	TAATACGACTCACTATAGctctcccaactgagctatttGTTTT	272
3f		chr6:28823365-28823388	TAATACGACTCACTATAGctctcccaactgagctatttGTTTT	272
3g		chr6:28981720-28981743	TAATACGACTCACTATAGctctcccaactgagctatttGTTTT	272
3h		chr11:59566428-59566451	TAATACGACTCACTATAGctctcccaactgagctatttGTTTT	272
4a		chr12:124927869-124927892	TAATACGACTCACTATAGcggttagactgaagatctaaGTTTT	273
4b		chr19:1383388-1383411	TAATACGACTCACTATAGcggttagactgaagatctaaGTTTT	273
4c		chr13:94549676-94549699	TAATACGACTCACTATAGcggttagactgaagatctaaGTTTT	273
4d		chr6:28764409-28764432	TAATACGACTCACTATAGcggttagactgaagatctaaGTTTT	273
5a		chr1:229431555-229431578	TAATACGACTCACTATAGGATCCACATCTGCTGGAAGGGTTT	274
5b		chr6:101430419-101430442	TAATACGACTCACTATAGGATCCACATCTGCTGGAAGGGTTT	274

6a	chr9:92006517-92006540	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6b	chr12:122365605-122365628	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6c	chr12:12155745-12155768	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6d	chr2:171587547-171587570	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6e	chr13:27256498-27256521	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6f	chr4:19814204-19814227	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6g	chr17:3665675-3665698	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6h	chr17:43079753-43079776	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6i	chr7:20003220-20003243	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6j	chr5:177265214-177265237	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6k	chr10:67794223-67794246	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6l	chr3:154024271-154024294	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6m	chr3:147277225-147277248	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
6n	chr3:32030651-32030674	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	275
7a	chr1:39963383-39963406	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	276
7b	chr6:43364273-43364296	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	276
7c	chr10:106455788-106455811	TAATACGACTCACTATAGGCCATGAATTCATAGGGAATGTTTTAGAGCTAGAAATAGC	276
8a	chr15:64593305-64593328	TAATACGACTCACTATAGCCCCCACCACCAAGCCCCATGTAGTTTTAGAGCTAGAAATAGC	277
8b	chr12:56044133-56044156	TAATACGACTCACTATAGCCCCCACCACCAAGCCCCATGTAGTTTTAGAGCTAGAAATAGC	277
8c	chr12:3469275-3469298	TAATACGACTCACTATAGCCCCCACCACCAAGCCCCATGTAGTTTTAGAGCTAGAAATAGC	277
8d	chr1:51980799-51980822	TAATACGACTCACTATAGCCCCCACCACCAAGCCCCATGTAGTTTTAGAGCTAGAAATAGC	277
8e	chrX:72044871-72044894	TAATACGACTCACTATAGCCCCCACCACCAAGCCCCATGTAGTTTTAGAGCTAGAAATAGC	277
8f	chr5:75374788-75374811	TAATACGACTCACTATAGCCCCCACCACCAAGCCCCATGTAGTTTTAGAGCTAGAAATAGC	277
8g	chr18:59761884-59761907	TAATACGACTCACTATAGCCCCCACCACCAAGCCCCATGTAGTTTTAGAGCTAGAAATAGC	277
9a	chr12:118246943-118246966	TAATACGACTCACTATAGTCCAGCTGTGGCTACAACATGTTTTAGAGCTAGAAATAGC	278
9b	chr19:12285163-12285186	TAATACGACTCACTATAGTCCAGCTGTGGCTACAACATGTTTTAGAGCTAGAAATAGC	278

9c	chr19:16413554-16413577	TAATACGACTCACTATAGTCCAGCTGTGGCTACAACATGTTTTAGAGCTAGAAATAGC	278
9d	chr1:198949783-198949806	TAATACGACTCACTATAGTCCAGCTGTGGCTACAACATGTTTTAGAGCTAGAAATAGC	278
9e	chr1:399633382-399633405	TAATACGACTCACTATAGTCCAGCTGTGGCTACAACATGTTTTAGAGCTAGAAATAGC	278
9f	chr6:43364272-43364295	TAATACGACTCACTATAGTCCAGCTGTGGCTACAACATGTTTTAGAGCTAGAAATAGC	278
9g	chr10:106455787-106455810	TAATACGACTCACTATAGTCCAGCTGTGGCTACAACATGTTTTAGAGCTAGAAATAGC	278
10a	chr22:19258809-19258832	TAATACGACTCACTATAGAGCAGAAAGCAGGTACCCCTGTTTTAGAGCTAGAAATAGC	279
10b	chr12:52952845-52952868	TAATACGACTCACTATAGAGCAGAAAGCAGGTACCCCTGTTTTAGAGCTAGAAATAGC	279
10c	chr6:34191085-34191108	TAATACGACTCACTATAGAGCAGAAAGCAGGTACCCCTGTTTTAGAGCTAGAAATAGC	279
11a	chr9:106978769-106978792	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
11b	chr9:133351349-133351372	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
11c	chr14:69885333-69885356	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
11d	chr1:147224336-147224359	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
11e	chr2:158549207-158549230	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
11f	chr16:22379779-22379802	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
11g	chr8:39860514-39860537	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
11h	chr10:63902437-63902460	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
11i	chr10:33057602-33057625	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
11j	chr3:121494891-121494914	TAATACGACTCACTATAGGTACATTTAACCACAGTTTAGGTTTTAGAGCTAGAAATAGC	280
12a	chr15:99285519-99285542	TAATACGACTCACTATAGCTGAGGAAGCTCTTCATTGGGTTTTAGAGCTAGAAATAGC	281
12b	chr1:115857760-115857783	TAATACGACTCACTATAGCTGAGGAAGCTCTTCATTGGGTTTTAGAGCTAGAAATAGC	281
13a	chr19:21233591-21233614	TAATACGACTCACTATAGAAAGATGCAAGCATTTTGAACGTTTTAGAGCTAGAAATAGC	282
13b	chr4:112171466-112171489	TAATACGACTCACTATAGAAAGATGCAAGCATTTTGAACGTTTTAGAGCTAGAAATAGC	282
13c	chr17:67224980-67225003	TAATACGACTCACTATAGAAAGATGCAAGCATTTTGAACGTTTTAGAGCTAGAAATAGC	282
13d	chr8:95403830-95403853	TAATACGACTCACTATAGAAAGATGCAAGCATTTTGAACGTTTTAGAGCTAGAAATAGC	282
13e	chrX:101395414-101395437	TAATACGACTCACTATAGAAAGATGCAAGCATTTTGAACGTTTTAGAGCTAGAAATAGC	282
14a	chr15:43989501-43989524	TAATACGACTCACTATAGCCATCTCCTGCTCGAAGTCCGTTTTAGAGCTAGAAATAGC	283

14b	chr17:81511307-81511330	TAATACGACTCACTATAGCCATCTCCTGCTCGAAGTCCGTTTTAGAGCTAGAAATAGC	283
14c	chrX:53143241-53143264	TAATACGACTCACTATAGCCATCTCCTGCTCGAAGTCCGTTTTAGAGCTAGAAATAGC	283
15a	chr12:30862020-30862043	TAATACGACTCACTATAGGGCAGTGCAGATGAAAAACTGTTTTAGAGCTAGAAATAGC	284
15b	chr2:129496621-129496644	TAATACGACTCACTATAGGGCAGTGCAGATGAAAAACTGTTTTAGAGCTAGAAATAGC	284
15c	chr2:68125583-68125606	TAATACGACTCACTATAGGGCAGTGCAGATGAAAAACTGTTTTAGAGCTAGAAATAGC	284
15d	chr20:53421489-53421512	TAATACGACTCACTATAGGGCAGTGCAGATGAAAAACTGTTTTAGAGCTAGAAATAGC	284
15e	chr7:76361951-76361974	TAATACGACTCACTATAGGGCAGTGCAGATGAAAAACTGTTTTAGAGCTAGAAATAGC	284
15f	chr7:44799840-44799863	TAATACGACTCACTATAGGGCAGTGCAGATGAAAAACTGTTTTAGAGCTAGAAATAGC	284
15g	chr11:43466710-43466733	TAATACGACTCACTATAGGGCAGTGCAGATGAAAAACTGTTTTAGAGCTAGAAATAGC	284
16a	chr14:76634801-76634824	TAATACGACTCACTATAGCCATTTCAGTGGCCTGAGCAGGTTTTTAGAGCTAGAAATAGC	285
16b	chr12:65758144-65758167	TAATACGACTCACTATAGCCATTTCAGTGGCCTGAGCAGGTTTTTAGAGCTAGAAATAGC	285
16c	chr1:154378226-154378249	TAATACGACTCACTATAGCCATTTCAGTGGCCTGAGCAGGTTTTTAGAGCTAGAAATAGC	285
16d	chr1:101786362-101786385	TAATACGACTCACTATAGCCATTTCAGTGGCCTGAGCAGGTTTTTAGAGCTAGAAATAGC	285
16e	chr1:160266365-160266388	TAATACGACTCACTATAGCCATTTCAGTGGCCTGAGCAGGTTTTTAGAGCTAGAAATAGC	285
16f	chr13:67266858-67266881	TAATACGACTCACTATAGCCATTTCAGTGGCCTGAGCAGGTTTTTAGAGCTAGAAATAGC	285
16g	chrX:87703369-87703392	TAATACGACTCACTATAGCCATTTCAGTGGCCTGAGCAGGTTTTTAGAGCTAGAAATAGC	285
16h	chr10:14723194-14723217	TAATACGACTCACTATAGCCATTTCAGTGGCCTGAGCAGGTTTTTAGAGCTAGAAATAGC	285
16i	chr3:110682312-110682335	TAATACGACTCACTATAGCCATTTCAGTGGCCTGAGCAGGTTTTTAGAGCTAGAAATAGC	285
17a	chr12:65758139-65758162	TAATACGACTCACTATAGGCTCAGGCCACTGAATGGTGTGTTTTAGAGCTAGAAATAGC	286
17b	chr1:101786357-101786380	TAATACGACTCACTATAGGCTCAGGCCACTGAATGGTGTGTTTTAGAGCTAGAAATAGC	286
17c	chr1:154378221-154378244	TAATACGACTCACTATAGGCTCAGGCCACTGAATGGTGTGTTTTAGAGCTAGAAATAGC	286
17d	chr1:160266360-160266383	TAATACGACTCACTATAGGCTCAGGCCACTGAATGGTGTGTTTTAGAGCTAGAAATAGC	286
17e	chrX:87703364-87703387	TAATACGACTCACTATAGGCTCAGGCCACTGAATGGTGTGTTTTAGAGCTAGAAATAGC	286
17f	chr3:110682307-110682330	TAATACGACTCACTATAGGCTCAGGCCACTGAATGGTGTGTTTTAGAGCTAGAAATAGC	286
18a	chr22:40106287-40106310	TAATACGACTCACTATAGCCAGCTTAGAAAAATAATCAGTTTTTAGAGCTAGAAATAGC	287
18b	chr1:81098229-81098252	TAATACGACTCACTATAGCCAGCTTAGAAAAATAATCAGTTTTTAGAGCTAGAAATAGC	287

18c	chr13:76981140-76981163	TAATACGACTCACTATAGCCAGCTTAGAAAAAATATCAGTTTTAGAGCTAGAAATAGC	287
18d	chr6:17530862-17530885	TAATACGACTCACTATAGCCAGCTTAGAAAAAATATCAGTTTTAGAGCTAGAAATAGC	287
19a	chr2:36299490-36299513	TAATACGACTCACTATAGAGCCAAAGAGAAAGGTACCTGTTTTAGAGCTAGAAATAGC	288
19b	chr13:110762776-110762799	TAATACGACTCACTATAGAGCCAAAGAGAAAGGTACCTGTTTTAGAGCTAGAAATAGC	288
19c	chr17:29716286-29716309	TAATACGACTCACTATAGAGCCAAAGAGAAAGGTACCTGTTTTAGAGCTAGAAATAGC	288
20a	chr15:78863705-78863728	TAATACGACTCACTATAGTGACGAACACAAAGGGAAAGGTTTTAGAGCTAGAAATAGC	289
20b	chr14:55414035-55414058	TAATACGACTCACTATAGTGACGAACACAAAGGGAAAGGTTTTAGAGCTAGAAATAGC	289
20c	chr14:90367155-90367178	TAATACGACTCACTATAGTGACGAACACAAAGGGAAAGGTTTTAGAGCTAGAAATAGC	289
20d	chr14:65267666-65267689	TAATACGACTCACTATAGTGACGAACACAAAGGGAAAGGTTTTAGAGCTAGAAATAGC	289
20e	chr6:159526504-159526527	TAATACGACTCACTATAGTGACGAACACAAAGGGAAAGGTTTTAGAGCTAGAAATAGC	289
20f	chr4:127813176-127813199	TAATACGACTCACTATAGTGACGAACACAAAGGGAAAGGTTTTAGAGCTAGAAATAGC	289
20g	chr4:13632574-13632597	TAATACGACTCACTATAGTGACGAACACAAAGGGAAAGGTTTTAGAGCTAGAAATAGC	289
20h	chr10:120355159-120355182	TAATACGACTCACTATAGTGACGAACACAAAGGGAAAGGTTTTAGAGCTAGAAATAGC	289
20i	chr18:21119637-21119660	TAATACGACTCACTATAGTGACGAACACAAAGGGAAAGGTTTTAGAGCTAGAAATAGC	289
21a	chr1:56619406-56619429	TAATACGACTCACTATAGCCCTATGTTGTAGCCACAGCGTTTTAGAGCTAGAAATAGC	290
21b	chr16:1962095-1962118	TAATACGACTCACTATAGCCCTATGTTGTAGCCACAGCGTTTTAGAGCTAGAAATAGC	290
21c	chr17:47159188-47159211	TAATACGACTCACTATAGCCCTATGTTGTAGCCACAGCGTTTTAGAGCTAGAAATAGC	290
21d	chr17:53756986-53757009	TAATACGACTCACTATAGCCCTATGTTGTAGCCACAGCGTTTTAGAGCTAGAAATAGC	290
21e	chrX:40934979-40935002	TAATACGACTCACTATAGCCCTATGTTGTAGCCACAGCGTTTTAGAGCTAGAAATAGC	290
21f	chr20:47160835-47160858	TAATACGACTCACTATAGCCCTATGTTGTAGCCACAGCGTTTTAGAGCTAGAAATAGC	290
22a	chr14:70595829-70595852	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291
22b	chr22:19258807-19258830	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291
22c	chr12:52952843-52952866	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291
22d	chr1:111648262-111648285	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291
22e	chr16:72729906-72729929	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291
22f	chr4:144572204-144572227	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291

22g	chr4:135371152-135371175	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291
22h	chr20:49958033-49958056	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291
22i	chr3:12787365-12787388	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291
22j	chr3:32260231-32260254	TAATACGACTCACTATAGAAAGGGTACCCCTGCTTCTGCGTTTTAGAGCTAGAAATAGC	291

Table 4A
Multiple Target Single Spacer (MTSS) sequencing library ILMN_P5 primers (PCR1)

Spacer group*	ILMN_P5	SEQ ID NO.:
1a	CACTCTTTCCCTACACGACGCTCTTCCGATCTACATAGGCATCGAAGACGCT	292
1b	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGGCATTGAAGACGCTCACT	293
1c	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGGGAAGCACATAGGCAT	294
2a	CACTCTTTCCCTACACGACGCTCTTCCGATCTACAAAGCACACATGCAACCT	295
2b	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCTGCACACCTTCAGACCCAG	296
3a	CACTCTTTCCCTACACGACGCTCTTCCGATCTAATCGGCTTCGTCTATGCAC	297
3b	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTGGAAATGACAGATTGGGA	298
3c	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGACCAATCCTGAACGAAAG	299
3d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGTATCGTGGACGGAGAGTC	300
3e	CACTCTTTCCCTACACGACGCTCTTCCGATCTcaaaaTAAGGGTTCTATTAGGCAAA	301
3f	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGATTGGTGTGCAAGTGTG	302
3g	CACTCTTTCCCTACACGACGCTCTTCCGATCTATAATTTGATTCGGTGGGC	303
3h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTAACGCCCTAAAAACGGAAAGC	304
4a	CACTCTTTCCCTACACGACGCTCTTCCGATCTAATCGGCTTCGTCTATGCAC	297
4b	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGACCAATCCTGAACGAAAG	299
4c	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGTATCGTGGACGGAGAGTC	300

4d	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGGCATTAGGAATACGCC	305
5a	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCTGGAGTGGAGTGTGTCT	306
5b	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCAGATACAAAGCTTCTGGGACA	307
6a	CACTCTTTCCCTACACGACGCTCTTCCGATCTaagagaaagGTGCCTGGGTT	308
6b	CACTCTTTCCCTACACGACGCTCTTCCGATCTAAGAAAGCCAAAAGAGCAAGGG	309
6c	CACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCAACTGAAGCGCCA	310
6d	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGAAAAGCTACCCGGGTTCA	311
6e	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTAATAGCCTGCTCCACCCA	312
6f	CACTCTTTCCCTACACGACGCTCTTCCGATCTACCTGGGTTCAACTAAAGCG	313
6g	CACTCTTTCCCTACACGACGCTCTTCCGATCTACTAAAGCACCAAGCCTGCTC	314
6h	CACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCAACTGAAGCACCAAGCC	315
6i	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAACTAAAGCGCCAGCCTAC	316
6j	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCACTAGCTTGTCTCCACTCA	317
6k	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCCAATGAGAAAGGTTCCCTG	318
6l	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGTACTTGGTGTGAGCCAGC	319
6m	CACTCTTTCCCTACACGACGCTCTTCCGATCTAACTGAAGCTCCAGCCTGC	320
6n	CACTCTTTCCCTACACGACGCTCTTCCGATCTGACAAACGTACCTGGGTTCAA	321
7a	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGACCTCTGGAAGGAGACTG	322
7b	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGACCTCTGGAAGGAGACAG	323
7c	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGTGGCTACAACATAGGGG	324

8a	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGGAATCGATCTCGTGAAGC	325
8b	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGTGAGAGGATGGTGTCAA	326
8c	CACTCTTTCCCTACACGACGCTCTTCCGATCTGAGTGTGCAATTCACAGCAAA	327
8d	CACTCTTTCCCTACACGACGCTCTTCCGATCTGAATCGATCTCATGAAGCCC	328
8e	CACTCTTTCCCTACACGACGCTCTTCCGATCTAATCGATCTCGTGAAGCCTG	329
8f	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGAATCGATCTCATGAAGG	330
8g	CACTCTTTCCCTACACGACGCTCTTCCGATCTATCAATCTTGTGTAGCCCCG	331
9a	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGACCTCTGGAAGGAGACTG	332
9b	CACTCTTTCCCTACACGACGCTCTTCCGATCTTAACCCCTGACCTCTGGAAG	332
9c	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTGACTCCCAACCTCTTGAA	333
9d	CACTCTTTCCCTACACGACGCTCTTCCGATCTGACCTACAGCTACCTGACCCC	334
9e	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGACCTCTGGAAGGAGACTG	322
9f	CACTCTTTCCCTACACGACGCTCTTCCGATCTCGACCTCTGGAAGGAGACAG	323
9g	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGTGGCTACAACATAGGGG	324
10a	CACTCTTTCCCTACACGACGCTCTTCCGATCTCCCTAGACAGCAACTCC	335
10b	CACTCTTTCCCTACACGACGCTCTTCCGATCTAACCATCCAAAAGACCACCA	336
10c	CACTCTTTCCCTACACGACGCTCTTCCGATCTCCGCCAGGATAGTGGATG	337
11a	CACTCTTTCCCTACACGACGCTCTTCCGATCTAAGCCCTGAACCTCTCTTTCAA	338
11b	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGTCTTTTCAGATCCGCC	339
11c	CACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTCAAAACCCAGTTCCAA	340

11d	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCAATGTCTGGGTCCTAAA	341
11e	CACTCTTTCCCTACACGACGCTCTTCCGATCTATGATGAGATCGGCCATCAC	342
11f	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTGTAAAGTGGTGATTTTCAGTTTGA	343
11g	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGTGAAAGCTATCAGGACCA	344
11h	CACTCTTTCCCTACACGACGCTCTTCCGATCTGTGACAAAAGGTGACCTGGG	345
11i	CACTCTTTCCCTACACGACGCTCTTCCGATCTTcttcatcccttacttctTTT	346
11j	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGGCTAAGCTGGTGAAAGTT	347
12a	CACTCTTTCCCTACACGACGCTCTTCCGATCTTAGTGACAAACCCAAAGCCC	348
12b	CACTCTTTCCCTACACGACGCTCTTCCGATCTCATTAACTCAGCCCCGTGA	349
13a	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGCAGAATGGCTATGATGGG	350
13b	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGGACAAGCTAAGCCAAATTTT	351
13c	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCGGAAGAAGGCTAAAACT	352
13d	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGCATGCATTACACACAAT	353
13e	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGCAATGCTTTGCTTTAAATA	354
14a	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGGTAGTTTTCATGGATGCCG	355
14b	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCTCACAAACCTACCCAGG	356
14c	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAAGAAGGAAGGCTGGAACA	357
15a	CACTCTTTCCCTACACGACGCTCTTCCGATCTATCCTGGAATTCGGTGAGG	358
15b	CACTCTTTCCCTACACGACGCTCTTCCGATCTCACCATAACAGCACTGGTGG	359
15c	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTGGTCCATCTATGGGGAGA	360

15d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTGGTAAAGCCCCATCT	361
15e	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGGGAGAAATTCGATGAAG	362
15f	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGTGACTTCACACGCCATAA	363
15g	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCGGAGTTTATATGCCAGG	364
16a	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGTCTCTACAAGGTCACATGCTT	365
16b	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGGCCATGGACTTAGGATGACT	366
16c	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGCCAAAGGCTGTTCACTAA	367
16d	CACTCTTTCCCTACACGACGCTCTTCCGATCTTcTGCACATGTATCCCGTG	368
16e	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCACGTAATGATGACTTCCAA	369
16f	CACTCTTTCCCTACACGACGCTCTTCCGATCTTgtaaaggctcaacatTTCTTT	370
16g	CACTCTTTCCCTACACGACGCTCTTCCGATCTAAACAATGCTAAATGATGGCAA	371
16h	CACTCTTTCCCTACACGACGCTCTTCCGATCTCCATCAGACTTGTGTCCACG	372
16i	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGGCTTATGCCCCAAGACTTT	373
17a	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGCCATGGACTTAGGATGACT	366
17b	CACTCTTTCCCTACACGACGCTCTTCCGATCTTcTGCACATGTATCCCGTG	368
17c	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGCCAAAGGCTGTTCACTAA	367
17d	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCACGTAATGATGACTTCCAA	369
17e	CACTCTTTCCCTACACGACGCTCTTCCGATCTAAACAATGCTAAATGATGGCAA	371
17f	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGGCTTATGCCCCAAGACTTT	373
18a	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTTTTATGTTGCCCTGCTTT	374

18b	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGGCATACACTCCAATGAA	375
18c	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGATGCTCCCTGAGTTTCTTC	376
18d	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCACCTGTTCCAAGAGACTG	377
19a	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCAATGACTCCAGAGGGAAG	378
19b	CACTCTTTCCCTACACGACGCTCTTCCGATCTATTACAGGGAAGGTCCC	379
19c	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCTCCTCAGCCTGGAAACAT	380
20a	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGTAACACTTGTGGGCATT	381
20b	CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGCCATTCCCTTGATGTCT	382
20c	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGTAACACTTGTGGGCATT	381
20d	CACTCTTTCCCTACACGACGCTCTTCCGATCTcgccGATATCAACTTTCCTT	383
20e	CACTCTTTCCCTACACGACGCTCTTCCGATCTGCATATACGTGGCCAAAGGA	384
20f	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGAACAGTACCCGTTCCC	385
20g	CACTCTTTCCCTACACGACGCTCTTCCGATCTaagcctggccTCACCTTT	386
20h	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGTAACACTTGTGGGCATT	381
20i	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTCTTCCAGTTTGGCCAAGG	387
21a	CACTCTTTCCCTACACGACGCTCTTCCGATCTAAAGTGCTCCCGTTCTGCTA	388
21b	CACTCTTTCCCTACACGACGCTCTTCCGATCTCACTAACCATGCAGGACACG	389
21c	CACTCTTTCCCTACACGACGCTCTTCCGATCTaccacgttgccTTAATTT	390
21d	CACTCTTTCCCTACACGACGCTCTTCCGATCTCTTCTGCATTTTTCACATTAGCAA	391
21e	CACTCTTTCCCTACACGACGCTCTTCCGATCTccacacctggcTCAGAGG	392

21f	CACTCTTTCCCTACACGACGCTCTTCCGATCTgatttcaccttgaagcctcC	393
22a	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGAAGACAGCCAGGACTTCA	394
22b	CACTCTTTCCCTACACGACGCTCTTCCGATCTCCCTAGACAGCAGCAACTCC	335
22c	CACTCTTTCCCTACACGACGCTCTTCCGATCTAACCATCCAAAAGACCACCA	336
22d	CACTCTTTCCCTACACGACGCTCTTCCGATCTgcctcccagggttcaaacA	395
22e	CACTCTTTCCCTACACGACGCTCTTCCGATCTTCCATGCAAAACCAATTCAAA	396
22f	CACTCTTTCCCTACACGACGCTCTTCCGATCTGAAATTACCTATTAACAGATGCTGACA	397
22g	CACTCTTTCCCTACACGACGCTCTTCCGATCTcaagccagggttaagttacacaG	398
22h	CACTCTTTCCCTACACGACGCTCTTCCGATCTAACCATCCAAAAGACCACCA	336
22i	CACTCTTTCCCTACACGACGCTCTTCCGATCTcaggagaccataTGTTTATTTATTGATT	399
22j	CACTCTTTCCCTACACGACGCTCTTCCGATCTATCCAAAAGACCACCACTG	400

*Spacer group subsets indicated by lower case letters; genomic location of spacer group subsets are provided in Table 3.

Table 4B
Multiple Target Single Spacer (MTSS) sequencing library ILMN_P7 primers (PCR1)

Spacer group*	ILMN_P7	SEQ ID NO.:
1a	GGAGTTCAGACGTGTGCTCTTCCGATCTTGCACATGAGCTCTCTCTGG	401
1b	GGAGTTCAGACGTGTGCTCTTCCGATCTGCTAACGAGGAACTTGGCAG	402
1c	GGAGTTCAGACGTGTGCTCTTCCGATCTCAGCATTTTGGCTCCTCTCT	403
2a	GGAGTTCAGACGTGTGCTCTTCCGATCTGGATGTGGCCCAAGAAGTTA	404
2b	GGAGTTCAGACGTGTGCTCTTCCGATCTTCCCTGTAAACACCCCGTGACA	405
3a	GGAGTTCAGACGTGTGCTCTTCCGATCTGCAAAACGACCAGGAGGAT	406
3b	GGAGTTCAGACGTGTGCTCTTCCGATCTAGCATTCCTACACAATTACTGCTG	407
3c	GGAGTTCAGACGTGTGCTCTTCCGATCTGCCCTGGGAGACCTTACAA	408
3d	GGAGTTCAGACGTGTGCTCTTCCGATCTGGAAAAGACAAGCAAGCCAG	409
3e	GGAGTTCAGACGTGTGCTCTTCCGATCTAGGGTGAGTGAATGACAGG	410
3f	GGAGTTCAGACGTGTGCTCTTCCGATCTGAAGACAACCGAATTAGGCG	411
3g	GGAGTTCAGACGTGTGCTCTTCCGATCTTTTGTCTTTTGGTTCCTTCGG	412
3h	GGAGTTCAGACGTGTGCTCTTCCGATCTGCCCTTGCCCTTATCACCTTTG	413
4a	GGAGTTCAGACGTGTGCTCTTCCGATCTGCAAAACGACCAGGAGGATA	414
4b	GGAGTTCAGACGTGTGCTCTTCCGATCTGCCCTGGGAGACCTTACAA	408
4c	GGAGTTCAGACGTGTGCTCTTCCGATCTGGAAAAGACAAGCAAGCCAG	409

4d	GGAGTTCAGACGTGTGCTCTTCCGATCTATACACTTCTGGGATTGGCG	415
5a	GGAGTTCAGACGTGTGCTCTTCCGATCTGAAGATCAAGGTGGGTGTG	416
5b	GGAGTTCAGACGTGTGCTCTTCCGATCTACACCCAAACACCACGAAGAT	417
6a	GGAGTTCAGACGTGTGCTCTTCCGATCTACAACCCCTTTGTTCCCCCTA	418
6b	GGAGTTCAGACGTGTGCTCTTCCGATCTccagccACAAGTTTGTTTT	419
6c	GGAGTTCAGACGTGTGCTCTTCCGATCTtGAAGAAACAAGGCAAGGCT	420
6d	GGAGTTCAGACGTGTGCTCTTCCGATCTTGTGGGAGATGACACCCACAC	421
6e	GGAGTTCAGACGTGTGCTCTTCCGATCTGGGAGAGGACACCCACACTTC	422
6f	GGAGTTCAGACGTGTGCTCTTCCGATCTactgtgctggccTGAATA	423
6g	GGAGTTCAGACGTGTGCTCTTCCGATCTCTCACGAACACAGTATCCACG	424
6h	GGAGTTCAGACGTGTGCTCTTCCGATCTGCATGAGGCCCATTTGTAGTAA	425
6i	GGAGTTCAGACGTGTGCTCTTCCGATCTACTCATCTTCTGTGATTTTGTTC	426
6j	GGAGTTCAGACGTGTGCTCTTCCGATCTATTTCCCATGCTGACTTTTGC	427
6k	GGAGTTCAGACGTGTGCTCTTCCGATCTGGAAGGGACACCCACACTTCT	428
6l	GGAGTTCAGACGTGTGCTCTTCCGATCTTGCCCTTTTATGTGTGTGTG	429
6m	GGAGTTCAGACGTGTGCTCTTCCGATCTggacacttatgtgattccagtGT	430
6n	GGAGTTCAGACGTGTGCTCTTCCGATCTTGCTGTTGGAGAAAAATGTGTTT	431
7a	GGAGTTCAGACGTGTGCTCTTCCGATCTgcaagactctgtctgtaaaCATT	432
7b	GGAGTTCAGACGTGTGCTCTTCCGATCTAAAAGTTCCATTGGCTGTGG	433
7c	GGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTAAACAAACGTTATCAAA	434

8a	GGAGTTCAGACGTGTGCTCTTCCGATCTTTCAGAAATTTCACTGCATCGTG	435
8b	GGAGTTCAGACGTGTGCTCTTCCGATCTAAATCCCCACCTTATCTGGC	436
8c	GGAGTTCAGACGTGTGCTCTTCCGATCTTCCGATCTccagcactGGAAGCTCTTTT	437
8d	GGAGTTCAGACGTGTGCTCTTCCGATCTAGGATATGCTTGAATTATTTCCG	438
8e	GGAGTTCAGACGTGTGCTCTTCCGATCTACCTTGGCCACTCTTCTTT	439
8f	GGAGTTCAGACGTGTGCTCTTCCGATCTTCACCCCTAGGTGGCACAGAT	440
8g	GGAGTTCAGACGTGTGCTCTTCCGATCTgcacttcgctaagaactgtcttC	441
9a	GGAGTTCAGACGTGTGCTCTTCCGATCTgatagcactgCTCCAGGGAT	442
9b	GGAGTTCAGACGTGTGCTCTTCCGATCTaacaacaacaaaCAATGCATAC	443
9c	GGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGCCCTGGATGGAAC	444
9d	GGAGTTCAGACGTGTGCTCTTCCGATCTATCATCCCTAATGCCAAAGC	445
9e	GGAGTTCAGACGTGTGCTCTTCCGATCTgcaagactgtctcgtaaaCATT	432
9f	GGAGTTCAGACGTGTGCTCTTCCGATCTAAAAGTTCCATTGGCTGTGG	433
9g	GGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTAAACAAACGTTATCAAA	434
10a	GGAGTTCAGACGTGTGCTCTTCCGATCTcattttatgattttaattgttGGGG	446
10b	GGAGTTCAGACGTGTGCTCTTCCGATCTGTTGGGTTAGGGCATTGTG	447
10c	GGAGTTCAGACGTGTGCTCTTCCGATCTattgcacctgacCTATGTCT	448
11a	GGAGTTCAGACGTGTGCTCTTCCGATCTTCCAAAATCTGTGGCTTGTG	449
11b	GGAGTTCAGACGTGTGCTCTTCCGATCTCAACAGACACTGGCTGAAGG	450
11c	GGAGTTCAGACGTGTGCTCTTCCGATCTGGTCCTAAGTCTGTGGCTCG	451

11d	GGAGTTCAGACGTGTGCTCTTCCGATCTTCCAAAGAATACCAGCCACC	452
11e	GGAGTTCAGACGTGTGCTCTTCCGATCTTCCACCCCTGTAAAAGTACCA	453
11f	GGAGTTCAGACGTGTGCTCTTCCGATCTGGTCCTAAGTCTGTGGCTCG	451
11g	GGAGTTCAGACGTGTGCTCTTCCGATCTTggttctctttagattcctTTT	454
11h	GGAGTTCAGACGTGTGCTCTTCCGATCTGGTGAAGCTATCAGGACCA	455
11i	GGAGTTCAGACGTGTGCTCTTCCGATCTAGATATGATGAGATCCGCCG	456
11j	GGAGTTCAGACGTGTGCTCTTCCGATCTccagcATCTTTACATGCTTTT	457
12a	GGAGTTCAGACGTGTGCTCTTCCGATCTAGTACACAGTGGCTGCCCAT	458
12b	GGAGTTCAGACGTGTGCTCTTCCGATCTGCCGAGGAAGCATTGTAAAG	459
13a	GGAGTTCAGACGTGTGCTCTTCCGATCTAATAAGGATGACACTTTAGAACTGGA	460
13b	GGAGTTCAGACGTGTGCTCTTCCGATCTaccggccATAAACTCAAG	461
13c	GGAGTTCAGACGTGTGCTCTTCCGATCTGTTAAACAATAATGCTTTGGGCT	462
13d	GGAGTTCAGACGTGTGCTCTTCCGATCTCCTTGTCTCCAACCTCCCAA	463
13e	GGAGTTCAGACGTGTGCTCTTCCGATCTTTGCACCTTCCACCCATAAT	464
14a	GGAGTTCAGACGTGTGCTCTTCCGATCTAGCTACATCTTCACCGCCAC	465
14b	GGAGTTCAGACGTGTGCTCTTCCGATCTGAGGCTACAGCTTCACCACC	466
14c	GGAGTTCAGACGTGTGCTCTTCCGATCTCTCACCACCAAGCTTCACCATCA	467
15a	GGAGTTCAGACGTGTGCTCTTCCGATCTGCCCTCCACAATAATTATGCC	468
15b	GGAGTTCAGACGTGTGCTCTTCCGATCTTGTGGCCTCCACAATATTCA	469
15c	GGAGTTCAGACGTGTGCTCTTCCGATCTTGGTGATCTTGTGCTGCTTTG	470

15d	GGAGTTCAGACGTGTGCTCTTCCGATCTATTGATTCATGCCCTCTTGC	471
15e	GGAGTTCAGACGTGTGCTCTTCCGATCTggccATAGCAATGGTGATCT	472
15f	GGAGTTCAGACGTGTGCTCTTCCGATCTTAGTGTTTCCGTTCCCC	473
15g	GGAGTTCAGACGTGTGCTCTTCCGATCTTTCACCCCTGCCAAAGATCA	474
16a	GGAGTTCAGACGTGTGCTCTTCCGATCTCCCTACTGAAGACTGGAGCG	475
16b	GGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGTGCCTATTACAGCAGT	476
16c	GGAGTTCAGACGTGTGCTCTTCCGATCTCCCTACTGAAGACTGGAGCG	475
16d	GGAGTTCAGACGTGTGCTCTTCCGATCTCCCTACTGAAGACTGGAGCG	475
16e	GGAGTTCAGACGTGTGCTCTTCCGATCTTGAACAATGGAGCACTCAGC	477
16f	GGAGTTCAGACGTGTGCTCTTCCGATCTCAGACTGGTCTGAAAGCGTG	478
16g	GGAGTTCAGACGTGTGCTCTTCCGATCTTGAAGACTGGAGCACTCAGC	479
16h	GGAGTTCAGACGTGTGCTCTTCCGATCTTCCAGCACCTGAGTTCACTG	480
16i	GGAGTTCAGACGTGTGCTCTTCCGATCTTGCCCTATTACAGAGTTCCCT	481
17a	GGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGTGCCTATTACAGCAGT	476
17b	GGAGTTCAGACGTGTGCTCTTCCGATCTCCCTACTGAAGACTGGAGCG	475
17c	GGAGTTCAGACGTGTGCTCTTCCGATCTCCCTACTGAAGACTGGAGCG	475
17d	GGAGTTCAGACGTGTGCTCTTCCGATCTTGAACAATGGAGCACTCAGC	477
17e	GGAGTTCAGACGTGTGCTCTTCCGATCTAAGACTGGAGCACTCAGCGT	482
17f	GGAGTTCAGACGTGTGCTCTTCCGATCTTGCCCTATTACAGAGTTCCCT	481
18a	GGAGTTCAGACGTGTGCTCTTCCGATCTTAGAAGGTGGAGATGCTGGC	483

18b	GGAGTTCAGACGTGTGCTCTTCCGATCTGGCAACAGACAGGACCAGAT	484
18c	GGAGTTCAGACGTGTGCTCTTCCGATCTTAGAAGGTGGAGATGCTGGC	483
18d	GGAGTTCAGACGTGTGCTCTTCCGATCTTAGAAGGTGGAGATGCTGGC	483
19a	GGAGTTCAGACGTGTGCTCTTCCGATCTTTGTGCACGTTAAGCACCTCTG	485
19b	GGAGTTCAGACGTGTGCTCTTCCGATCTAAGAGAGCTTCCTGACACGC	486
19c	GGAGTTCAGACGTGTGCTCTTCCGATCTTCTAAGAGCCAAAGACAGCTTCC	487
20a	GGAGTTCAGACGTGTGCTCTTCCGATCTAGCTTGTTCCTTTTCAGCCG	488
20b	GGAGTTCAGACGTGTGCTCTTCCGATCTCAATGCGTTTCCTTTTAGCC	489
20c	GGAGTTCAGACGTGTGCTCTTCCGATCTAGTATTTCTTTTGCCCTTTCGGC	490
20d	GGAGTTCAGACGTGTGCTCTTCCGATCTTATCCATTTCTGCCTCTGCC	491
20e	GGAGTTCAGACGTGTGCTCTTCCGATCTCCCTTTTCTTTTCCAGAGG	492
20f	GGAGTTCAGACGTGTGCTCTTCCGATCTAAAAGAAAATCGCCTTTTCGG	493
20g	GGAGTTCAGACGTGTGCTCTTCCGATCTtcttTTCCTTTTCAGCCCGGA	494
20h	GGAGTTCAGACGTGTGCTCTTCCGATCTTTGTCTGCCTCAGCTCAAGA	495
20i	GGAGTTCAGACGTGTGCTCTTCCGATCTaaaacTCCCTTCCCTTTTGGC	496
21a	GGAGTTCAGACGTGTGCTCTTCCGATCTCGACCTCTGGAAGGAGACTG	497
21b	GGAGTTCAGACGTGTGCTCTTCCGATCTCGACCTCTGGAAGGAGACTG	497
21c	GGAGTTCAGACGTGTGCTCTTCCGATCTCACTGACCACCTCGTCAAGA	498
21d	GGAGTTCAGACGTGTGCTCTTCCGATCTCACTGACCACCTCGTCAAGA	498
21e	GGAGTTCAGACGTGTGCTCTTCCGATCTCGACCTCTGGAAGGAGACTG	497

21f	GGAGTTCAGACGTGTGCTCTTCCGATCTGACCTACAGCTACCCGACCC	499
22a	GGAGTTCAGACGTGTGCTCTTCCGATCTGAACGTC AAGCATCCCAAGT	500
22b	GGAGTTCAGACGTGTGCTCTTCCGATCTcatttataattgattgGGGG	446
22c	GGAGTTCAGACGTGTGCTCTTCCGATCTGTTGGTTAGGGCATTGTG	447
22d	GGAGTTCAGACGTGTGCTCTTCCGATCTAAAAGACGACCATCCACCAG	501
22e	GGAGTTCAGACGTGTGCTCTTCCGATCTatgatctaccacagaGACC	502
22f	GGAGTTCAGACGTGTGCTCTTCCGATCTGAGCAGCTCCTCCTTGAGAG	503
22g	GGAGTTCAGACGTGTGCTCTTCCGATCTAGCAAAGTGGGGTGTGAGAC	504
22h	GGAGTTCAGACGTGTGCTCTTCCGATCTtgccagtgaggatgataagaaA	505
22i	GGAGTTCAGACGTGTGCTCTTCCGATCTCGGATAGTGGATGGCAAAGT	506
22j	GGAGTTCAGACGTGTGCTCTTCCGATCTTCATTATATTAGCTTTGTAGCGAGA	507

*Spacer group subsets indicated by lower case letters; genomic location of spacer group subsets are provided in Table 3.

Table 5
sgRNA assembly primers for sgRNPs

- 5 Primer 1: AGTAATAATACGACTCACTATAG (SEQ ID NO:508)
 Primer 2: AAAAAAGCACCAGCTCGGTGCCACTTTTCAAGTTGATAACGGACTAGC (SEQ ID NO: 509)
 Primer 3: Unique to sgRNA and listed in Table 1 as SEQ ID NOS: 1-4, respectively
 Primer 4: GTTTTAGAGCTAGAAATAGCAAGTTAAATAAAGGCTAGTCCGTTATCAAC (SEQ ID NO: 510)
 Primer 5: AAAAAAGCACCAGCTCGGTGCC (SEQ ID NO: 511)

Table 6
Illumina sequencing index primers (i5) for PCR2 (i5 index sequences are bolded)

<u>Well</u>	<u>Name</u>	<u>Sequence</u>	<u>SEQ ID NO.:</u>
5			
A1	ILMN_AMP_FORi5_BC1	AATGATACGGCGACCAACCGAGATCTACACT AGATCGCACACTCTTCCCTACACGACG	512
B1	ILMN_AMP_FORi5_BC2	AATGATACGGCGACCAACCGAGATCTACACT CTCTATACACTCTTTCCCTACACGACG	513
C1	ILMN_AMP_FORi5_BC3	AATGATACGGCGACCAACCGAGATCTACACT TATCCTCTACACTCTTTCCCTACACGACG	514
D1	ILMN_AMP_FORi5_BC4	AATGATACGGCGACCAACCGAGATCTACAC AGAGTAGAGACACTCTTTCCCTACACGACG	515
E1	ILMN_AMP_FORi5_BC5	AATGATACGGCGACCAACCGAGATCTACAC GTAAAGGAGACACTCTTTCCCTACACGACG	516
F1	ILMN_AMP_FORi5_BC6	AATGATACGGCGACCAACCGAGATCTACAC ACTGCATTAACACTCTTTCCCTACACGACG	517
G1	ILMN_AMP_FORi5_BC7	AATGATACGGCGACCAACCGAGATCTACAC AAAGGAGTAACACTCTTTCCCTACACGACG	518
H1	ILMN_AMP_FORi5_BC8	AATGATACGGCGACCAACCGAGATCTACAC CTAAGCCCTACACTCTTTCCCTACACGACG	519
A2	ILMN_AMP_FORi5_BC9	AATGATACGGCGACCAACCGAGATCTACACT GAACCTTACACTCTTTCCCTACACGACG	520
B2	ILMN_AMP_FORi5_BC10	AATGATACGGCGACCAACCGAGATCTACACT GTAAAGTACACTCTTTCCCTACACGACG	521
C2	ILMN_AMP_FORi5_BC11	AATGATACGGCGACCAACCGAGATCTACACT TAAGTTCCACACTCTTTCCCTACACGACG	522
D2	ILMN_AMP_FORi5_BC12	AATGATACGGCGACCAACCGAGATCTACAC ATAGAGGCACACTCTTTCCCTACACGACG	523
E2	ILMN_AMP_FORi5_BC13	AATGATACGGCGACCAACCGAGATCTACAC GGCTCTGAACACTCTTTCCCTACACGACG	524
F2	ILMN_AMP_FORi5_BC14	AATGATACGGCGACCAACCGAGATCTACAC AGGCGAAGACACTCTTTCCCTACACGACG	525
G2	ILMN_AMP_FORi5_BC15	AATGATACGGCGACCAACCGAGATCTACACT TAATCTTAACACTCTTTCCCTACACGACG	526
H2	ILMN_AMP_FORi5_BC16	AATGATACGGCGACCAACCGAGATCTACAC CAGGACGTACACTCTTTCCCTACACGACG	527

Table 7

Illumina sequencing index primers (i7) for PCR2 (i7 index sequences are bolded)

<u>Well position</u>	<u>Name</u>	<u>Sequence</u>	<u>SEQ ID NO.:</u>
A1	ILMN_AMP_REV17_BC1	CAAGCAGAAAGACGGCATAACGAGAT ATTGGTCA GTGACTGGAGTTCAGACGTGTGCTC	528
B1	ILMN_AMP_REV17_BC2	CAAGCAGAAAGACGGCATAACGAGAT TAAAAATG GTGACTGGAGTTCAGACGTGTGCTC	529
C1	ILMN_AMP_REV17_BC3	CAAGCAGAAAGACGGCATAACGAGAT ATCACTGT GTGACTGGAGTTCAGACGTGTGCTC	530
D1	ILMN_AMP_REV17_BC4	CAAGCAGAAAGACGGCATAACGAGAT TATTTCA CGTGTGACTGGAGTTCAGACGTGTGCTC	531
E1	ILMN_AMP_REV17_BC5	CAAGCAGAAAGACGGCATAACGAGAT ATATGG CGTGTGACTGGAGTTCAGACGTGTGCTC	532
F1	ILMN_AMP_REV17_BC6	CAAGCAGAAAGACGGCATAACGAGAT TATAC CAAGGTGACTGGAGTTCAGACGTGTGCTC	533
G1	ILMN_AMP_REV17_BC7	CAAGCAGAAAGACGGCATAACGAGAT ATGATCT GGTGTGACTGGAGTTCAGACGTGTGCTC	534
H1	ILMN_AMP_REV17_BC8	CAAGCAGAAAGACGGCATAACGAGAT TACTCT ACGTGTGACTGGAGTTCAGACGTGTGCTC	535
A2	ILMN_AMP_REV17_BC9	CAAGCAGAAAGACGGCATAACGAGAT ATAAG CTAGTGTGACTGGAGTTCAGACGTGTGCTC	536
B2	ILMN_AMP_REV17_BC10	CAAGCAGAAAGACGGCATAACGAGAT TAGTAT AGGTGACTGGAGTTCAGACGTGTGCTC	537
C2	ILMN_AMP_REV17_BC11	CAAGCAGAAAGACGGCATAACGAGAT ATTACA AGGTGACTGGAGTTCAGACGTGTGCTC	538
D2	ILMN_AMP_REV17_BC12	CAAGCAGAAAGACGGCATAACGAGAT TAAAT GGCGTGTGACTGGAGTTCAGACGTGTGCTC	539
E2	ILMN_AMP_REV17_BC13	CAAGCAGAAAGACGGCATAACGAGAT ATCTCT ACGTGTGACTGGAGTTCAGACGTGTGCTC	540
F2	ILMN_AMP_REV17_BC14	CAAGCAGAAAGACGGCATAACGAGAT TAGATCT GGTGTGACTGGAGTTCAGACGTGTGCTC	541
G2	ILMN_AMP_REV17_BC15	CAAGCAGAAAGACGGCATAACGAGAT ATGCG ACGTGTGACTGGAGTTCAGACGTGTGCTC	542
H2	ILMN_AMP_REV17_BC16	CAAGCAGAAAGACGGCATAACGAGAT TAAAT CACTGTGTGACTGGAGTTCAGACGTGTGCTC	543
A3	ILMN_AMP_REV17_BC17	CAAGCAGAAAGACGGCATAACGAGAT ATTTT CACTGTGTGACTGGAGTTCAGACGTGTGCTC	544

B3	ILMN_AMP_REV17_BC18	CAAGCAGAAAGACGGGCATACGAGATTACACTGTGTGCTGACTGGAGTTCAGACCGTGTGCTC	545
C3	ILMN_AMP_REV17_BC19	CAAGCAGAAAGACGGGCATACGAGATATGGCCACCTGACTGGAGTTCAGACCGTGTGCTC	546
D3	ILMN_AMP_REV17_BC20	CAAGCAGAAAGACGGGCATACGAGATTACCGGTGGTGTGACTGGAGTTCAGACCGTGTGCTC	547
E3	ILMN_AMP_REV17_BC21	CAAGCAGAAAGACGGGCATACGAGATATCGAAACCTGACTGGAGTTCAGACCGTGTGCTC	548
F3	ILMN_AMP_REV17_BC22	CAAGCAGAAAGACGGGCATACGAGATTATAGTTGGTGTGACTGGAGTTCAGACCGTGTGCTC	549
G3	ILMN_AMP_REV17_BC23	CAAGCAGAAAGACGGGCATACGAGATATCGTACGGTGTGACTGGAGTTCAGACCGTGTGCTC	550
H3	ILMN_AMP_REV17_BC24	CAAGCAGAAAGACGGGCATACGAGATTAGAATGACTGACTGGAGTTCAGACCGTGTGCTC	551
A4	ILMN_AMP_REV17_BC25	CAAGCAGAAAGACGGGCATACGAGATATGCTACCGTGTGACTGGAGTTCAGACCGTGTGCTC	552
B4	ILMN_AMP_REV17_BC26	CAAGCAGAAAGACGGGCATACGAGATTAAATCGTGTGACTGGAGTTCAGACCGTGTGCTC	553
C4	ILMN_AMP_REV17_BC27	CAAGCAGAAAGACGGGCATACGAGATATATCACTGTGACTGGAGTTCAGACCGTGTGCTC	554
D4	ILMN_AMP_REV17_BC28	CAAGCAGAAAGACGGGCATACGAGATTAGCGGACGTGACTGGAGTTCAGACCGTGTGCTC	555
E4	ILMN_AMP_REV17_BC29	CAAGCAGAAAGACGGGCATACGAGATATGCTCATGTGACTGGAGTTCAGACCGTGTGCTC	556
F4	ILMN_AMP_REV17_BC30	CAAGCAGAAAGACGGGCATACGAGATTACGATTAGTGTGACTGGAGTTCAGACCGTGTGCTC	557
G4	ILMN_AMP_REV17_BC31	CAAGCAGAAAGACGGGCATACGAGATATAGGAATGTGACTGGAGTTCAGACCGTGTGCTC	558
H4	ILMN_AMP_REV17_BC32	CAAGCAGAAAGACGGGCATACGAGATATTAGTTGGTGTGACTGGAGTTCAGACCGTGTGCTC	559
A5	ILMN_AMP_REV17_BC33	CAAGCAGAAAGACGGGCATACGAGATTACGAAACGTGACTGGAGTTCAGACCGTGTGCTC	560
B5	ILMN_AMP_REV17_BC34	CAAGCAGAAAGACGGGCATACGAGATATCCGGTGGTGTGACTGGAGTTCAGACCGTGTGCTC	561
C5	ILMN_AMP_REV17_BC35	CAAGCAGAAAGACGGGCATACGAGATTAGGCCACGTGACTGGAGTTCAGACCGTGTGCTC	562
D5	ILMN_AMP_REV17_BC36	CAAGCAGAAAGACGGGCATACGAGATATATCGTGTGACTGGAGTTCAGACCGTGTGCTC	563
E5	ILMN_AMP_REV17_BC37	CAAGCAGAAAGACGGGCATACGAGATTAGCTACCGTGTGACTGGAGTTCAGACCGTGTGCTC	564

F5	ILMN_AMP_REV17_BC38	CAAGCAGAAAGACGGCATACGAGATATCGCCCTGGTGACTGGAGTTCAGACGTTGTGCTC	565
G5	ILMN_AMP_REV17_BC39	CAAGCAGAAAGACGGCATACGAGATATAAAATGGTGACTGGAGTTCAGACGTTGTGCTC	566
H5	ILMN_AMP_REV17_BC40	CAAGCAGAAAGACGGCATACGAGATTATGGTCACTGACTGGAGTTCAGACGTTGTGCTC	567
A6	ILMN_AMP_REV17_BC41	CAAGCAGAAAGACGGCATACGAGATATATTCGGTGACTGGAGTTCAGACGTTGTGCTC	568
B6	ILMN_AMP_REV17_BC42	CAAGCAGAAAGACGGCATACGAGATATGTATAGTGACTGGAGTTCAGACGTTGTGCTC	569
C6	ILMN_AMP_REV17_BC43	CAAGCAGAAAGACGGCATACGAGATTAAAGCTAGTGACTGGAGTTCAGACGTTGTGCTC	570
D6	ILMN_AMP_REV17_BC44	CAAGCAGAAAGACGGCATACGAGATATCGATTAGTGACTGGAGTTCAGACGTTGTGCTC	571
E6	ILMN_AMP_REV17_BC45	CAAGCAGAAAGACGGCATACGAGATTAGCTCATGTGACTGGAGTTCAGACGTTGTGCTC	572
F6	ILMN_AMP_REV17_BC46	CAAGCAGAAAGACGGCATACGAGATATGAATGAGTGACTGGAGTTCAGACGTTGTGCTC	573
G6	ILMN_AMP_REV17_BC47	CAAGCAGAAAGACGGCATACGAGATTACGTACGGTGACTGGAGTTCAGACGTTGTGCTC	574
H6	ILMN_AMP_REV17_BC48	CAAGCAGAAAGACGGCATACGAGATATCGTGATGTGACTGGAGTTCAGACGTTGTGCTC	575
A7	ILMN_AMP_REV17_BC49	CAAGCAGAAAGACGGCATACGAGATATACATCGGTGACTGGAGTTCAGACGTTGTGCTC	576
B7	ILMN_AMP_REV17_BC50	CAAGCAGAAAGACGGCATACGAGATATGCCCTAAGTGACTGGAGTTCAGACGTTGTGCTC	577
C7	ILMN_AMP_REV17_BC51	CAAGCAGAAAGACGGCATACGAGATATTCAAAGTGACTGGAGTTCAGACGTTGTGCTC	578
D7	ILMN_AMP_REV17_BC52	CAAGCAGAAAGACGGCATACGAGATATCTGATCGTGACTGGAGTTCAGACGTTGTGCTC	579
E7	ILMN_AMP_REV17_BC53	CAAGCAGAAAGACGGCATACGAGATATGTAGCCGTGACTGGAGTTCAGACGTTGTGCTC	580
F7	ILMN_AMP_REV17_BC54	CAAGCAGAAAGACGGCATACGAGATATTTGACTGTGACTGGAGTTCAGACGTTGTGCTC	581
G7	ILMN_AMP_REV17_BC55	CAAGCAGAAAGACGGCATACGAGATATGGAACGTGTGACTGGAGTTCAGACGTTGTGCTC	582
H7	ILMN_AMP_REV17_BC56	CAAGCAGAAAGACGGCATACGAGATATTGACATGTGACTGGAGTTCAGACGTTGTGCTC	583
A8	ILMN_AMP_REV17_BC57	CAAGCAGAAAGACGGCATACGAGATATGGACGGGTGACTGGAGTTCAGACGTTGTGCTC	584

B8	ILMN_AMP_REV17_BC58	CAAGCAGAAAGACGGCATACGAGATATCCACTCGTGA	585
C8	ILMN_AMP_REV17_BC59	CAAGCAGAAAGACGGCATACGAGATATCTTTTGGTGACTGGAGTTCAGACG	586
D8	ILMN_AMP_REV17_BC60	CAAGCAGAAAGACGGCATACGAGATATTGAGTGGTGA	587
E8	ILMN_AMP_REV17_BC61	CAAGCAGAAAGACGGCATACGAGATATGCCATGGTGA	588
F8	ILMN_AMP_REV17_BC62	CAAGCAGAAAGACGGCATACGAGATATTGTTGGTGACTGGAGTTCAGACG	589
G8	ILMN_AMP_REV17_BC63	CAAGCAGAAAGACGGCATACGAGATATAGCTAGGTGA	590
H8	ILMN_AMP_REV17_BC64	CAAGCAGAAAGACGGCATACGAGATATTCTGAGGTGA	591
A9	ILMN_AMP_REV17_BC65	CAAGCAGAAAGACGGCATACGAGATATGTCTGCTGA	592
B9	ILMN_AMP_REV17_BC66	CAAGCAGAAAGACGGCATACGAGATATGCTGTAGTGA	593
C9	ILMN_AMP_REV17_BC67	CAAGCAGAAAGACGGCATACGAGATATATTATAGTGA	594
D9	ILMN_AMP_REV17_BC68	CAAGCAGAAAGACGGCATACGAGATATTGCGGAGTGA	595
E9	ILMN_AMP_REV17_BC69	CAAGCAGAAAGACGGCATACGAGATATCTTCGAGTGA	596
F9	ILMN_AMP_REV17_BC70	CAAGCAGAAAGACGGCATACGAGATATTGCCGAGTGA	597
G9	ILMN_AMP_REV17_BC71	CAAGCAGAAAGACGGCATACGAGATTACGTGATGTGA	598
H9	ILMN_AMP_REV17_BC72	CAAGCAGAAAGACGGCATACGAGATTAACATCGGTGA	599
A10	ILMN_AMP_REV17_BC73	CAAGCAGAAAGACGGCATACGAGATTAGCCTAAGTGA	600
B10	ILMN_AMP_REV17_BC74	CAAGCAGAAAGACGGCATACGAGATTATCAAGTGTGA	601
C10	ILMN_AMP_REV17_BC75	CAAGCAGAAAGACGGCATACGAGATTACTGATCGTGA	602
D10	ILMN_AMP_REV17_BC76	CAAGCAGAAAGACGGCATACGAGATTAGTAGCCGTGA	603
E10	ILMN_AMP_REV17_BC77	CAAGCAGAAAGACGGCATACGAGATTATTGACTGTGA	604

F10	ILMN_AMP_REV17_BC78	CAAGCAGAAAGACGGCATAACGAGATTAGGA ACT GTGTGACTGGAGTTCAGACGTTGTGCTC	605
G10	ILMN_AMP_REV17_BC79	CAAGCAGAAAGACGGCATAACGAGATTATGACATGTGACTGGAGTTCAGACGTTGTGCTC	606
H10	ILMN_AMP_REV17_BC80	CAAGCAGAAAGACGGCATAACGAGATTAGGACGGGTGACTGGAGTTCAGACGTTGTGCTC	607
A11	ILMN_AMP_REV17_BC81	CAAGCAGAAAGACGGCATAACGAGATTACCACTCGTGACTGGAGTTCAGACGTTGTGCTC	608
B11	ILMN_AMP_REV17_BC82	CAAGCAGAAAGACGGCATAACGAGATTAAAGAAATGTGACTGGAGTTCAGACGTTGTGCTC	609
C11	ILMN_AMP_REV17_BC83	CAAGCAGAAAGACGGCATAACGAGATTACTTTTGGTGACTGGAGTTCAGACGTTGTGCTC	610
D11	ILMN_AMP_REV17_BC84	CAAGCAGAAAGACGGCATAACGAGATTATGAGTGGTGACTGGAGTTCAGACGTTGTGCTC	611
E11	ILMN_AMP_REV17_BC85	CAAGCAGAAAGACGGCATAACGAGATTACGCCCTGGTGACTGGAGTTCAGACGTTGTGCTC	612
F11	ILMN_AMP_REV17_BC86	CAAGCAGAAAGACGGCATAACGAGATTAGCCATGGTGACTGGAGTTCAGACGTTGTGCTC	613
G11	ILMN_AMP_REV17_BC87	CAAGCAGAAAGACGGCATAACGAGATTATGTTGGGTGACTGGAGTTCAGACGTTGTGCTC	614
H11	ILMN_AMP_REV17_BC88	CAAGCAGAAAGACGGCATAACGAGATTAAATCCGGTGACTGGAGTTCAGACGTTGTGCTC	615
A12	ILMN_AMP_REV17_BC89	CAAGCAGAAAGACGGCATAACGAGATTAAAGCTAGGTGACTGGAGTTCAGACGTTGTGCTC	616
B12	ILMN_AMP_REV17_BC90	CAAGCAGAAAGACGGCATAACGAGATTATCTGAGGTGACTGGAGTTCAGACGTTGTGCTC	617
C12	ILMN_AMP_REV17_BC91	CAAGCAGAAAGACGGCATAACGAGATTAGTCGTCGTGACTGGAGTTCAGACGTTGTGCTC	618
D12	ILMN_AMP_REV17_BC92	CAAGCAGAAAGACGGCATAACGAGATTAGCTGTAGTGACTGGAGTTCAGACGTTGTGCTC	619
E12	ILMN_AMP_REV17_BC93	CAAGCAGAAAGACGGCATAACGAGATTAAATTATAGTGACTGGAGTTCAGACGTTGTGCTC	620
F12	ILMN_AMP_REV17_BC94	CAAGCAGAAAGACGGCATAACGAGATTATCGGGAGTGACTGGAGTTCAGACGTTGTGCTC	621
G12	ILMN_AMP_REV17_BC95	CAAGCAGAAAGACGGCATAACGAGATTACTTCGAGTGACTGGAGTTCAGACGTTGTGCTC	622
H12	ILMN_AMP_REV17_BC96	CAAGCAGAAAGACGGCATAACGAGATTATGCGAGTGACTGGAGTTCAGACGTTGTGCTC	623

CLAIMS

1. A method of modulating DNA repair outcomes comprising:

(1) contacting a selected target nucleic acid comprising a region to be modified with

(a) an agent that suppresses non-homologous end-joining (NHEJ), thereby
favoring alternative end-joining (alt-EJ)/microhomology mediated end-joining
(MMEJ), and

(b) a DNA binding molecule that targets said selected target nucleic acid; and

(2) producing one or more double-strand breaks in the targeted region using a
programmable endonuclease, thereby triggering DNA repair pathways to repair the breaks,
whereby repair of the cleaved target nucleic acid proceeds substantially by MMEJ and is
done in the absence of a donor polynucleotide, thereby modulating the DNA repair outcome.

2. The method of claim 1, wherein the agent in (a) interacts directly with the region to be
modified.

3. The method of claim 1, wherein the agent in (a) does not interact directly with the
region to be modified.

4. A method of modulating DNA repair outcomes comprising:

(1) contacting a selected target nucleic acid comprising a region to be modified with

(a) an agent that suppresses alternative end-joining (alt-EJ)/microhomology
mediated end-joining (MMEJ), thereby favoring non-homologous end-joining
(NHEJ), and

(b) a DNA binding molecule that targets said selected target nucleic acid; and

(2) producing one or more double-strand breaks in the targeted region using a
programmable endonuclease, thereby triggering DNA repair pathways to repair the breaks,
whereby repair of the cleaved target nucleic acid proceeds substantially by NHEJ and is
done in the absence of a donor polynucleotide, thereby modulating the DNA repair outcome.

5. The method of claim 4, wherein the agent in (a) interacts directly with the region to be
modified.

6. The method of claim 1, wherein the agent in (a) does not interact directly with the region to be modified.

5 7. A method of modulating DNA repair outcomes comprising:

(1) contacting a selected target nucleic acid comprising a region to be modified with
 (a) a first agent that suppresses alternative end-joining (alt-EJ)/microhomology
 mediated end-joining (MMEJ), and a second agent that suppresses non-
 homologous end-joining (NHEJ), thereby favoring homology directed repair
10 (HDR), and

 (b) a DNA binding molecule that targets said selected target nucleic acid;

(2) producing one or more double-strand breaks in the targeted region using a
programmable endonuclease; and

(3) inserting at least a portion of a donor polynucleotide into said target nucleic acid at
15 the double-strand break by an HDR DNA repair pathway,
 thereby modulating the DNA repair outcome.

8. The method of claim 7, wherein the first and/or second agents interact directly with
the region to be modified.

20 9. The method of claim 7, wherein the first and/or second agents do not interact directly
with the region to be modified.

10 10. The method of any one of claims 7-9, wherein the first and second agents are
25 different.

11. The method of any one of claims 7-9, wherein the first and second agents are the
same.

30 12. The method of any one of the preceding claims, wherein said method is performed in
a cell.

13. The method of claim 12, wherein said cell comprises a eukaryotic cell.

14. The method of either one of claims 12 or 13, wherein the cell constitutively
5 expresses the Cas endonuclease.

15. The method of any one of claims 1-13, wherein the programmable endonuclease is
complexed to the guide polynucleotide prior to producing one or more double-strand breaks in
the targeted region using a programmable endonuclease.

16. The method of any one of the preceding claims, wherein the DNA binding molecule
comprises a guide polynucleotide and the programmable endonuclease comprises a Cas
endonuclease.

17. The method of claim 16, wherein the Cas endonuclease comprises a Cas9
endonuclease.

18. The method of any one of claims 1-17, wherein the guide polynucleotide comprises a
sgRNA.

19. The method of either one of claims 17 or 18, wherein the DNA binding molecule
comprises sgRNA and the Cas endonuclease comprises Cas9.

20. The method of any one of the preceding claims wherein the method further
25 comprises reducing or restoring function of a gene product.

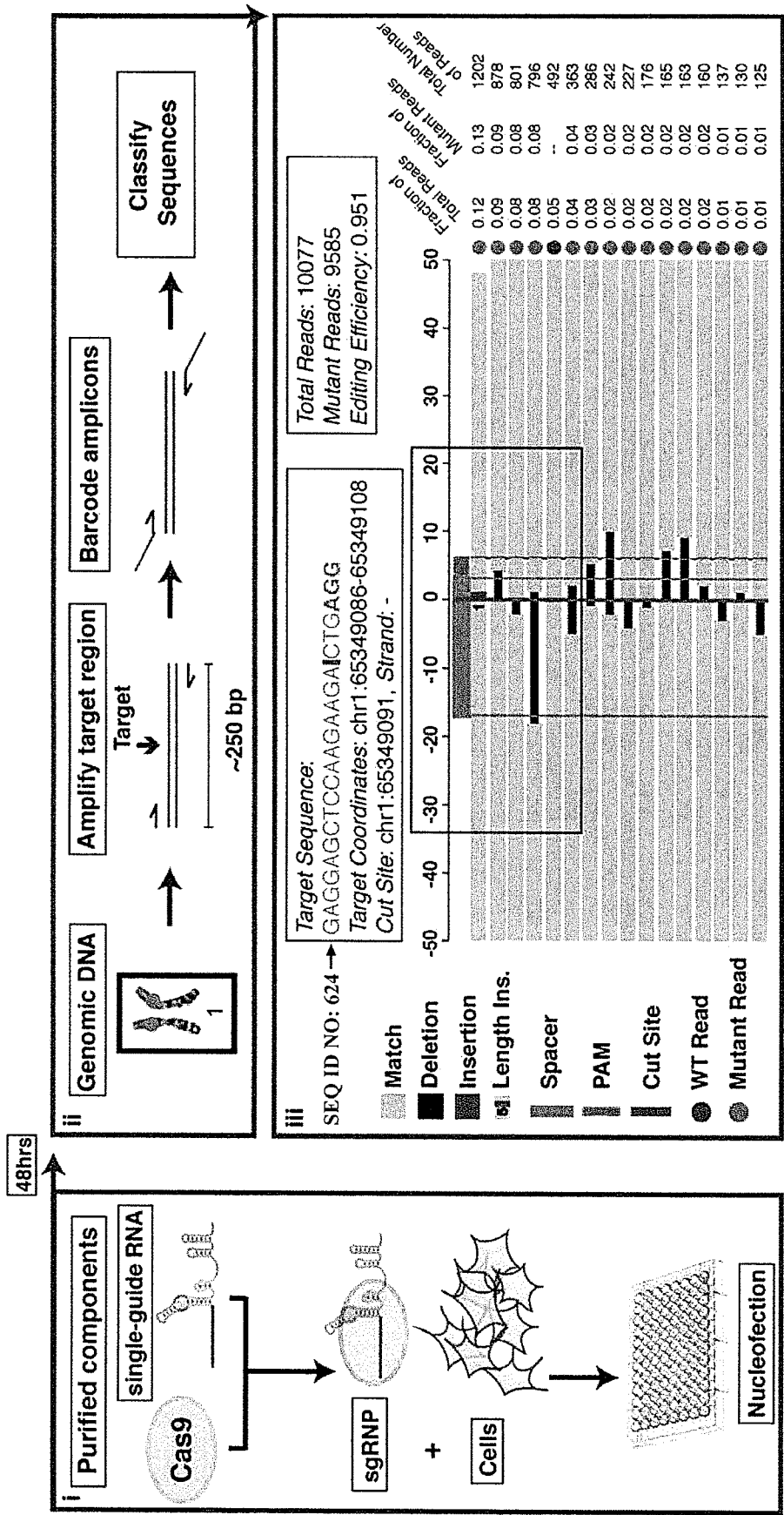
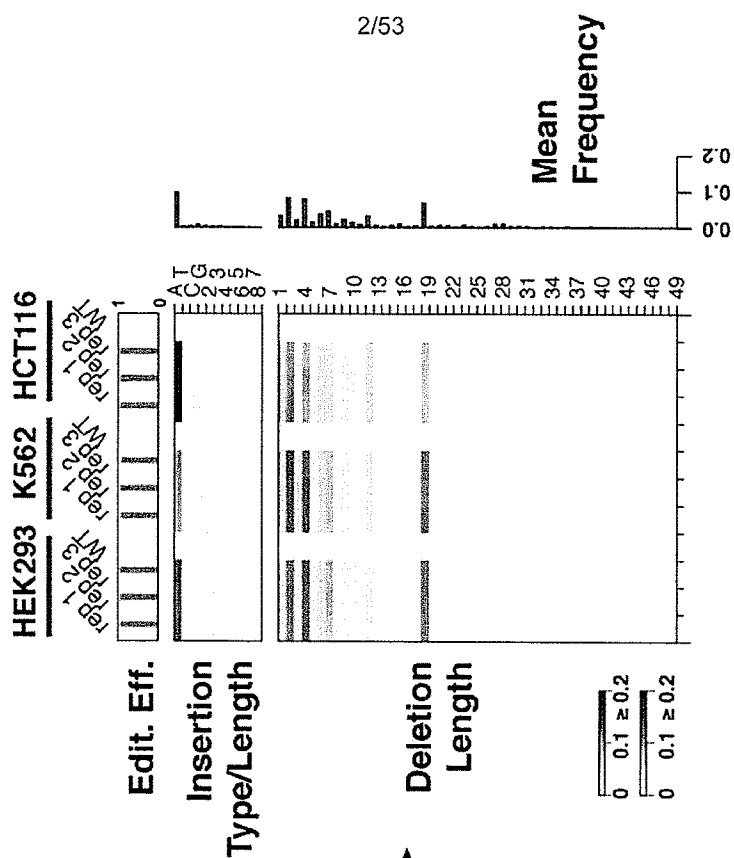
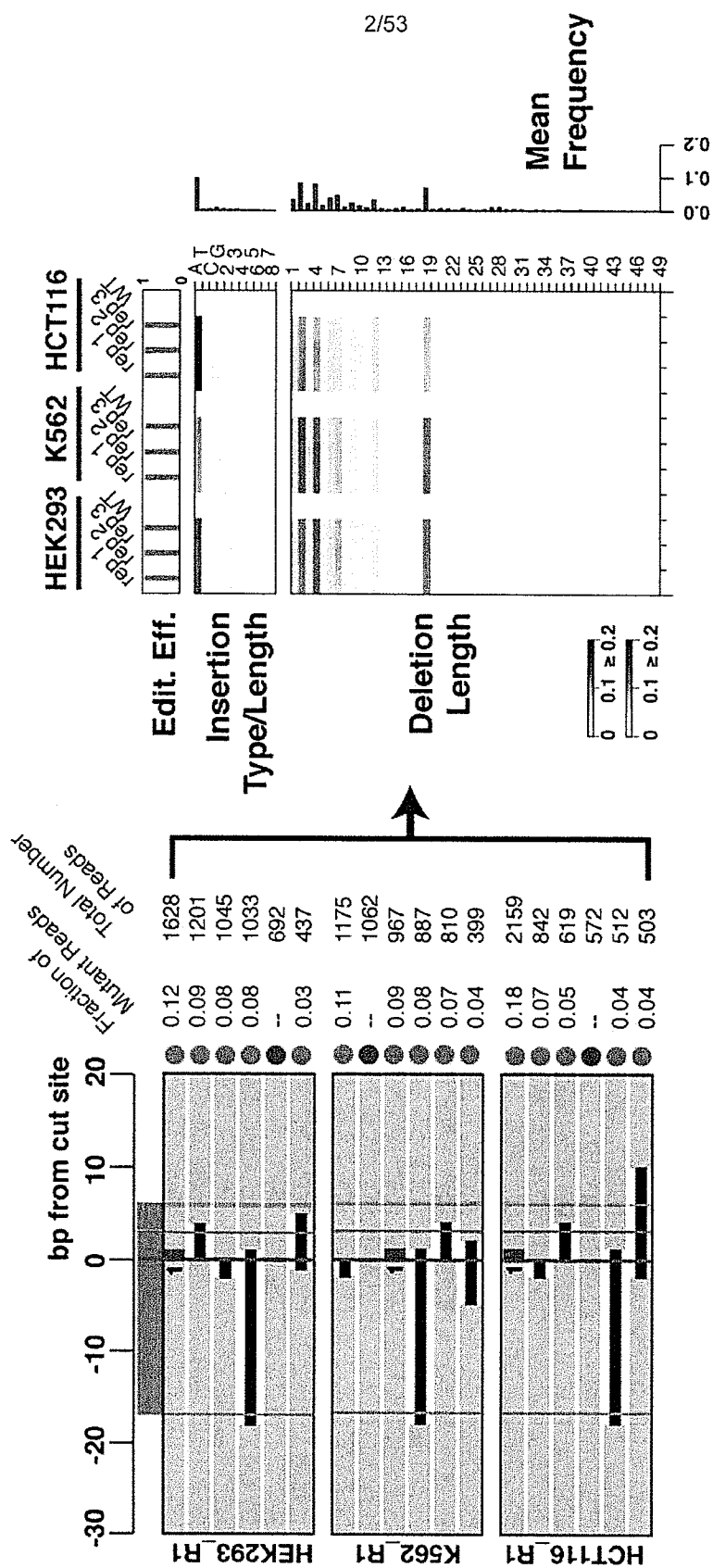


FIG. 1A



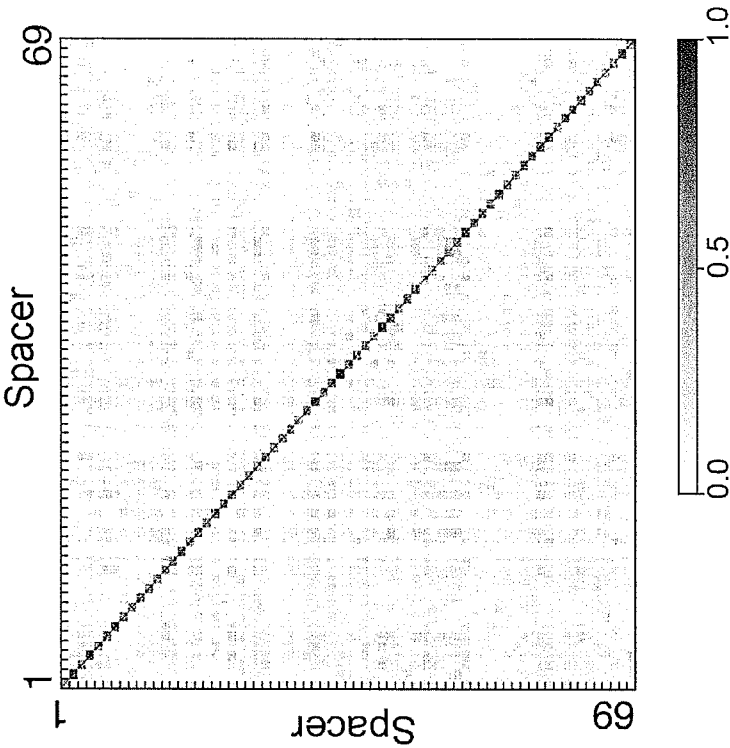


FIG. 2B

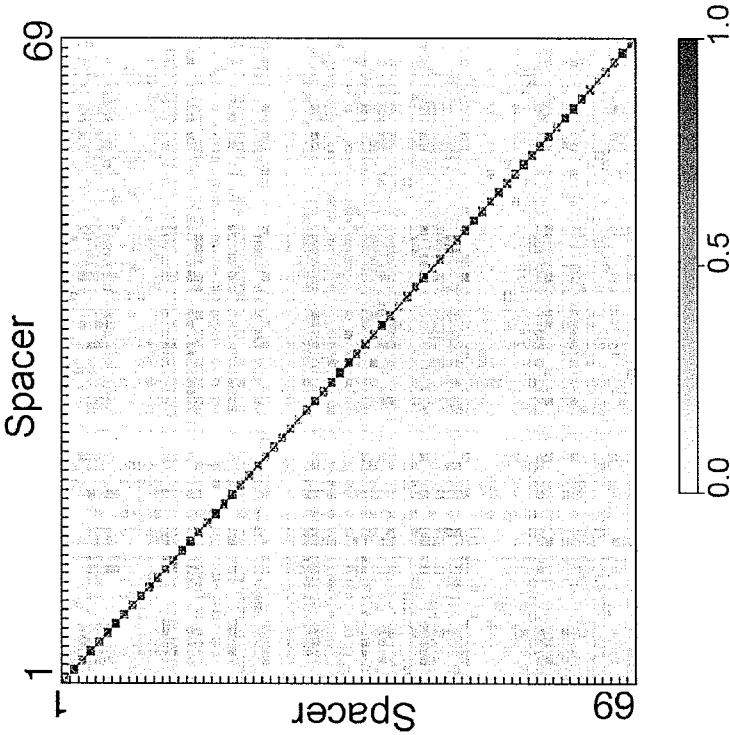


FIG. 2A

ARI	
sgRNP-only	0.920
sgRNP + Lentivirus	0.764

FIG. 2C

Spacer Group 15 Sequence:
GGCAGTGCAGATGAAA**A**ACTGGG ←SEQ ID NO: 649

Target	hg38 coordinates	Region	Gene
15a	chr12:30862020-30862043	noncoding	--
15b	chr2:129496621-129496644	noncoding	--
15c	chr2:68125583-68125606	intron	WDR92
15d	chr20:53421490-53421512	intron	TSHZ2
15e	chr7:76361952-76361974	noncoding	--
15f	chr7:44799840-44799863	exon	PPIA
15g	chr11:43466710-43466733	intron	TTC17

FIG. 3A

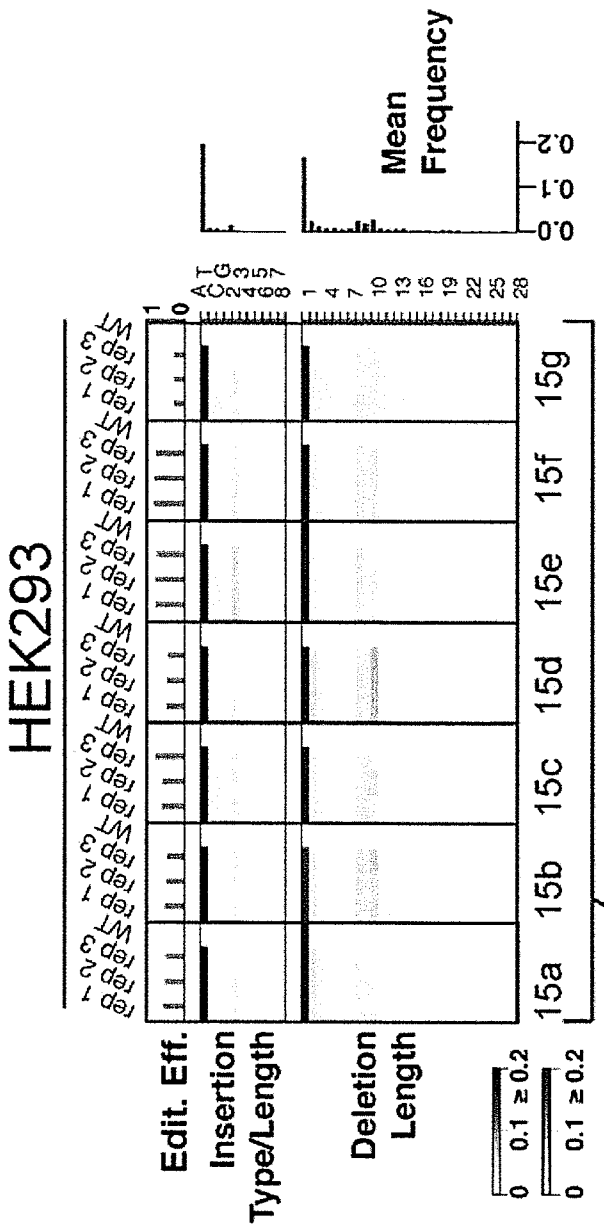


FIG. 3B

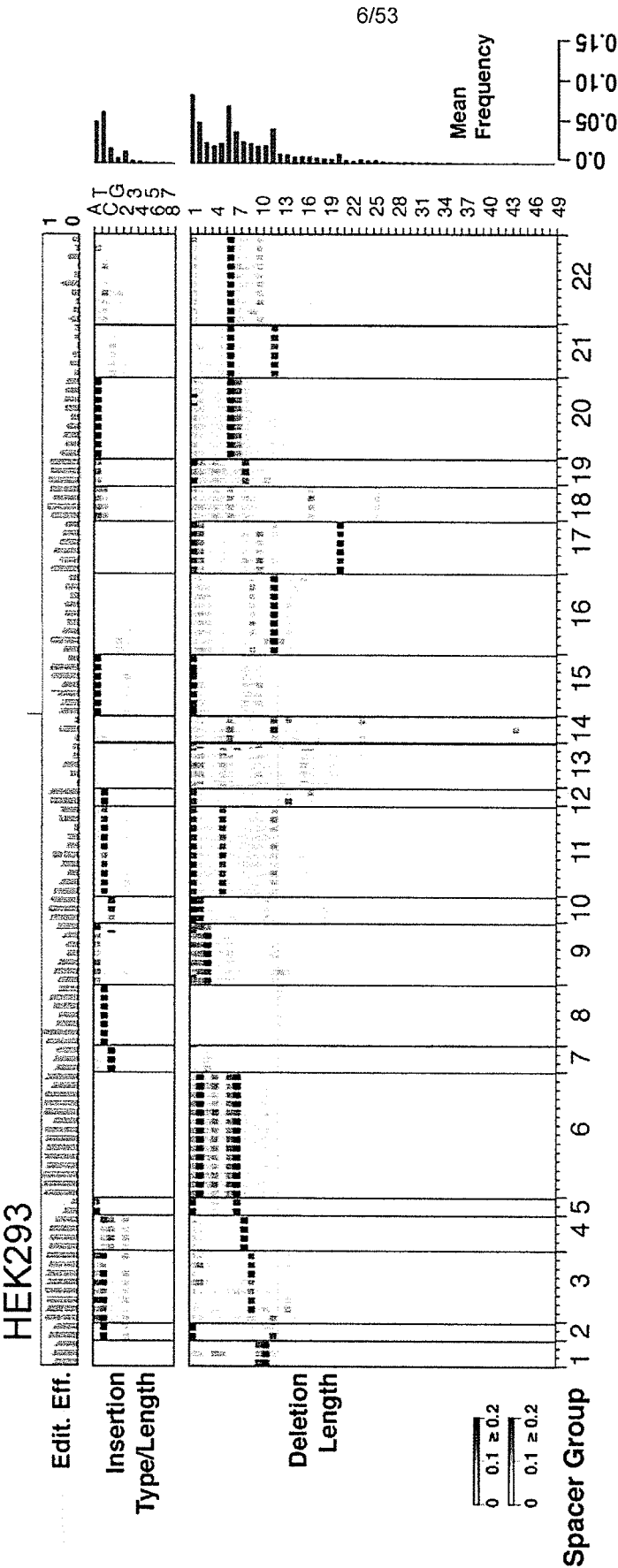


FIG. 3C

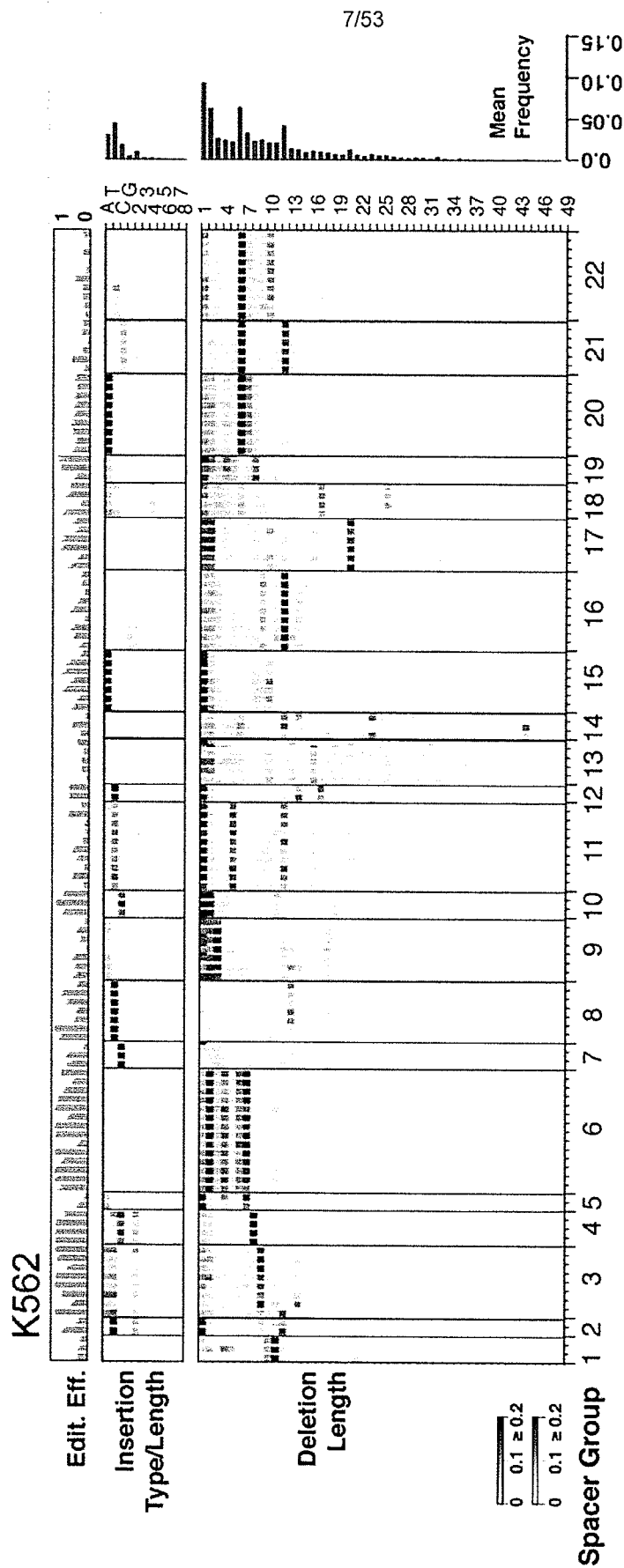


FIG. 3D

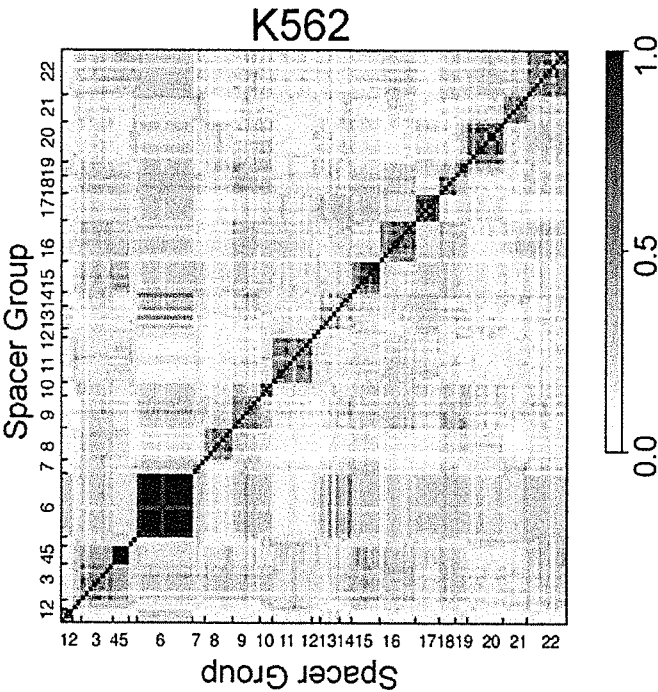


FIG. 3F

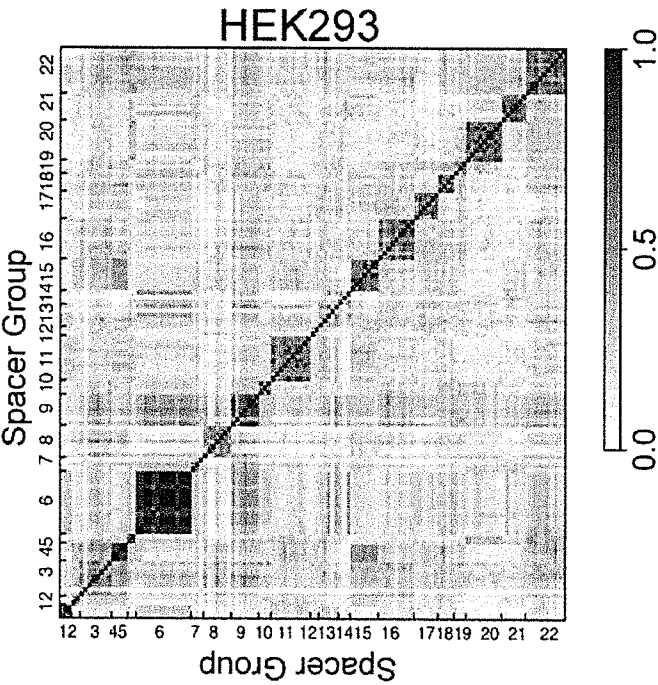


FIG. 3E

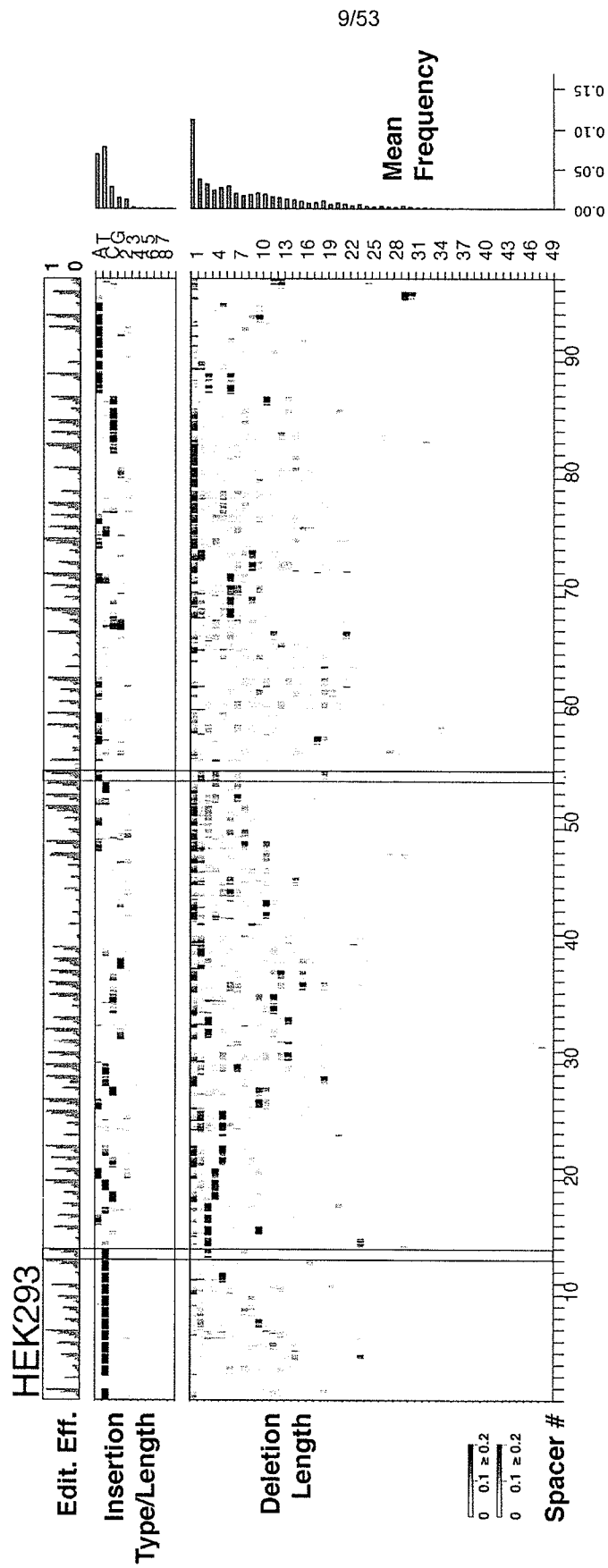
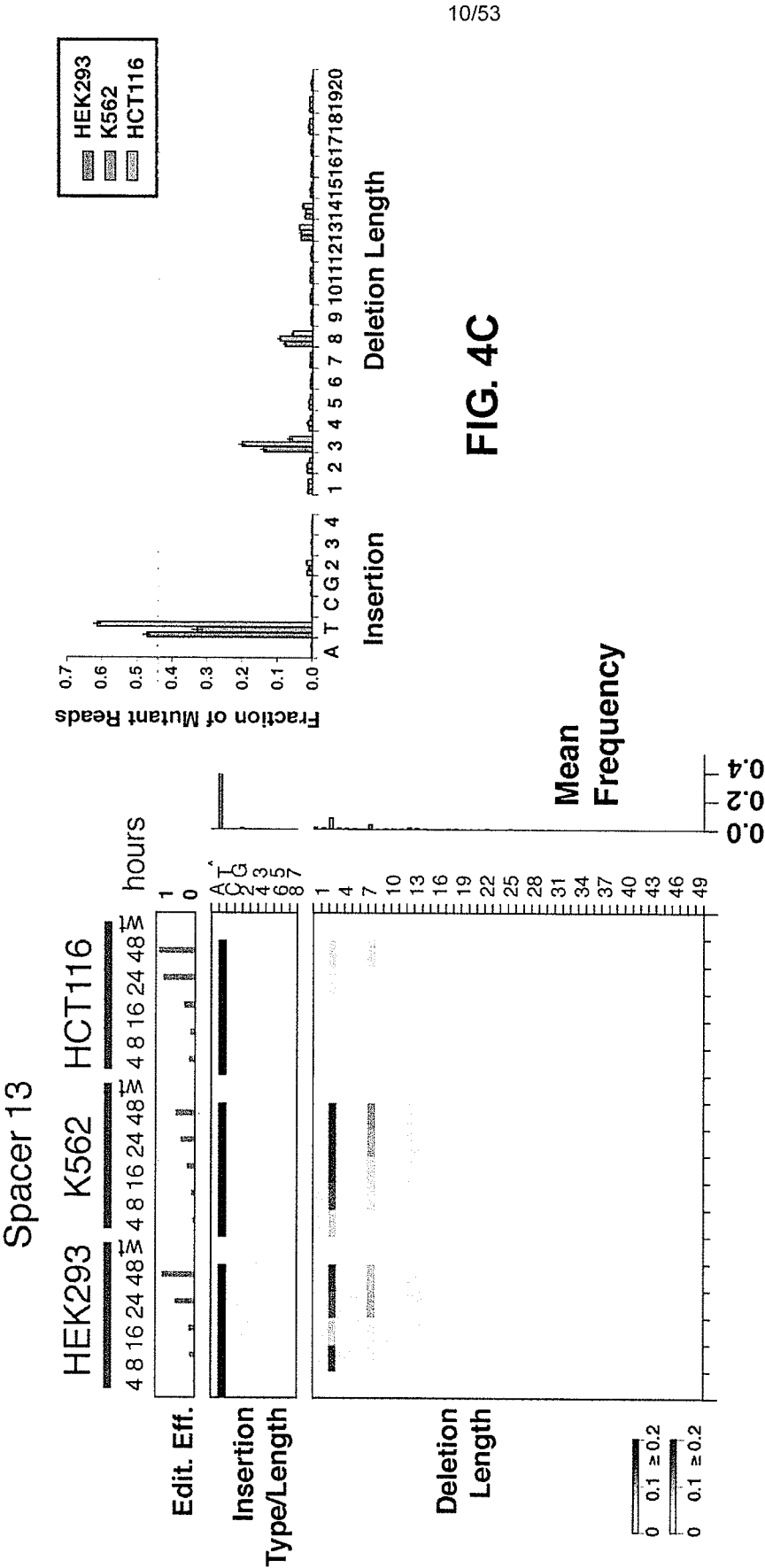
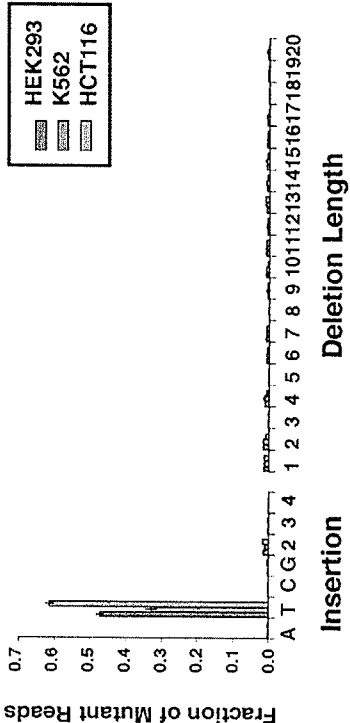
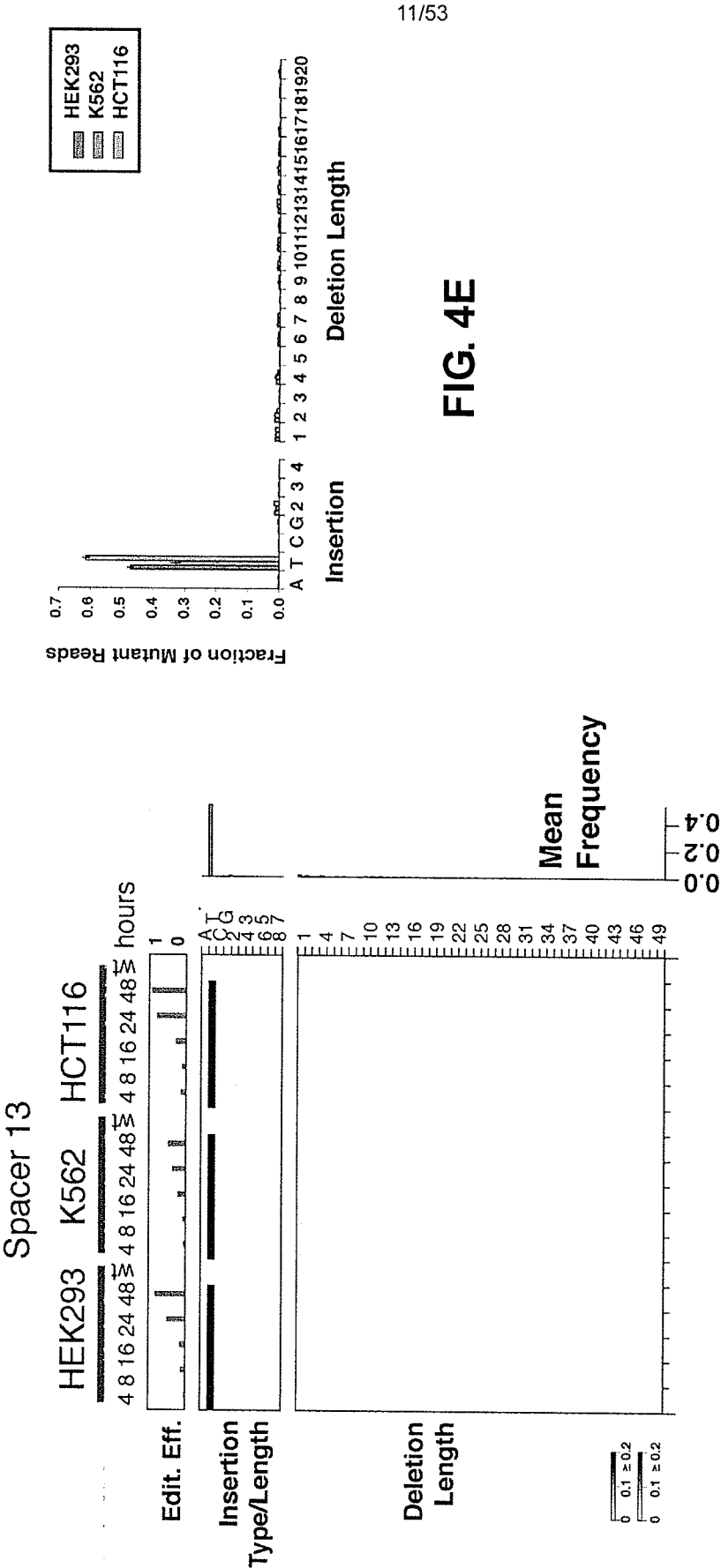


FIG. 4A





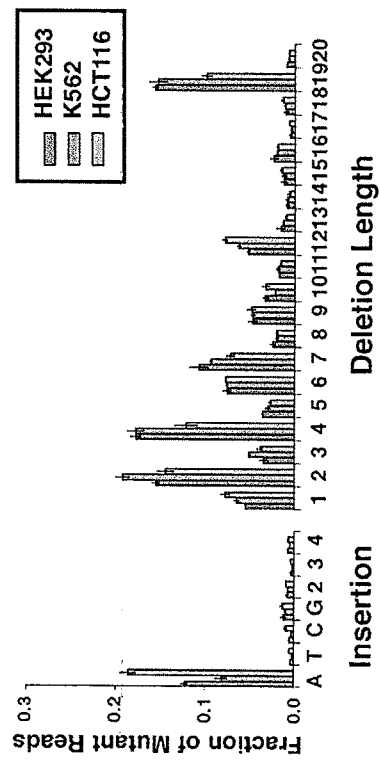


FIG. 4G

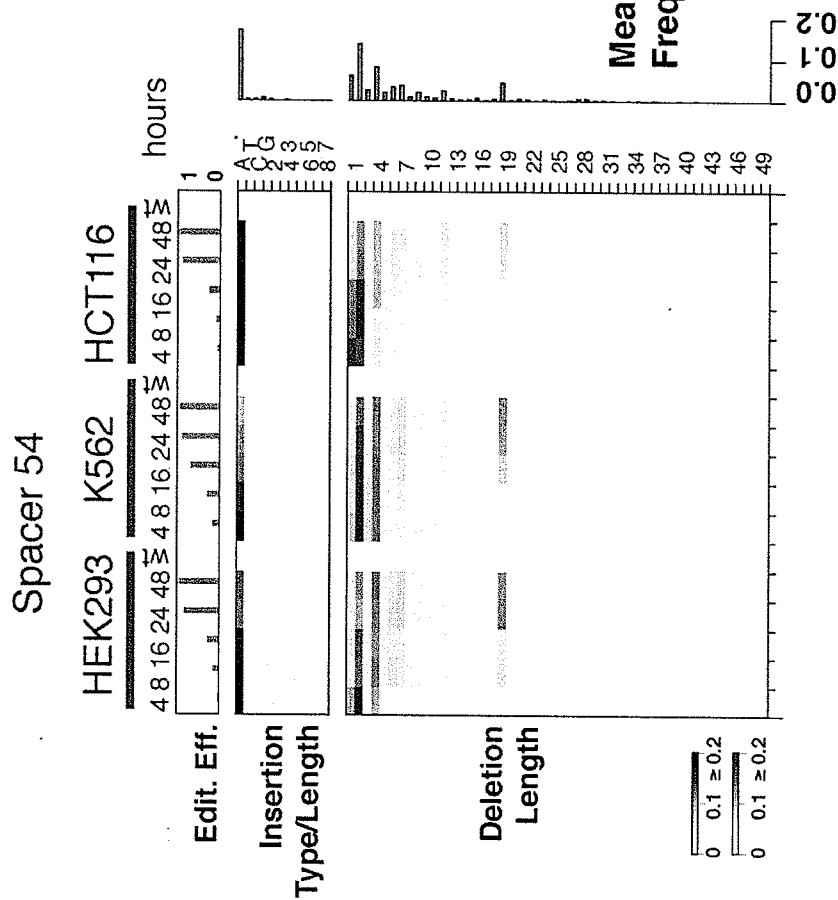


FIG. 4F

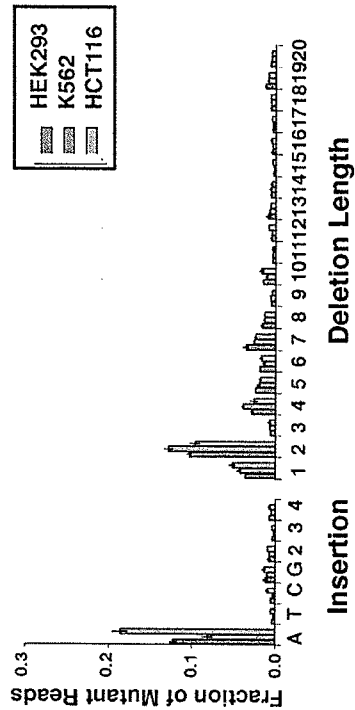


FIG. 4I

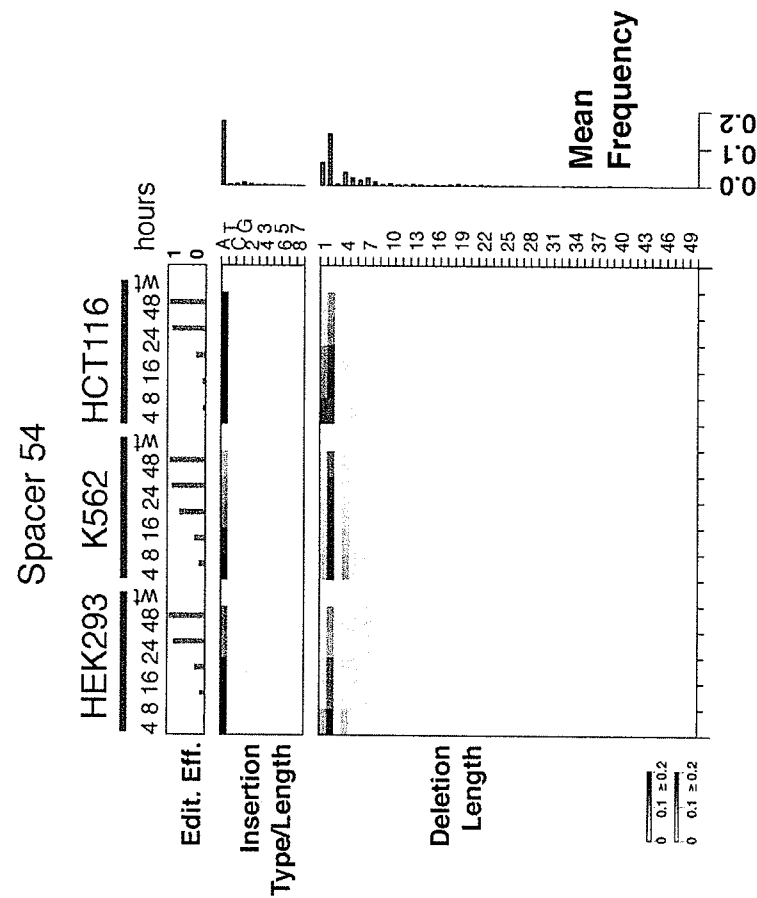


FIG. 4H

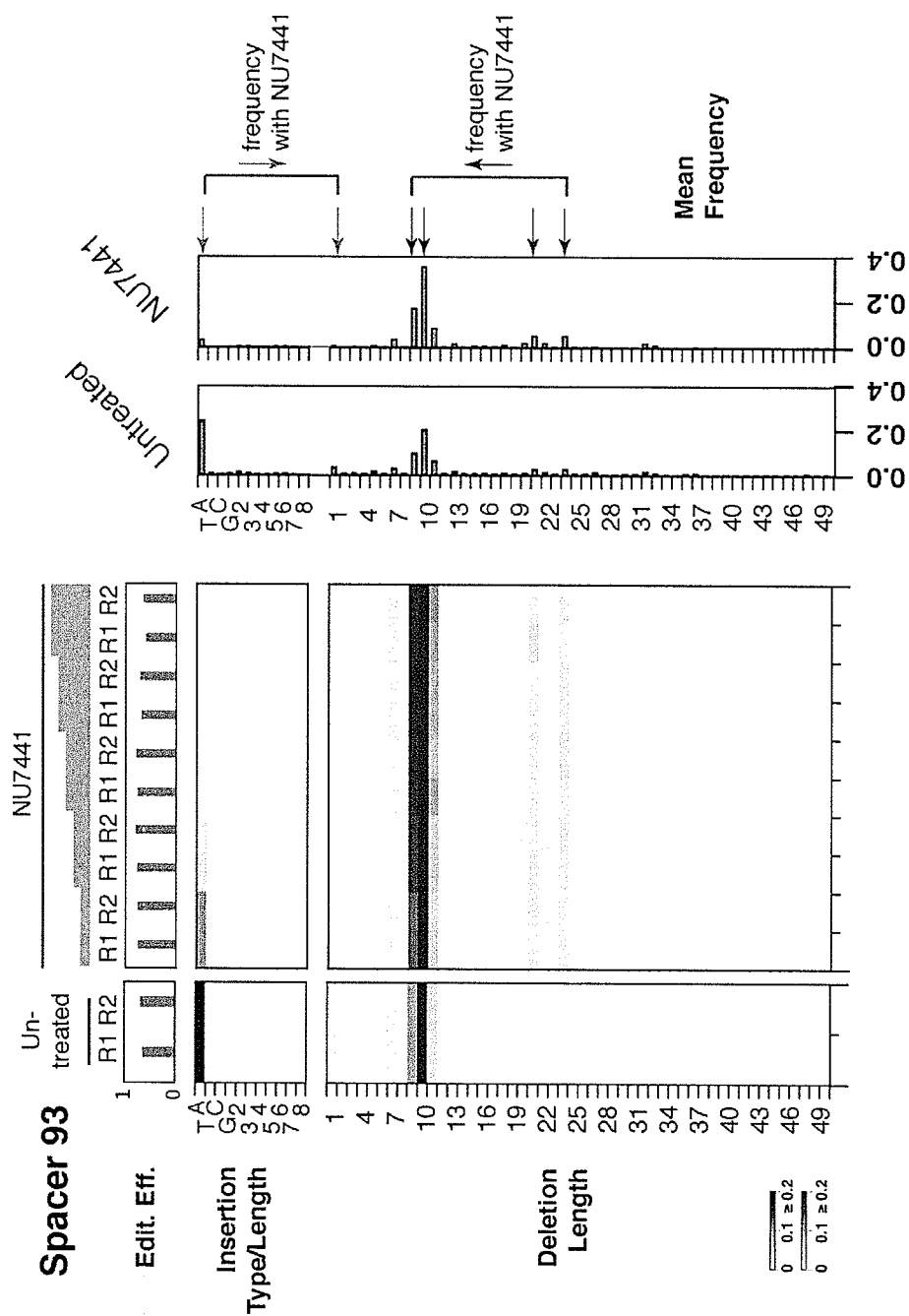


FIG. 5A

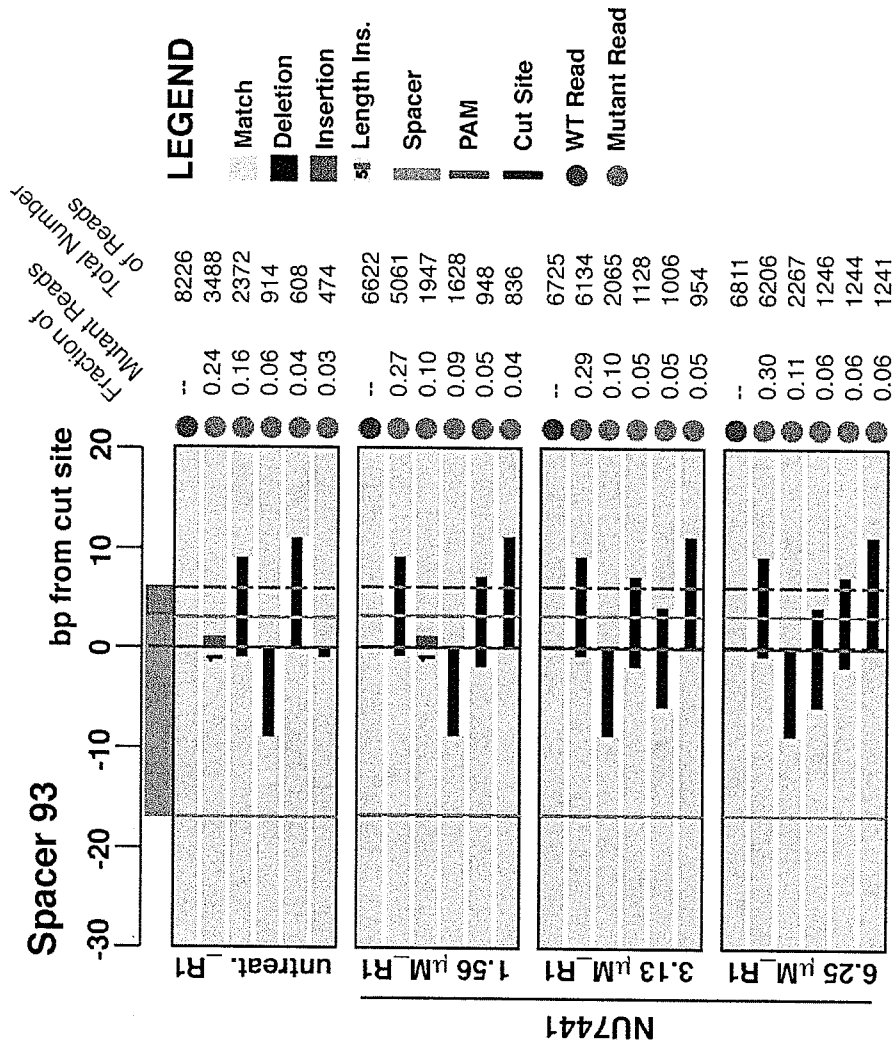


FIG. 5B

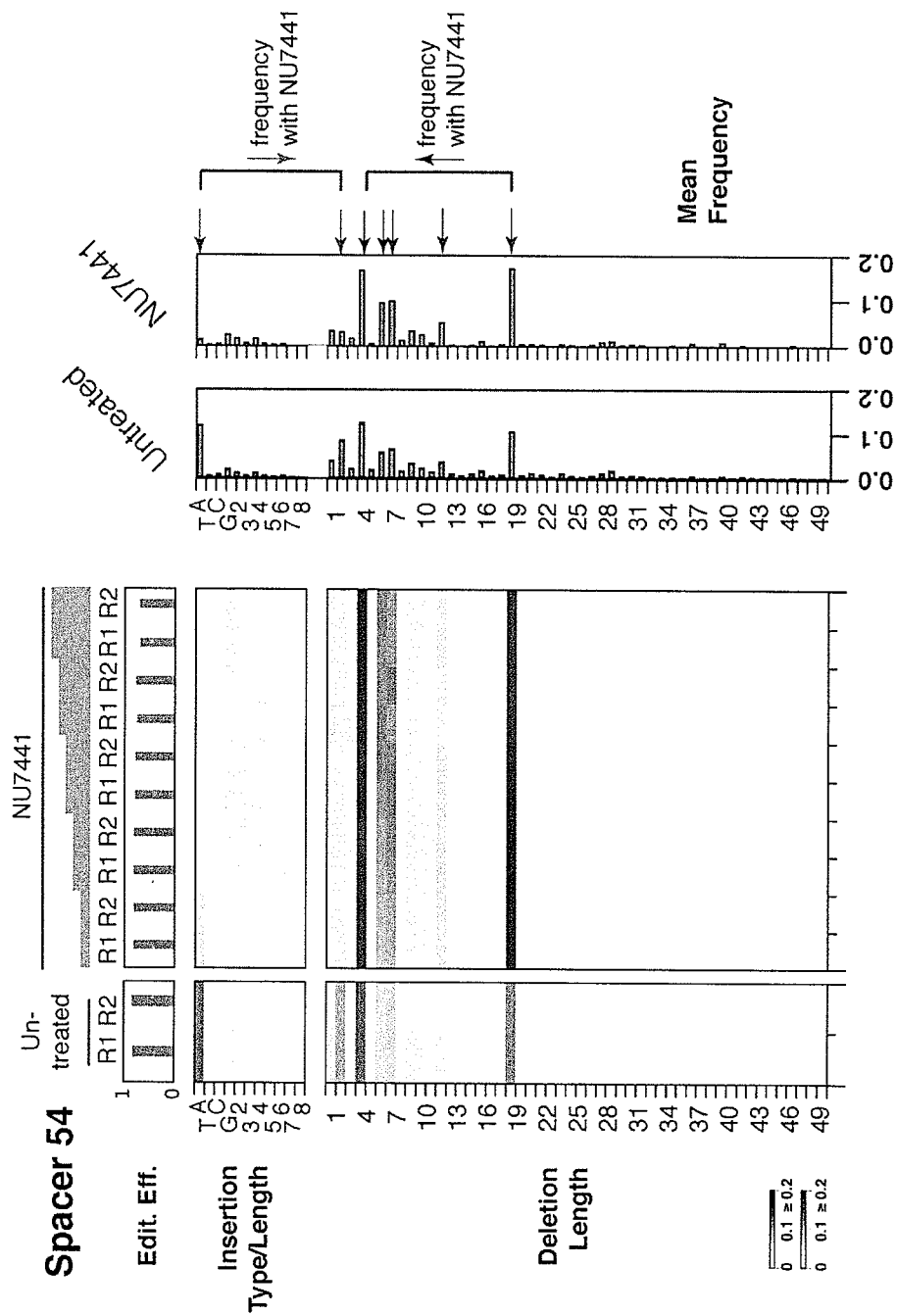


FIG. 5C

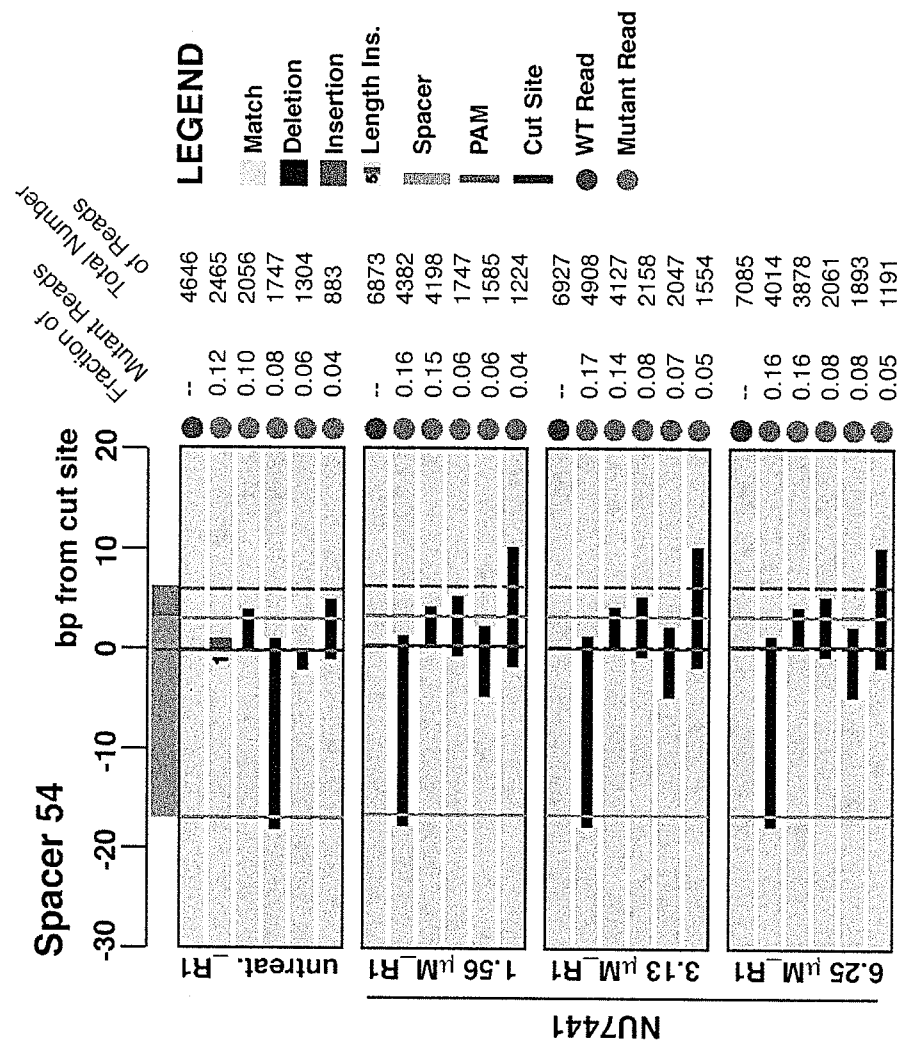


FIG. 5D

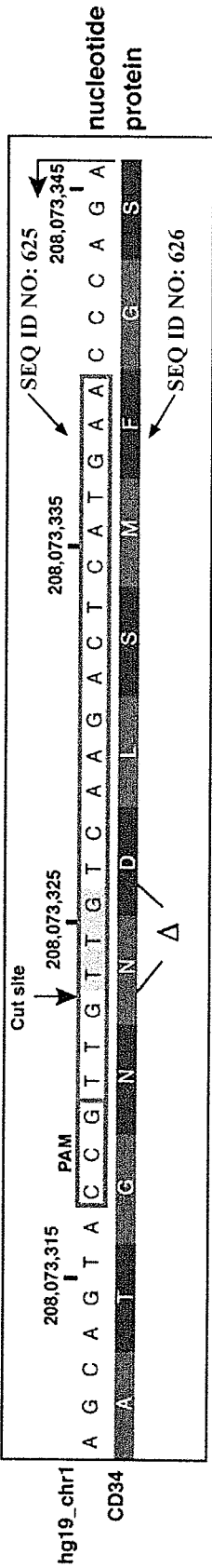


FIG. 6A

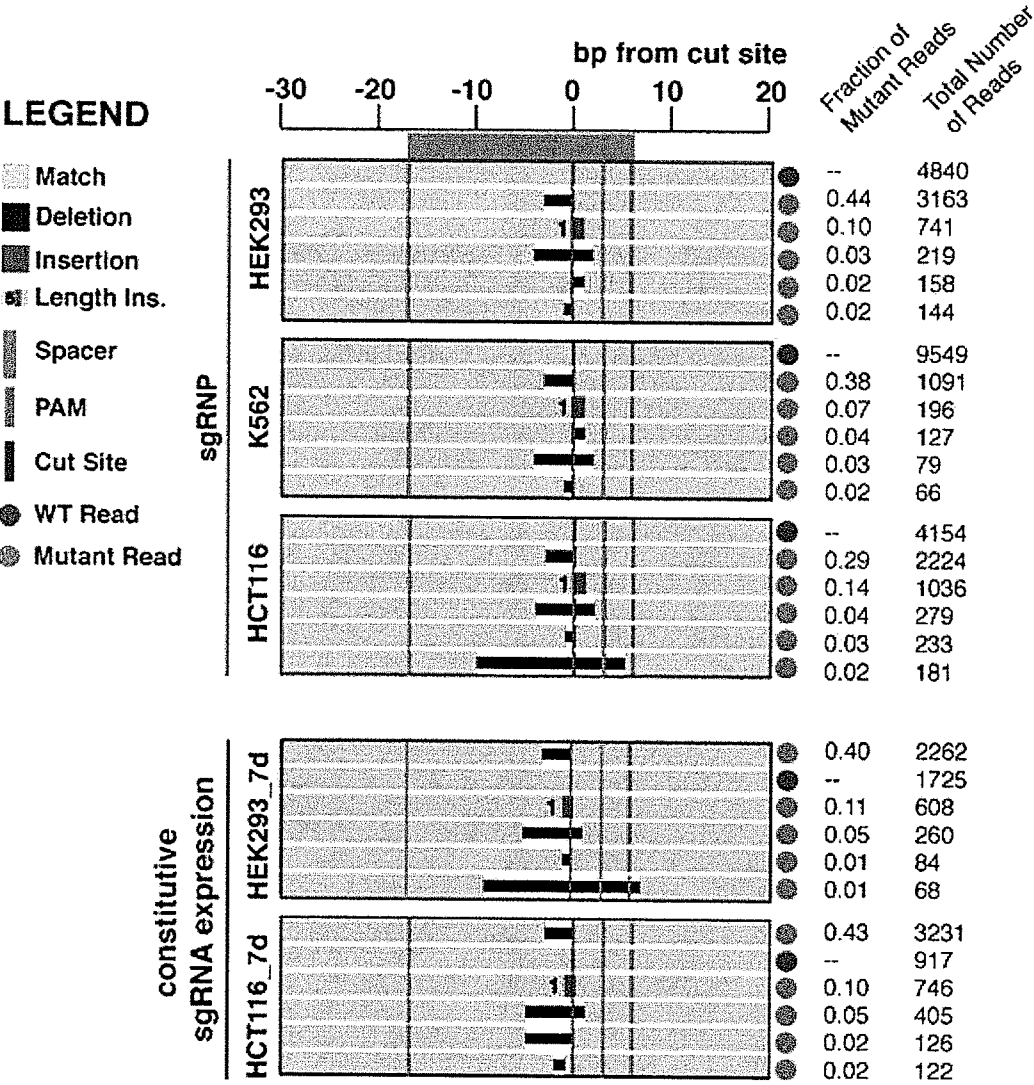


FIG. 6B

FIG. 6D

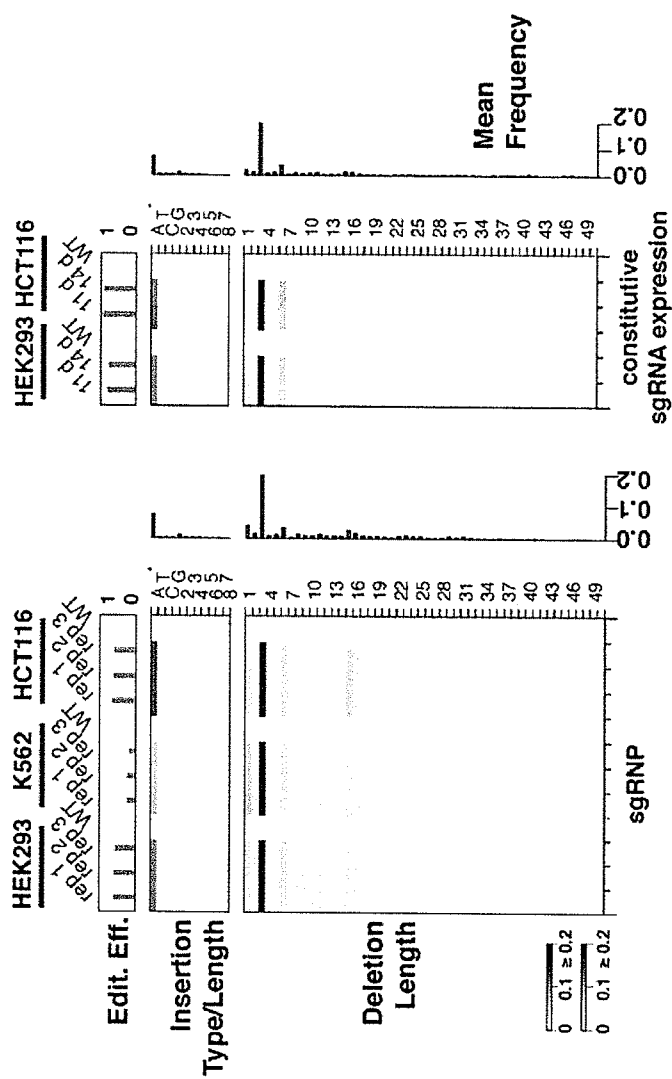


FIG. 6C

FIG. 6E

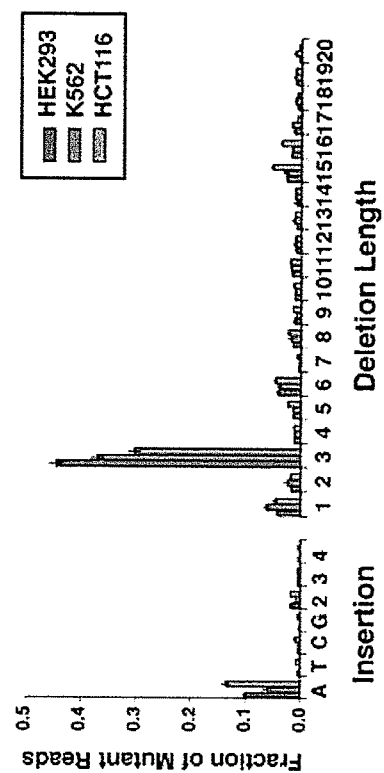




FIG. 7A

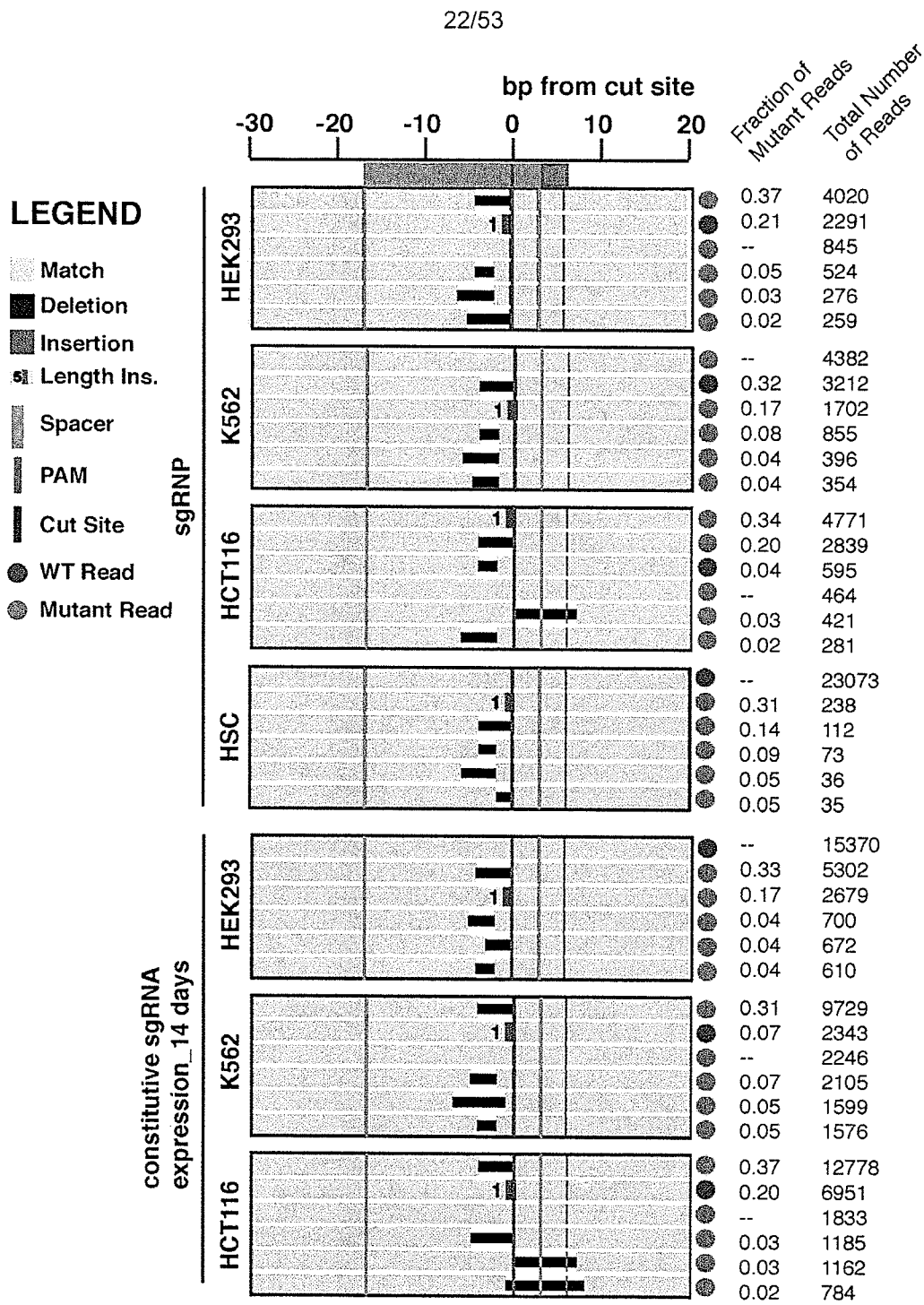


FIG. 7B

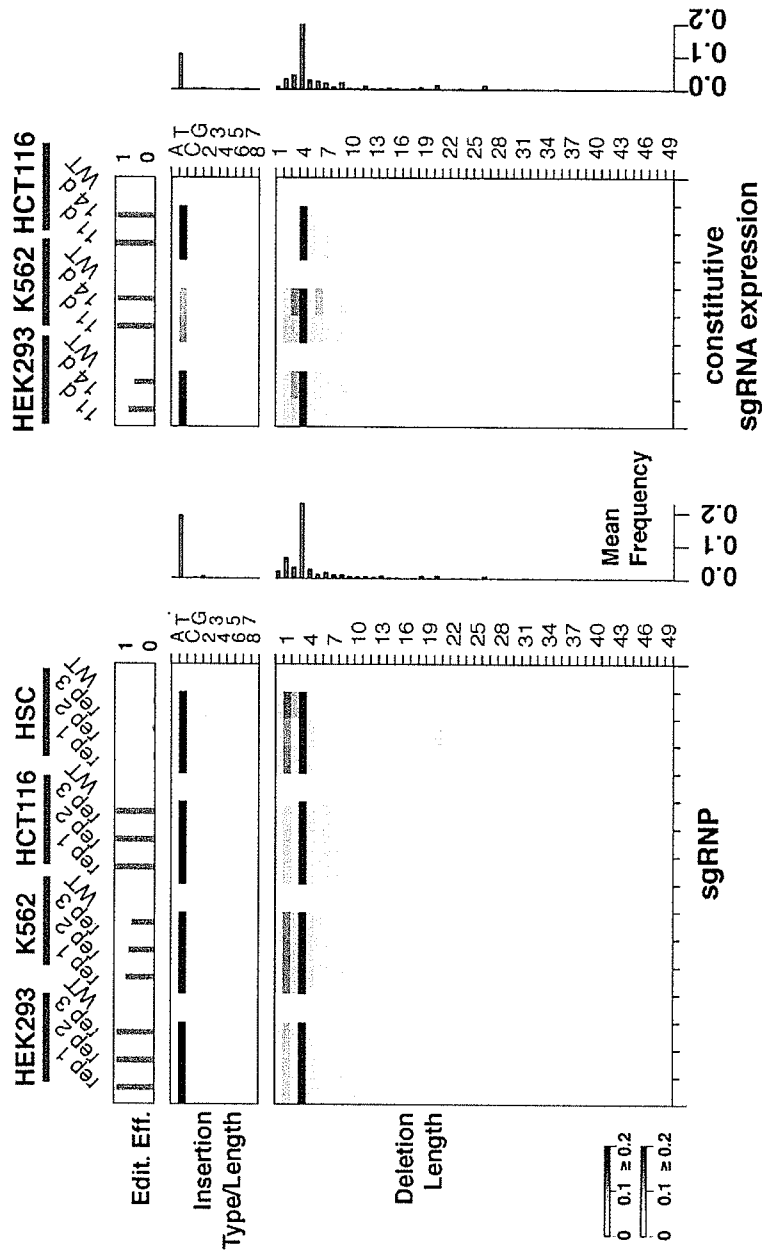


FIG. 7D

FIG. 7C

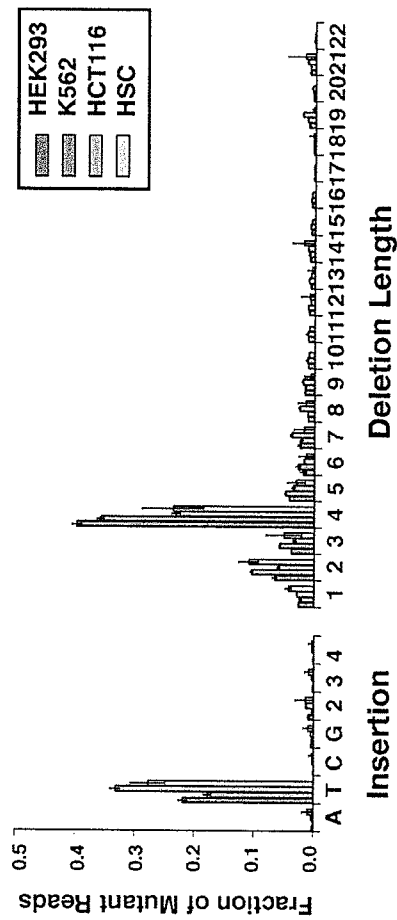


FIG. 7E

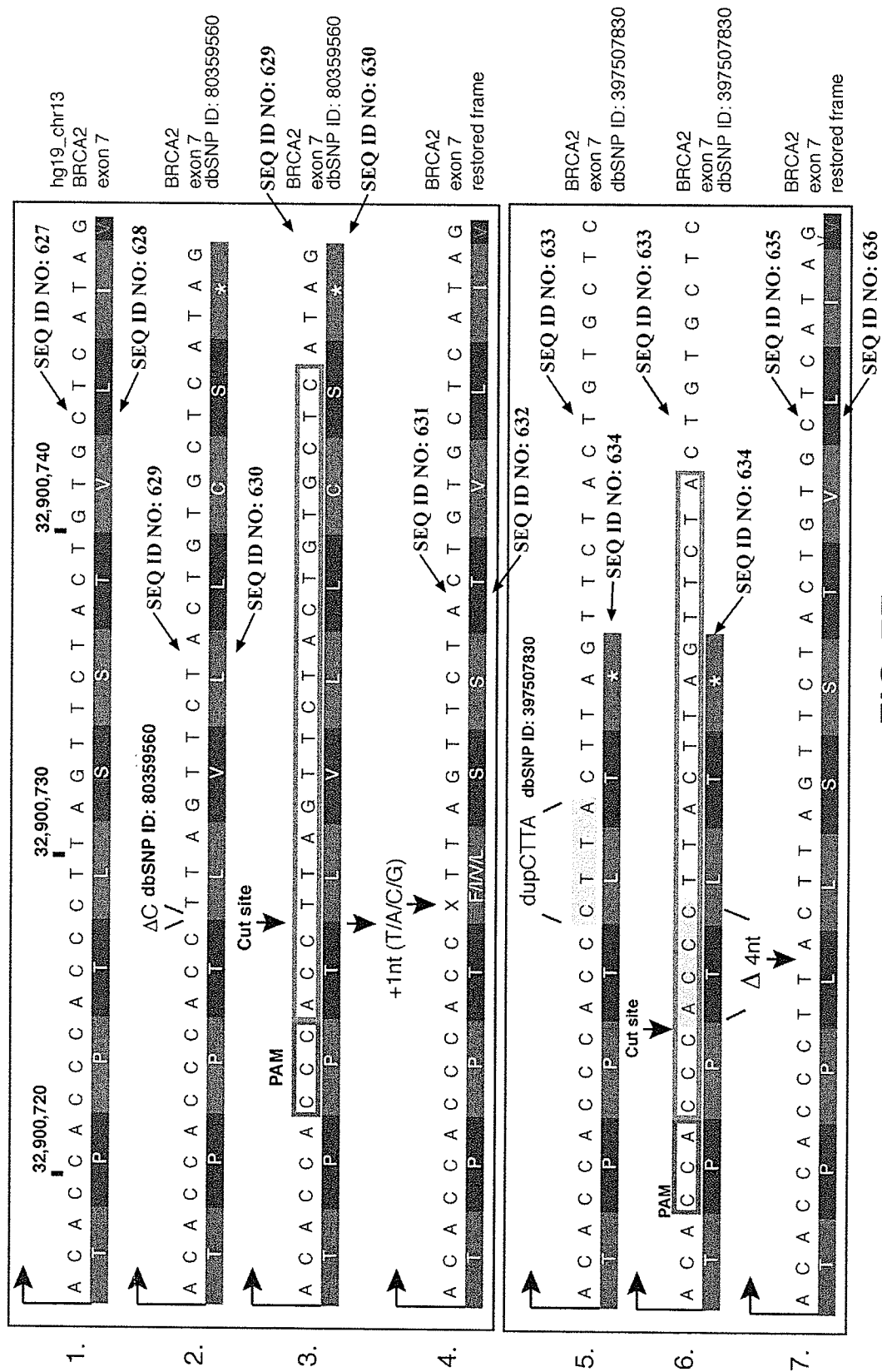


FIG. 7F

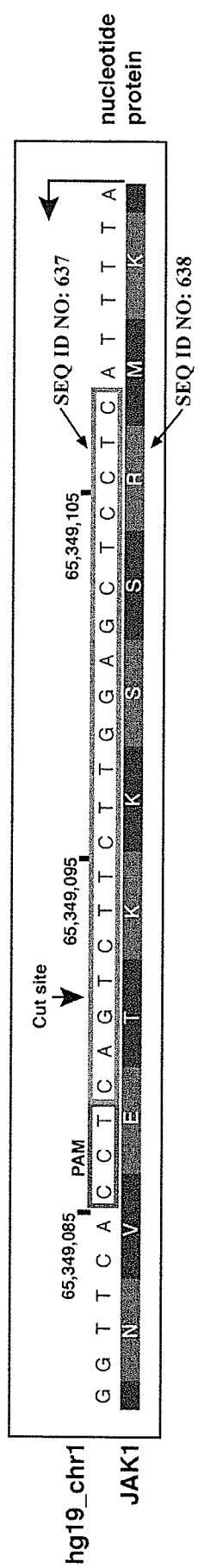


FIG. 8A

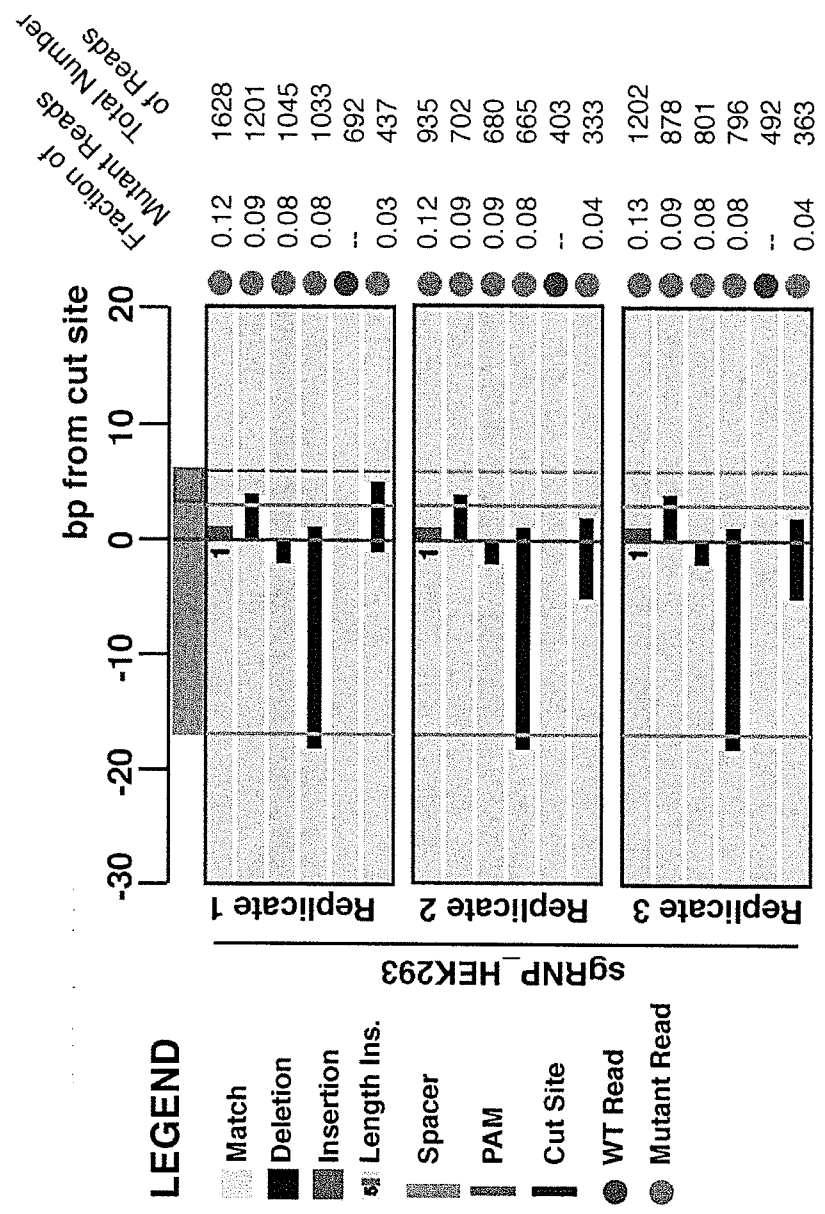


FIG. 8B

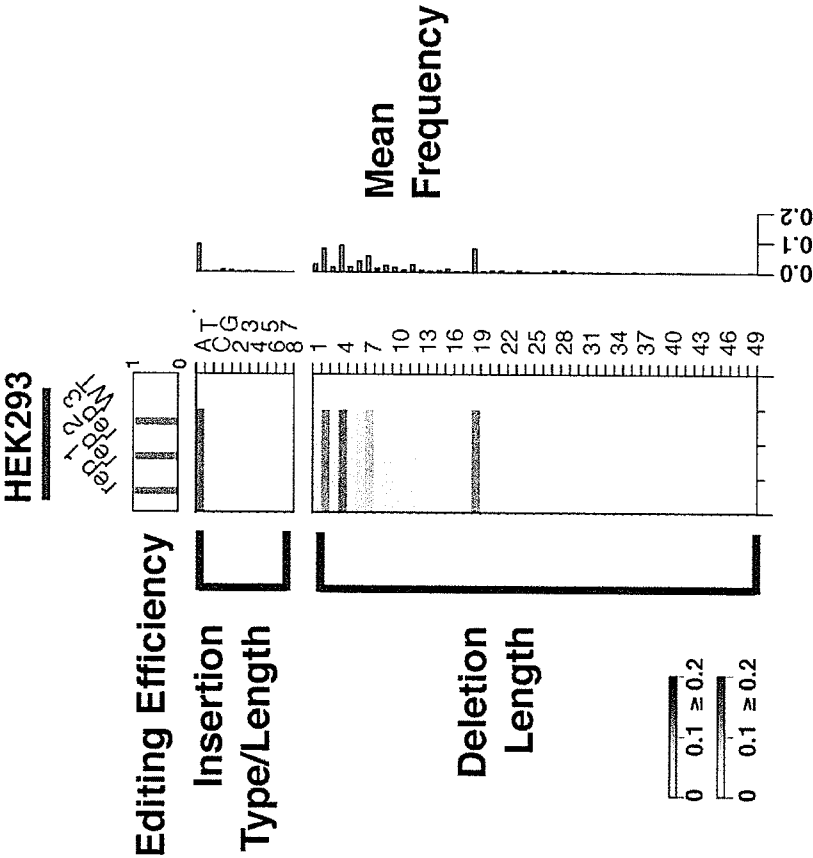


FIG. 8C

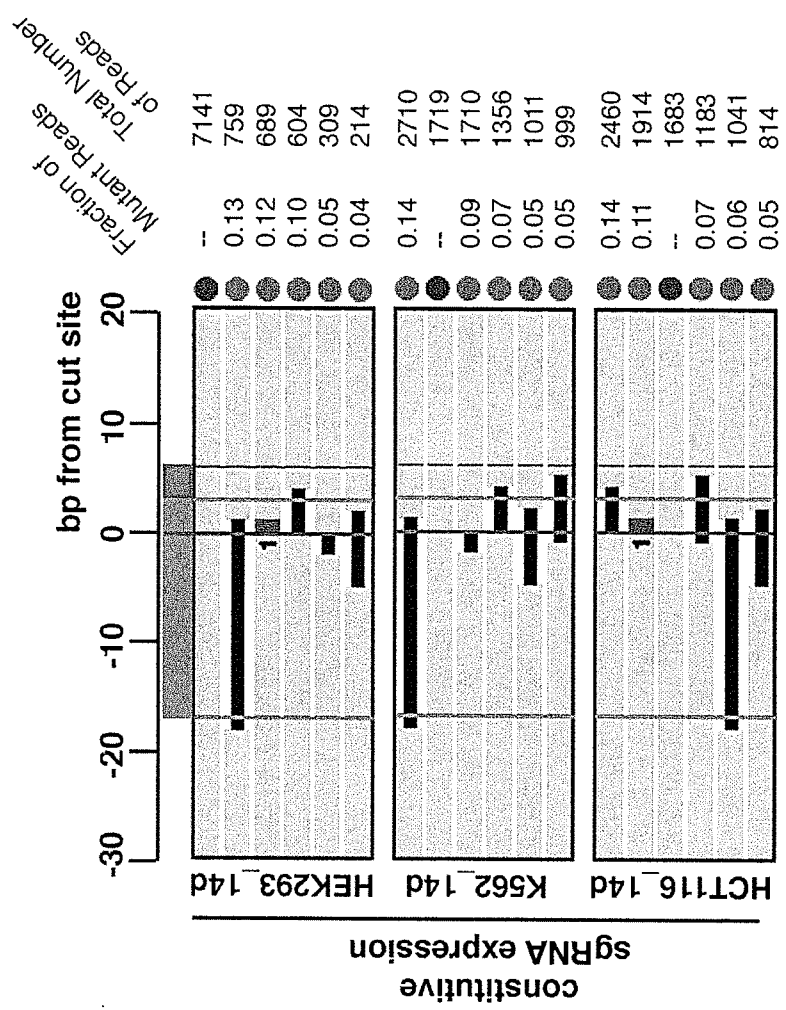


FIG. 8D

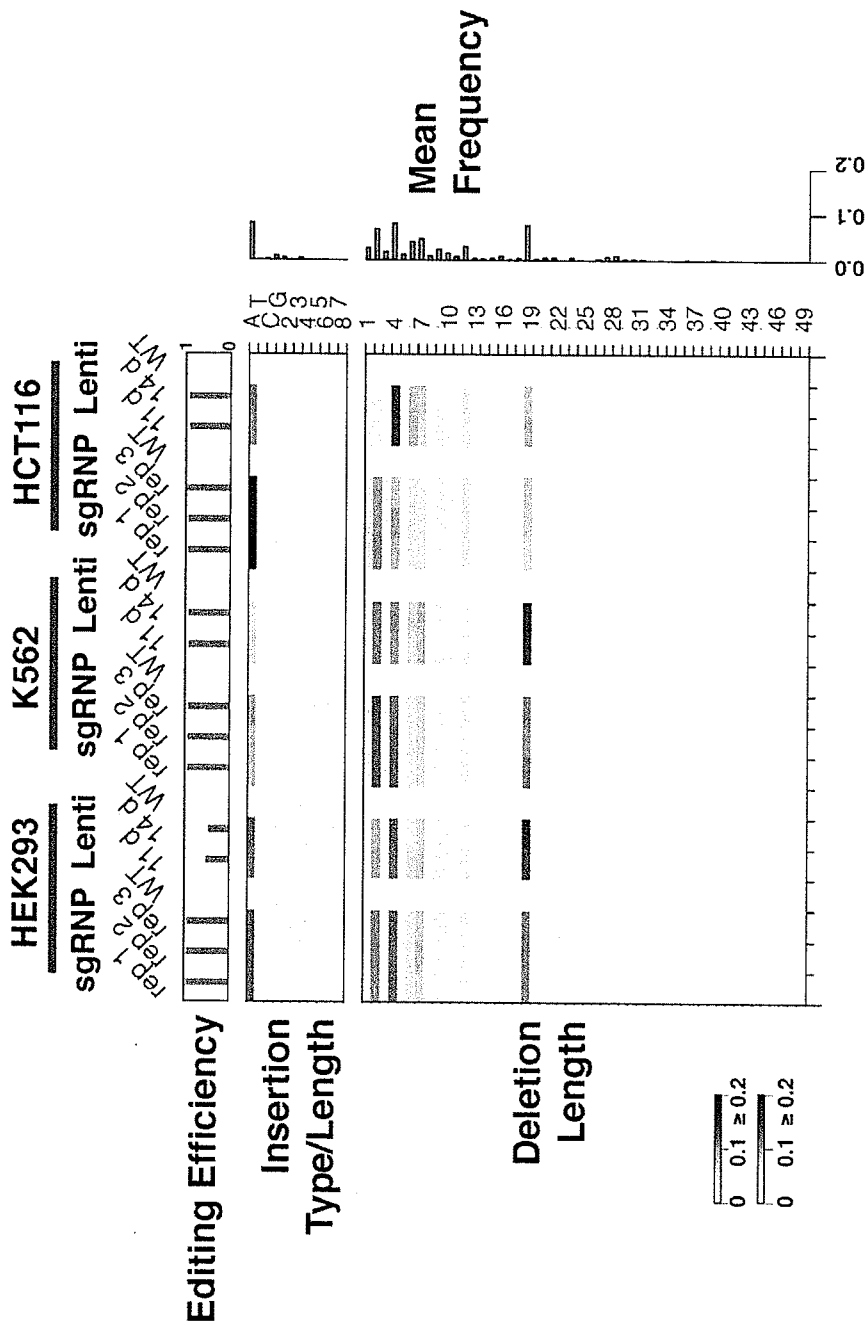


FIG. 8E

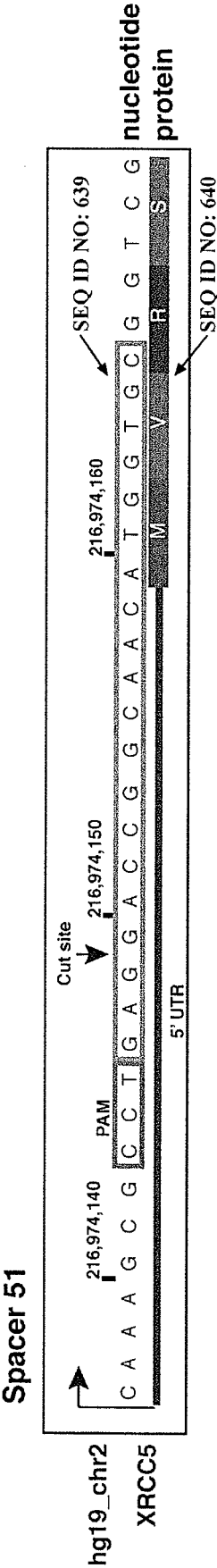


FIG. 9A

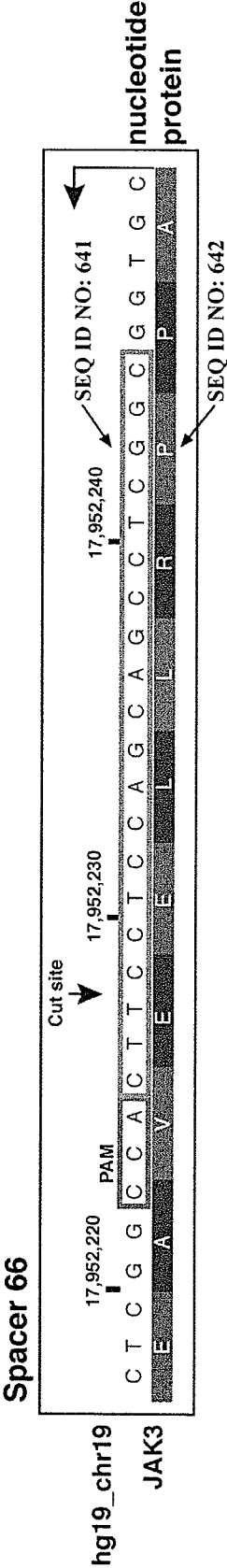


FIG. 9B

32/53

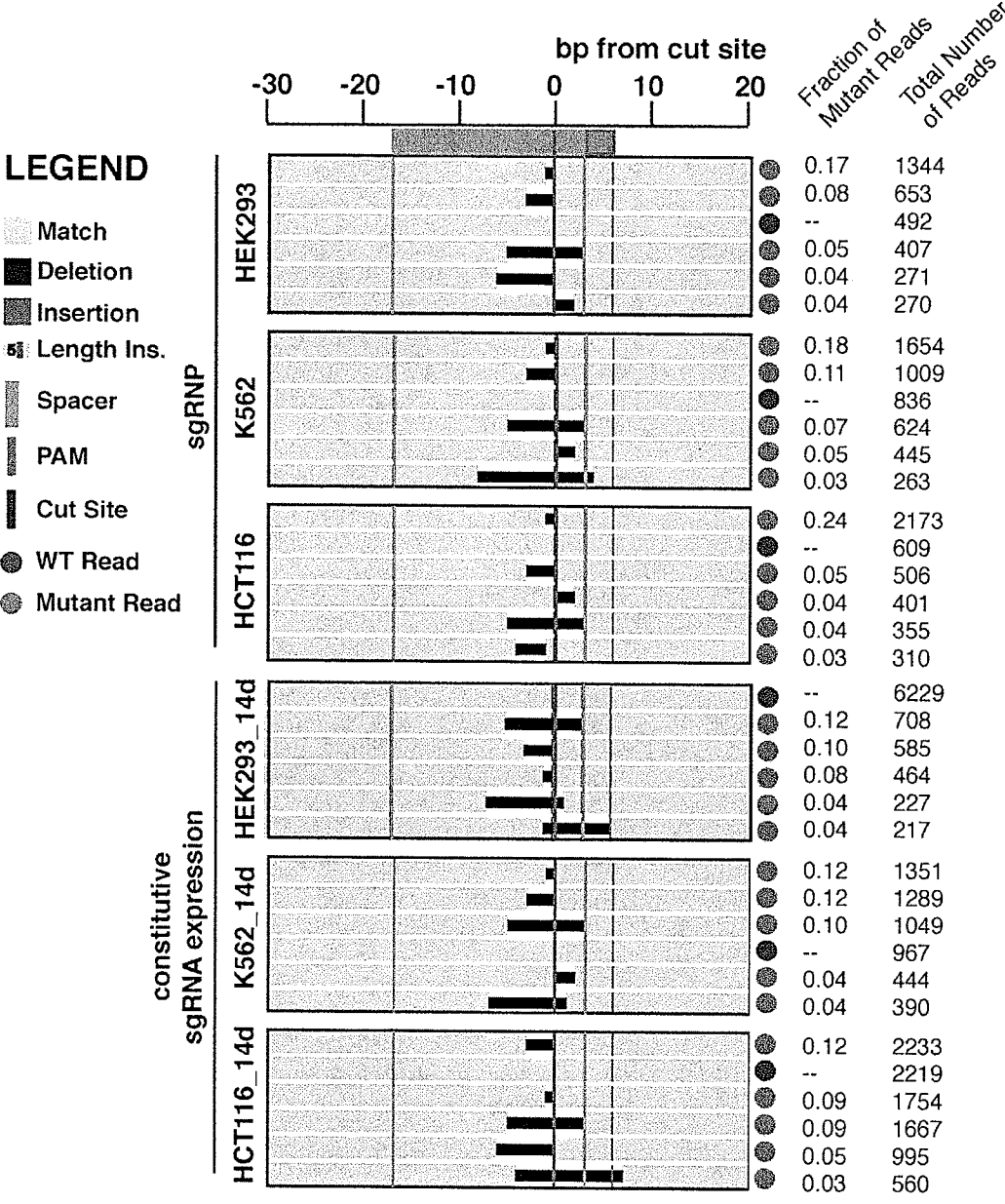


FIG. 9C

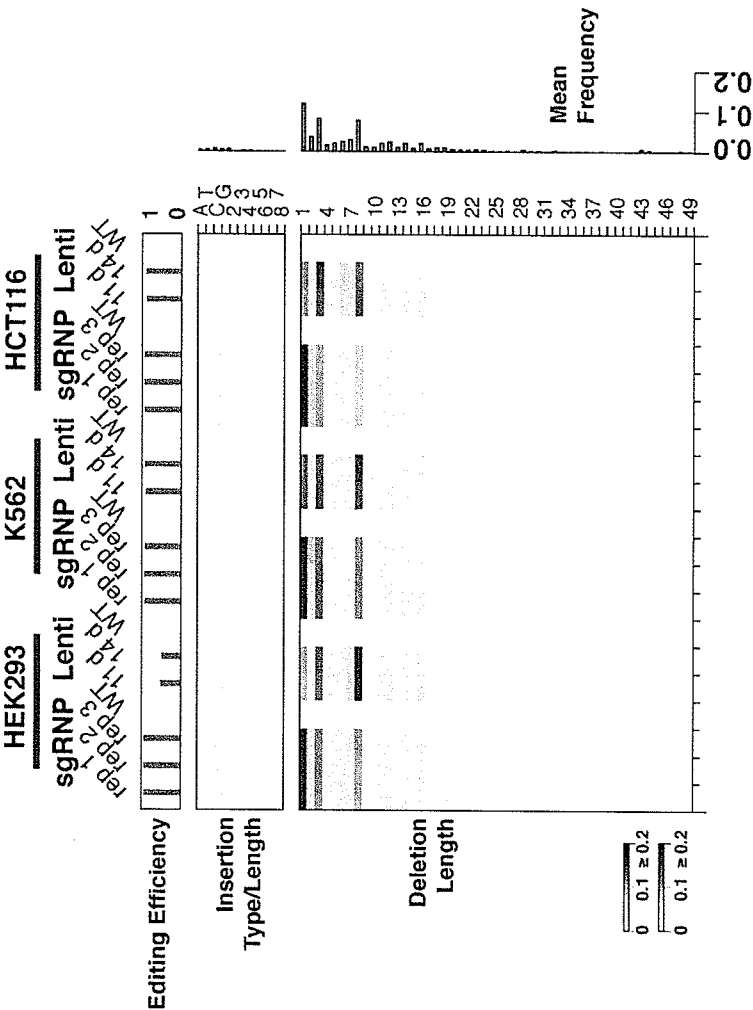


FIG. 9D

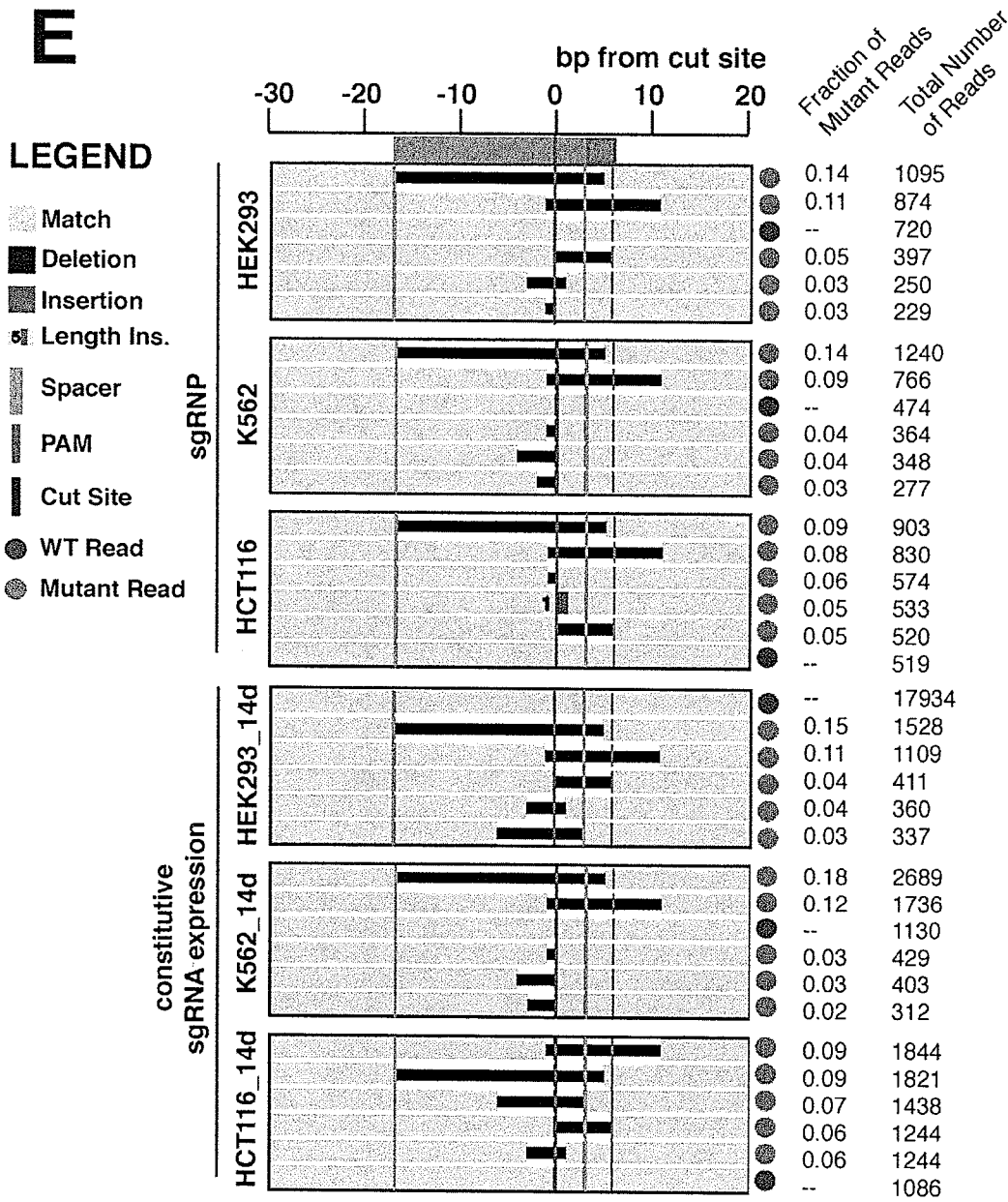


FIG. 9E

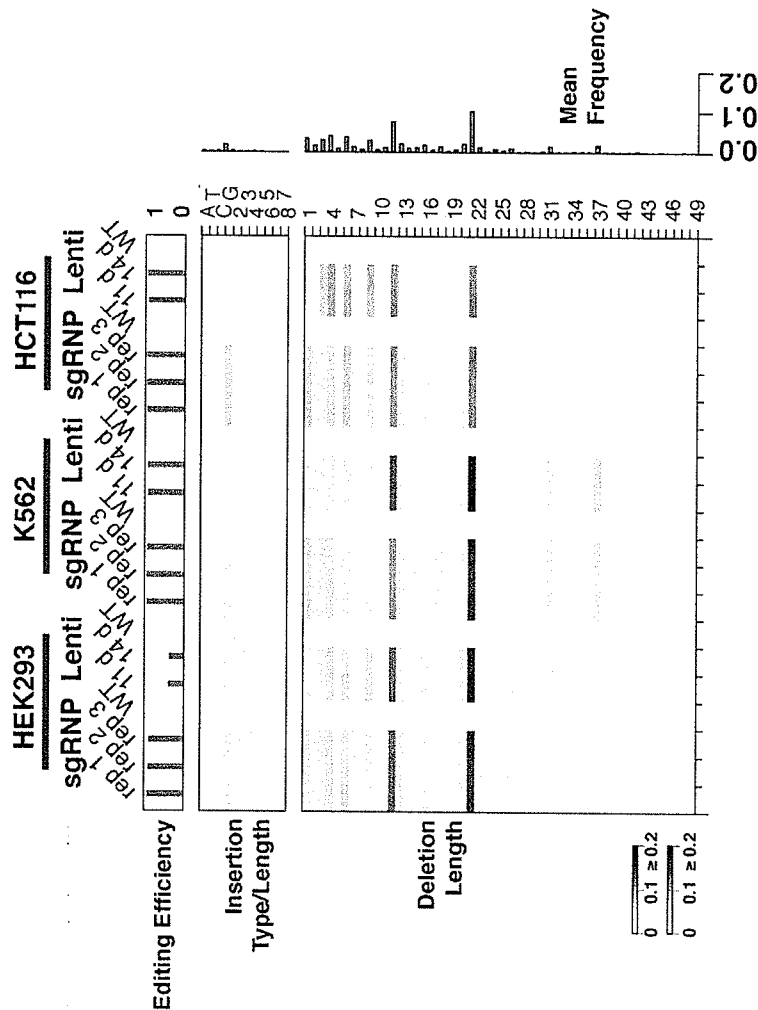


FIG. 9F

36/53

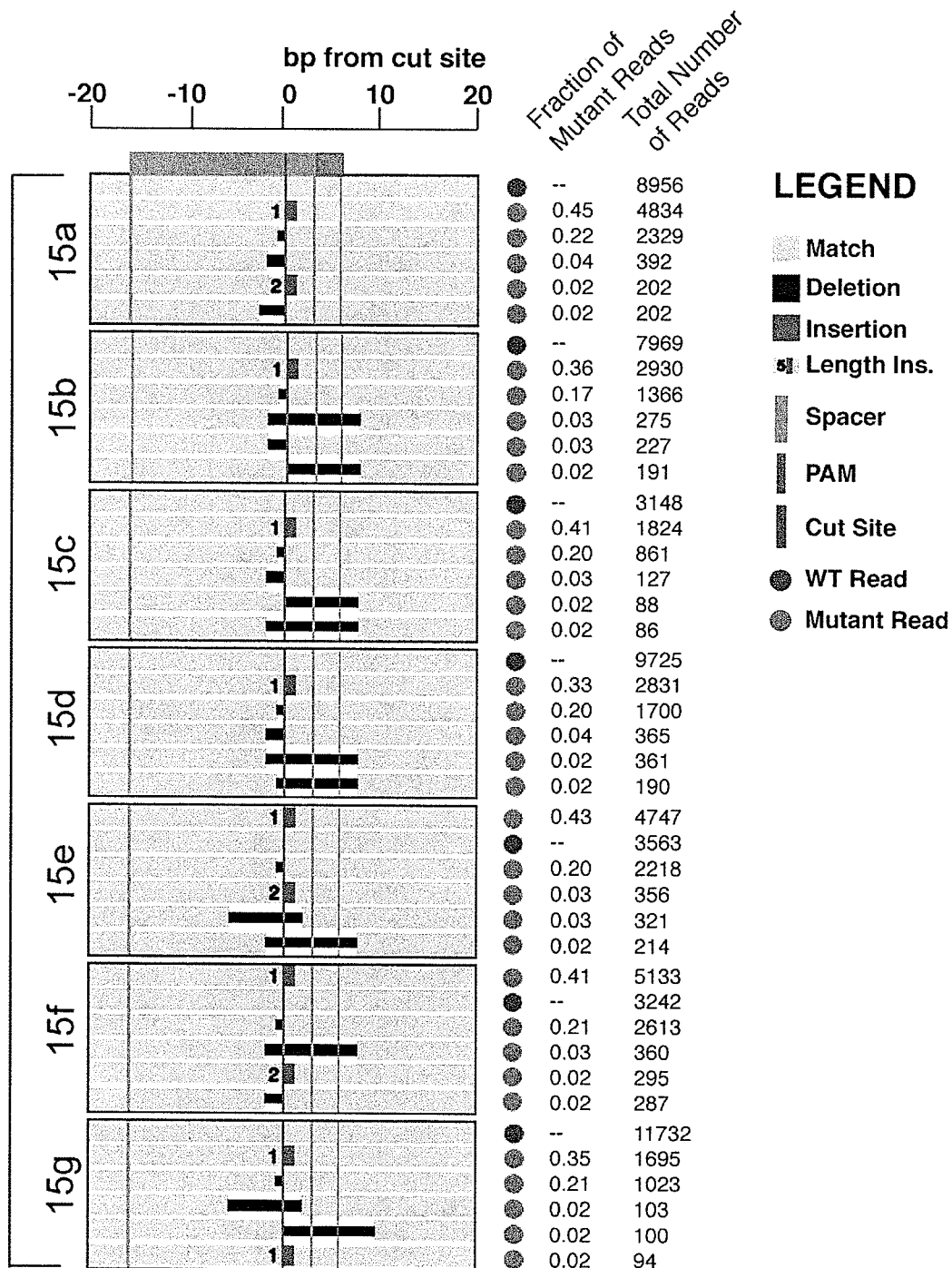


FIG. 10A

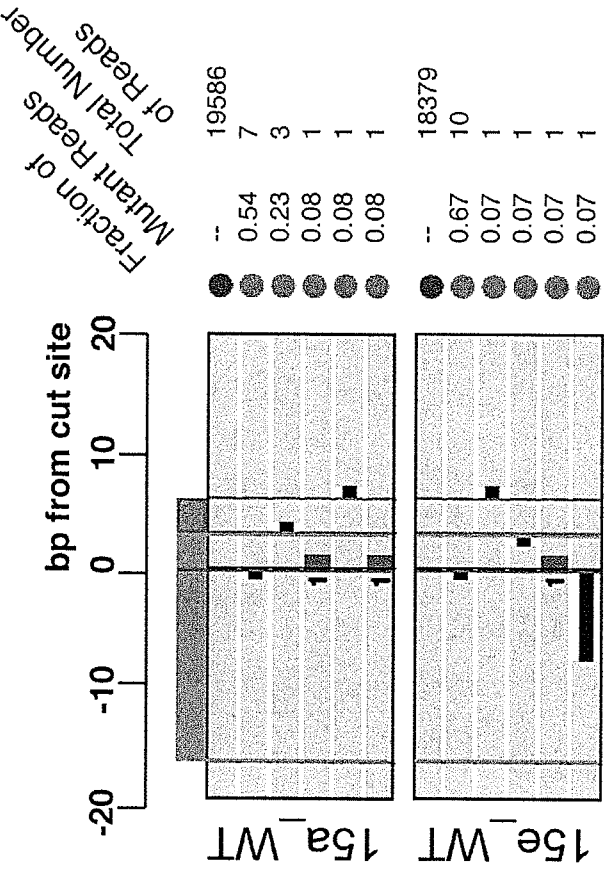


FIG. 10B

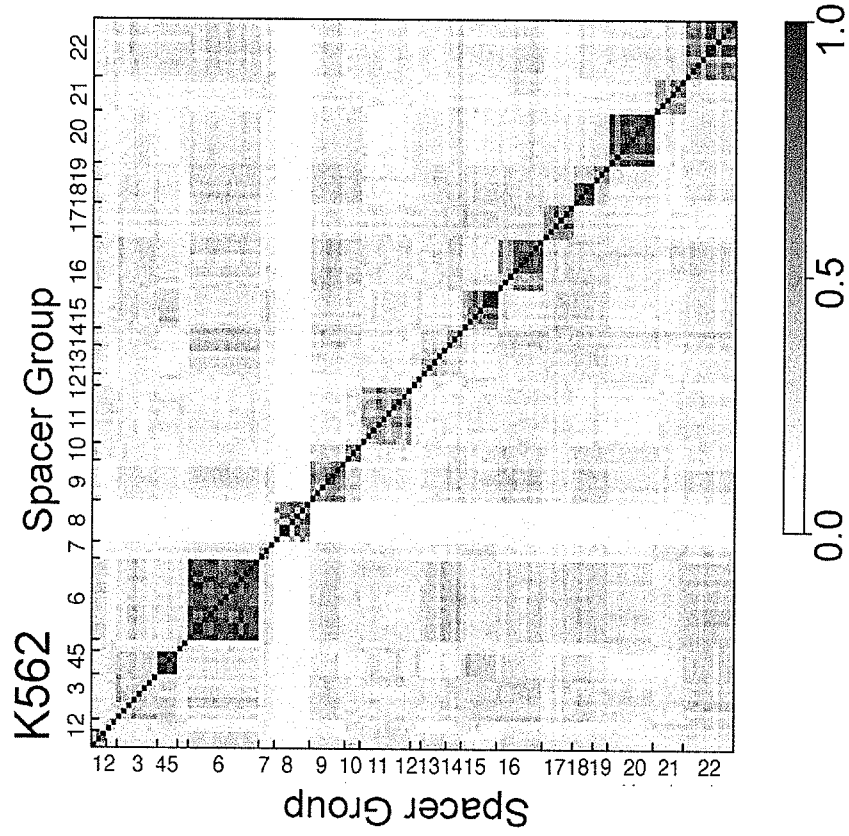


FIG. 11B

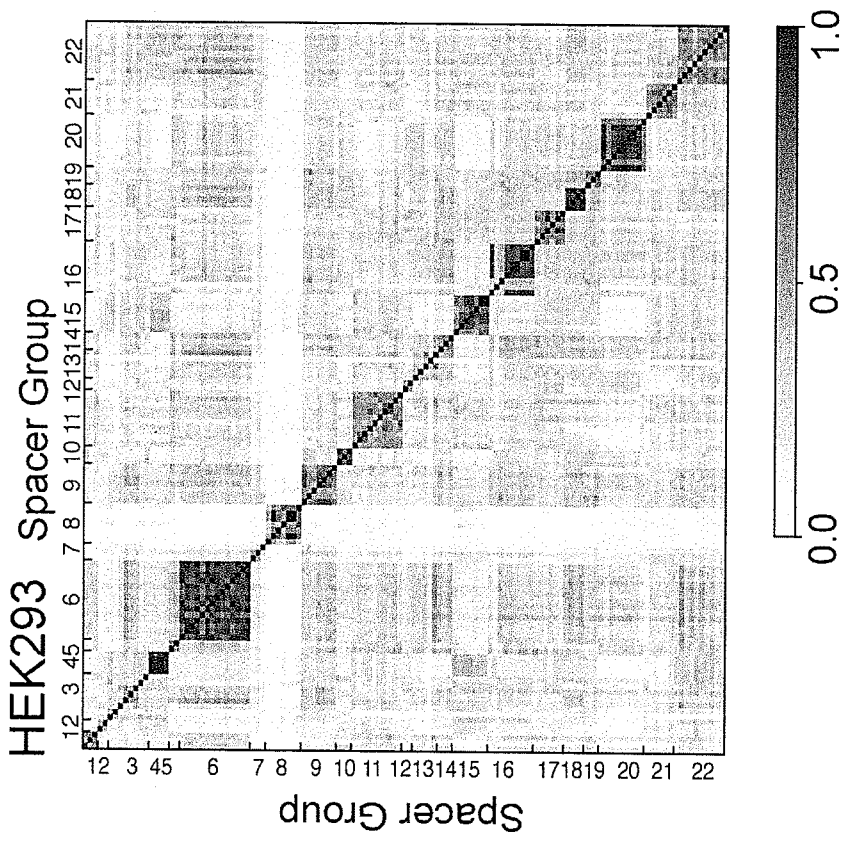


FIG. 11A

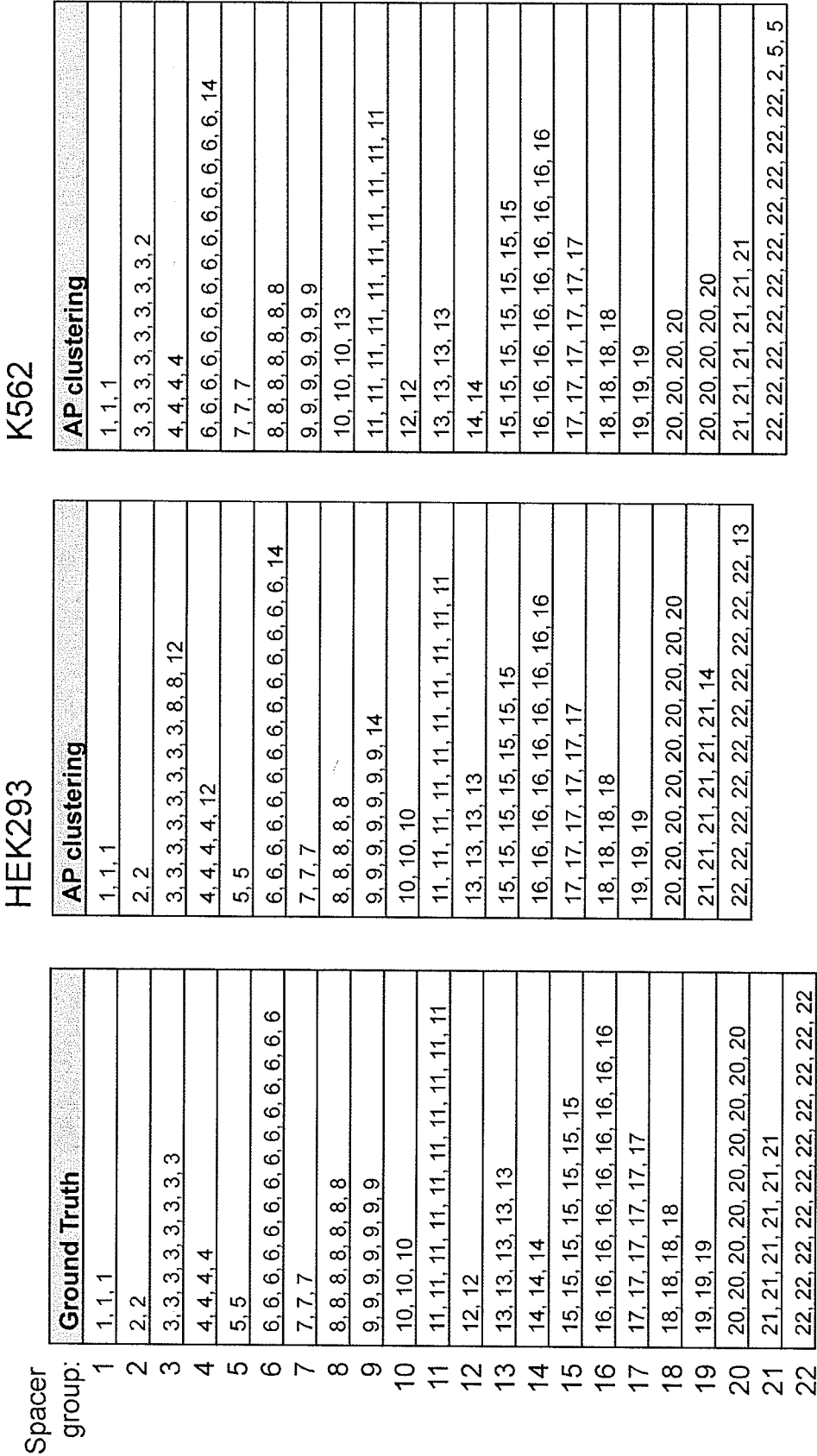


FIG. 11C

	ARI clustering of top 10 indels	ARI clustering of indels ranked by frequency
HEK293 (3 replicates)	0.870 ± 0.024	0.897 ± 0.054
K562 (3 replicates)	0.865 ± 0.057	0.893 ± 0.044

FIG. 11D

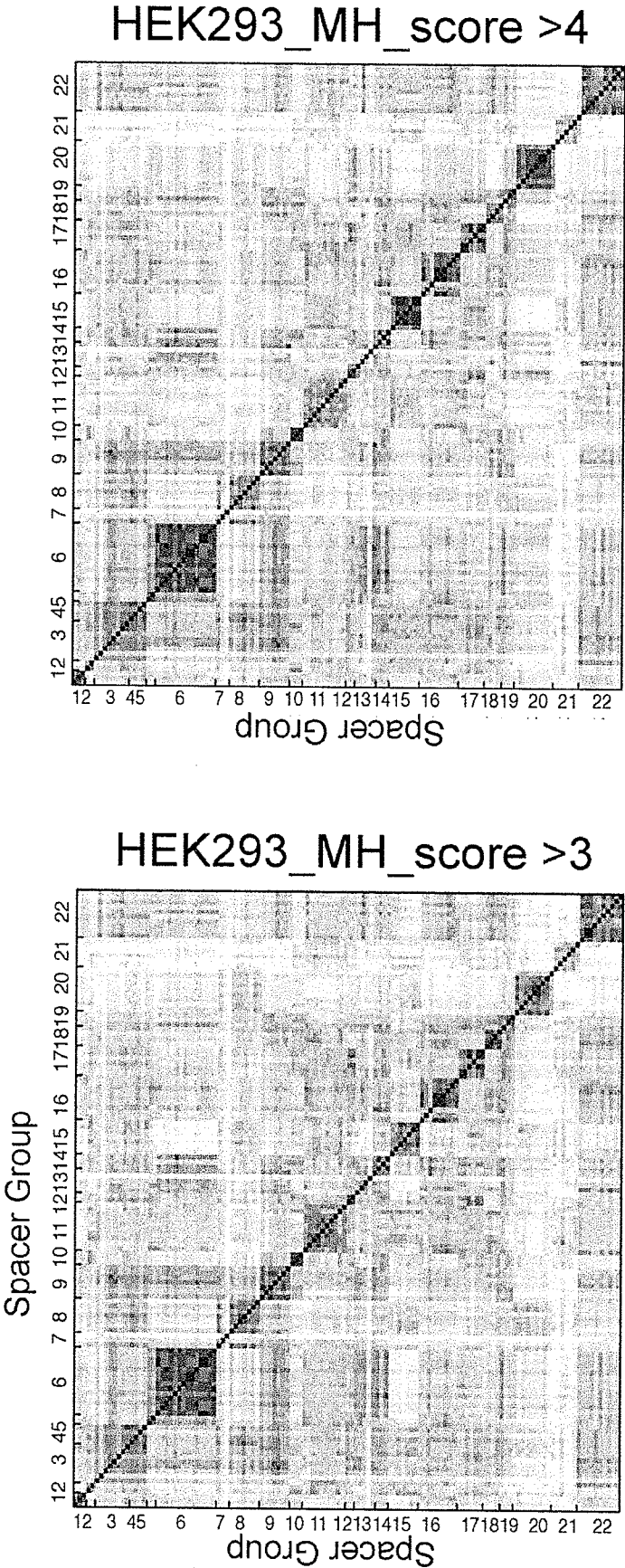


FIG. 12A

42/53

K562_MH_score >3

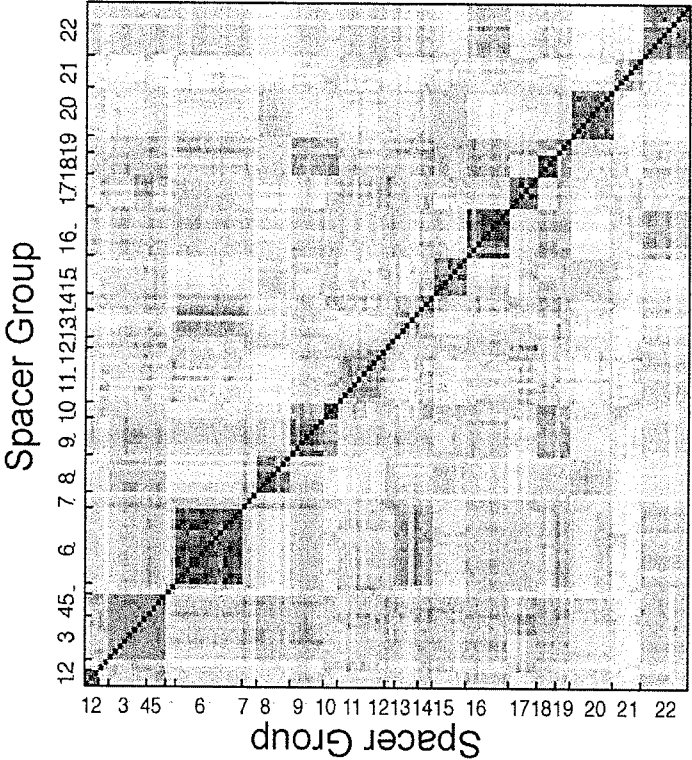


FIG. 12B

HEK293_MH_score >6

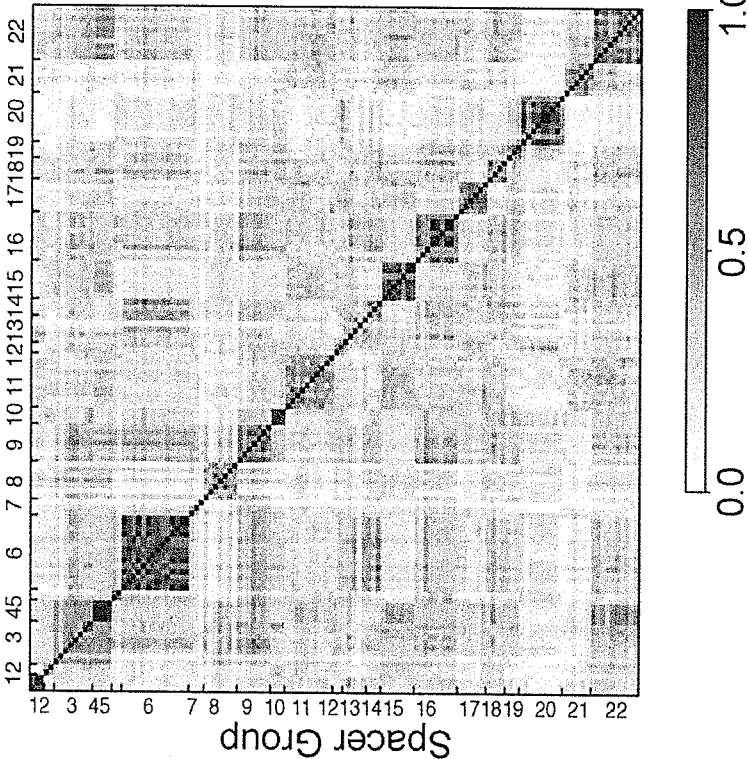


FIG. 12A cont.

43/53

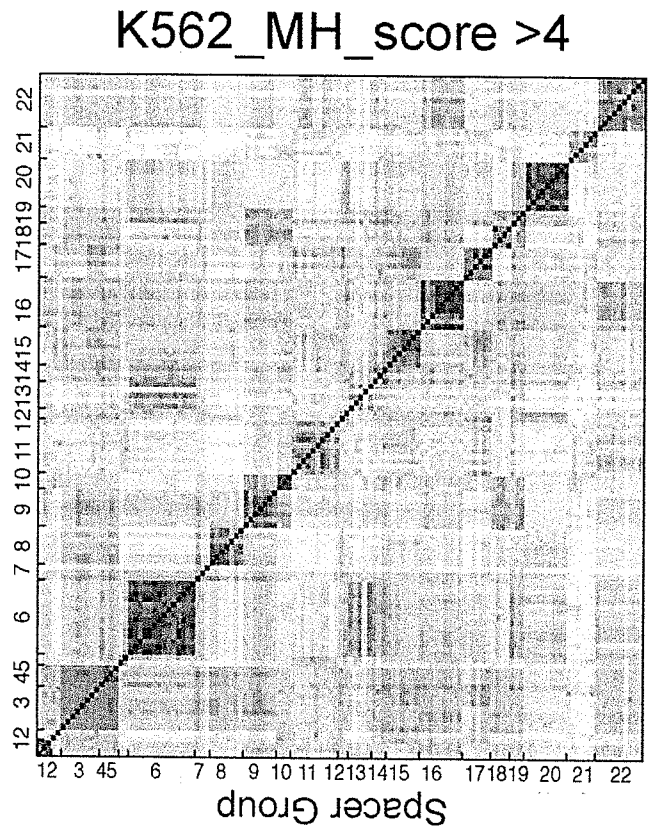
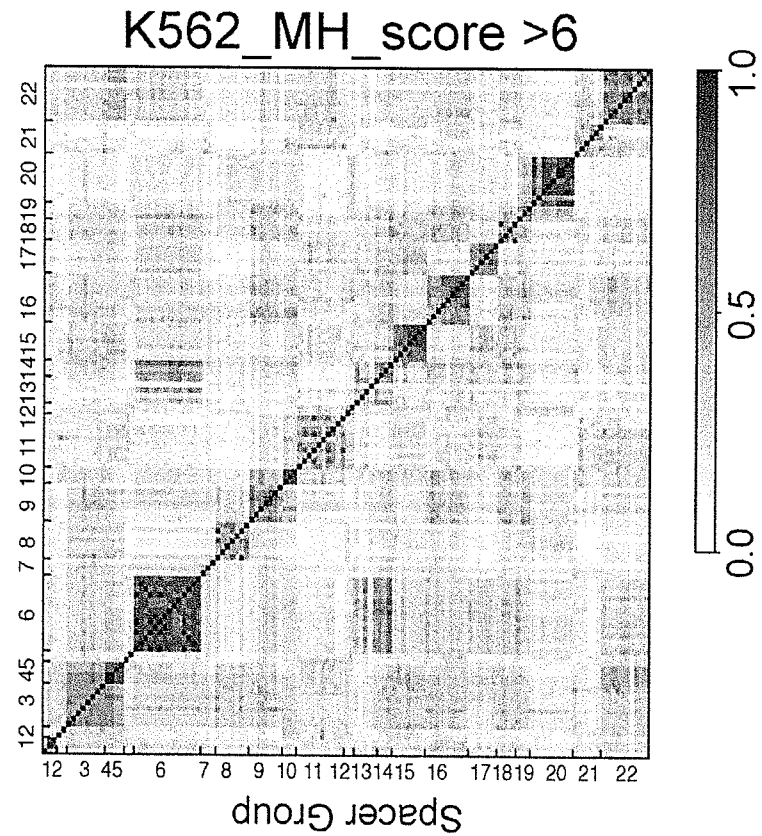


FIG. 12B cont.

	All <i>MH</i> _score deletions	Removed <i>MH</i> _score > 6 deletions	Removed <i>MH</i> _score > 4 deletions	Removed <i>MH</i> _score > 3 deletions
HEK293 (3 replicates)	0.870 ± 0.024	0.704 ± 0.039	0.674 ± 0.012	0.661 ± 0.041
K562 (3 replicates)	0.865 ± 0.057	0.705 ± 0.057	0.723 ± 0.018	0.743 ± 0.021

FIG. 12C

<i>MH</i> _score cutoff	Microhomology allowed (IUPAC symbols*)
6	N, NN, WWW
4	N, WW
3	N

*IUPAC symbols

N: aNy (A, T, C, G)
W: Weak (A, T)
S: Strong (C, G)

FIG. 12D

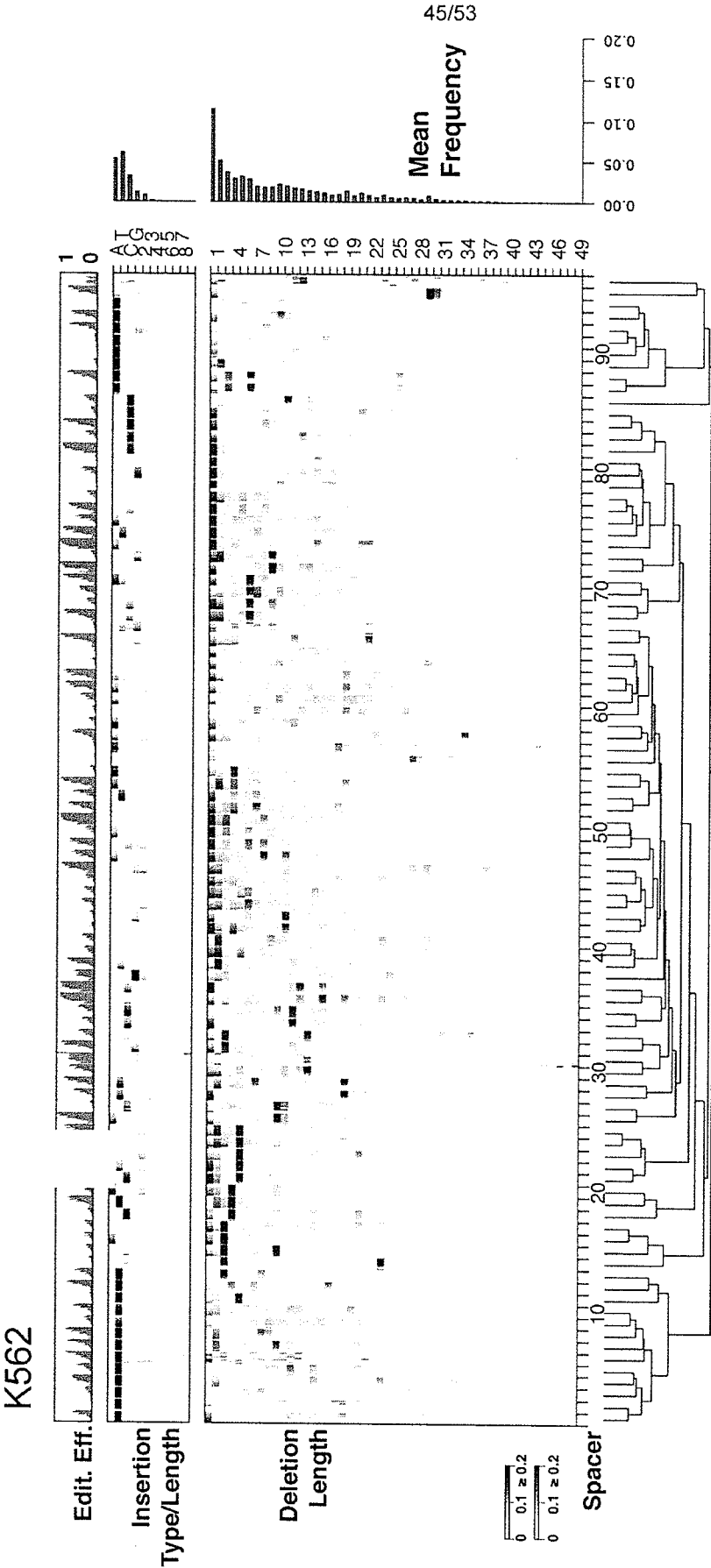


FIG. 13A

HCT116

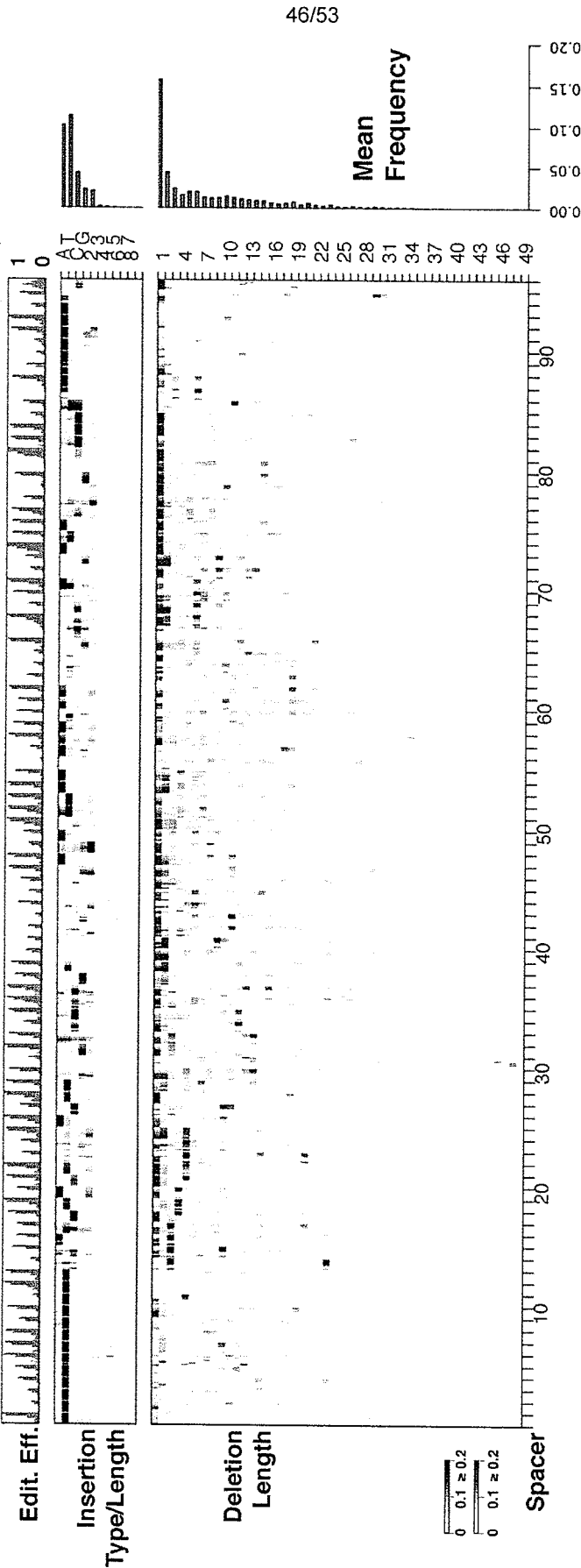
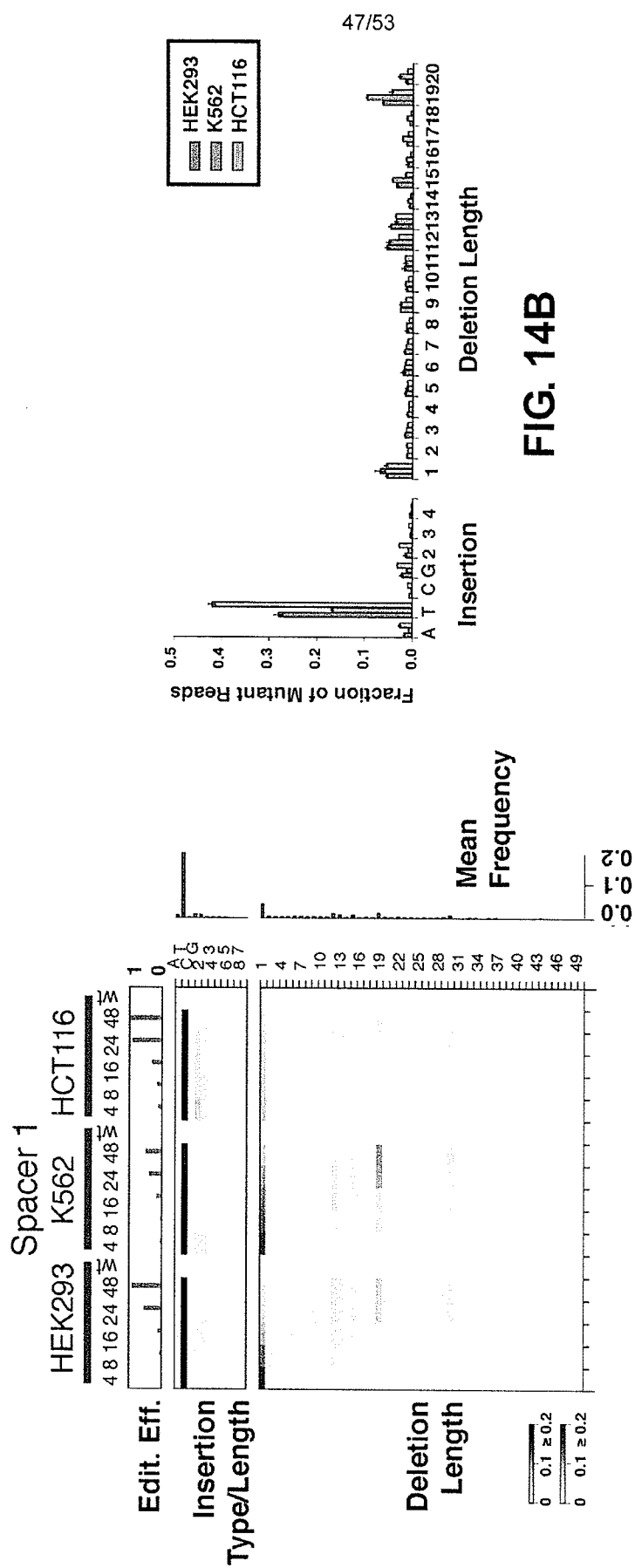


FIG. 13B



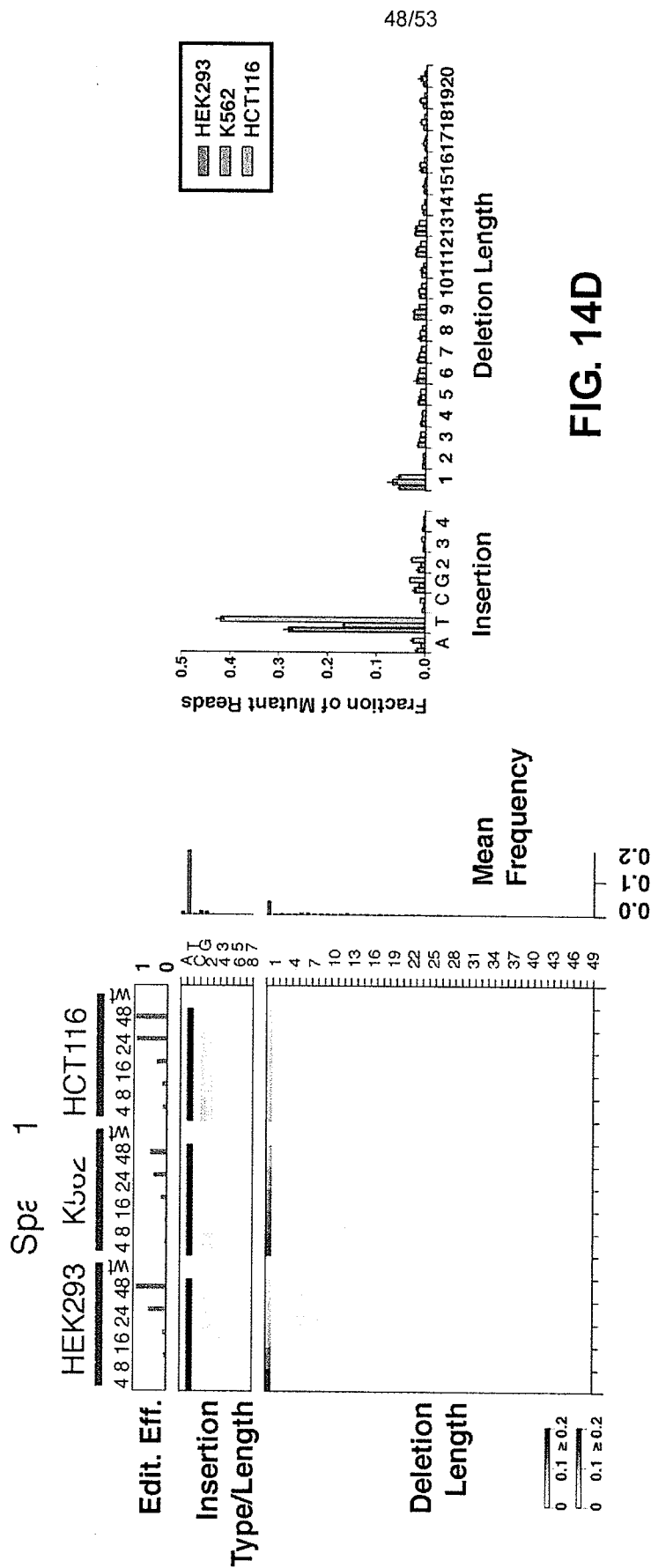


FIG. 14C

FIG. 14D

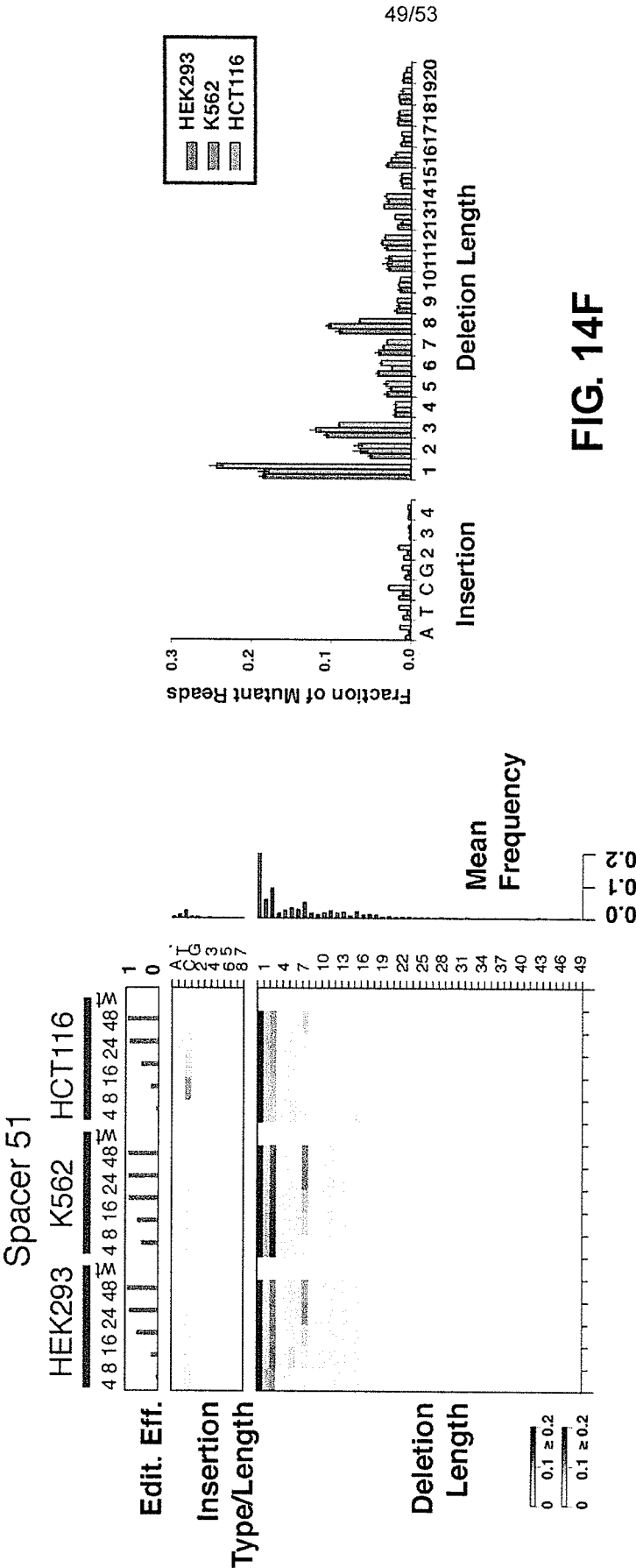
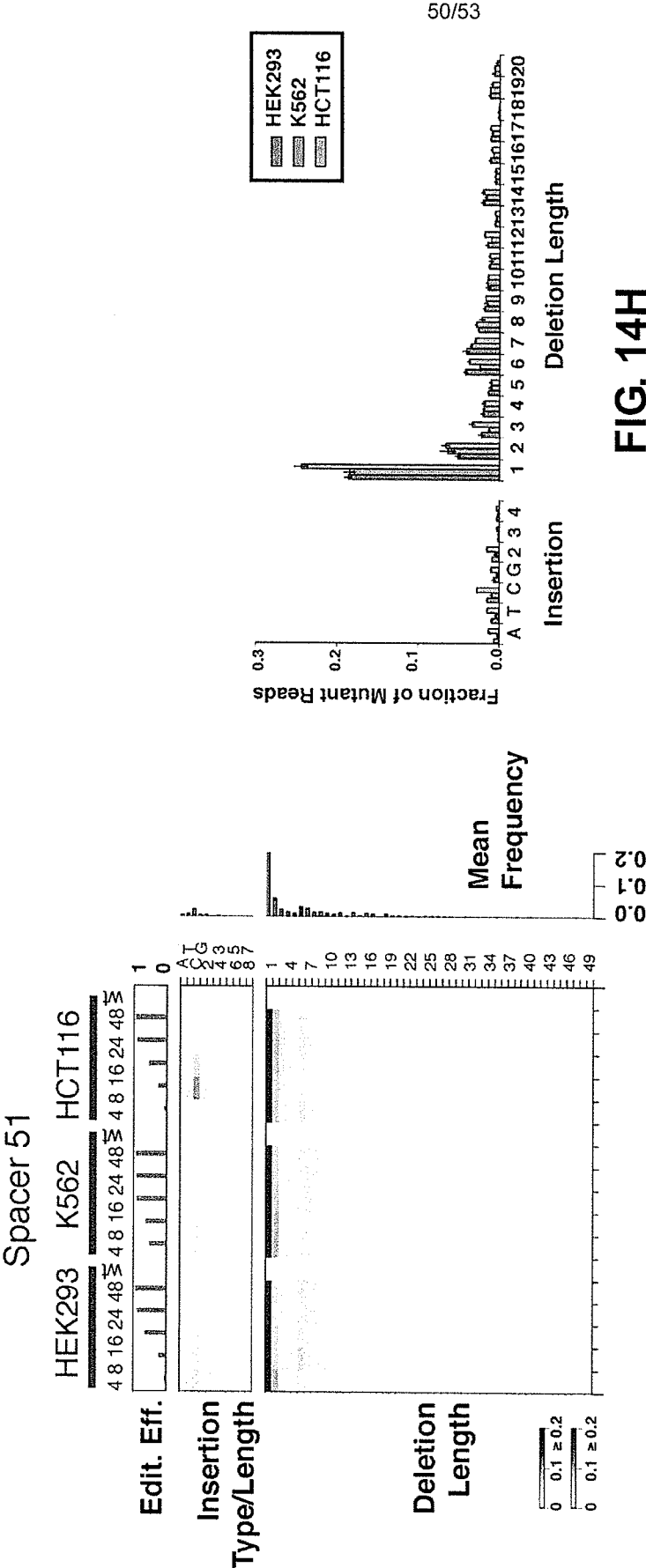
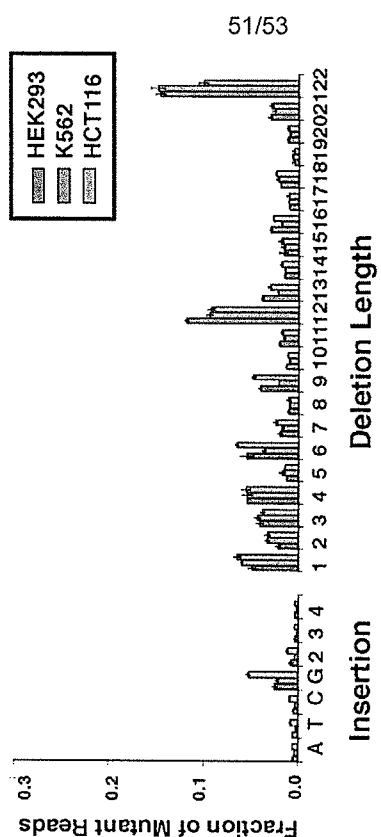
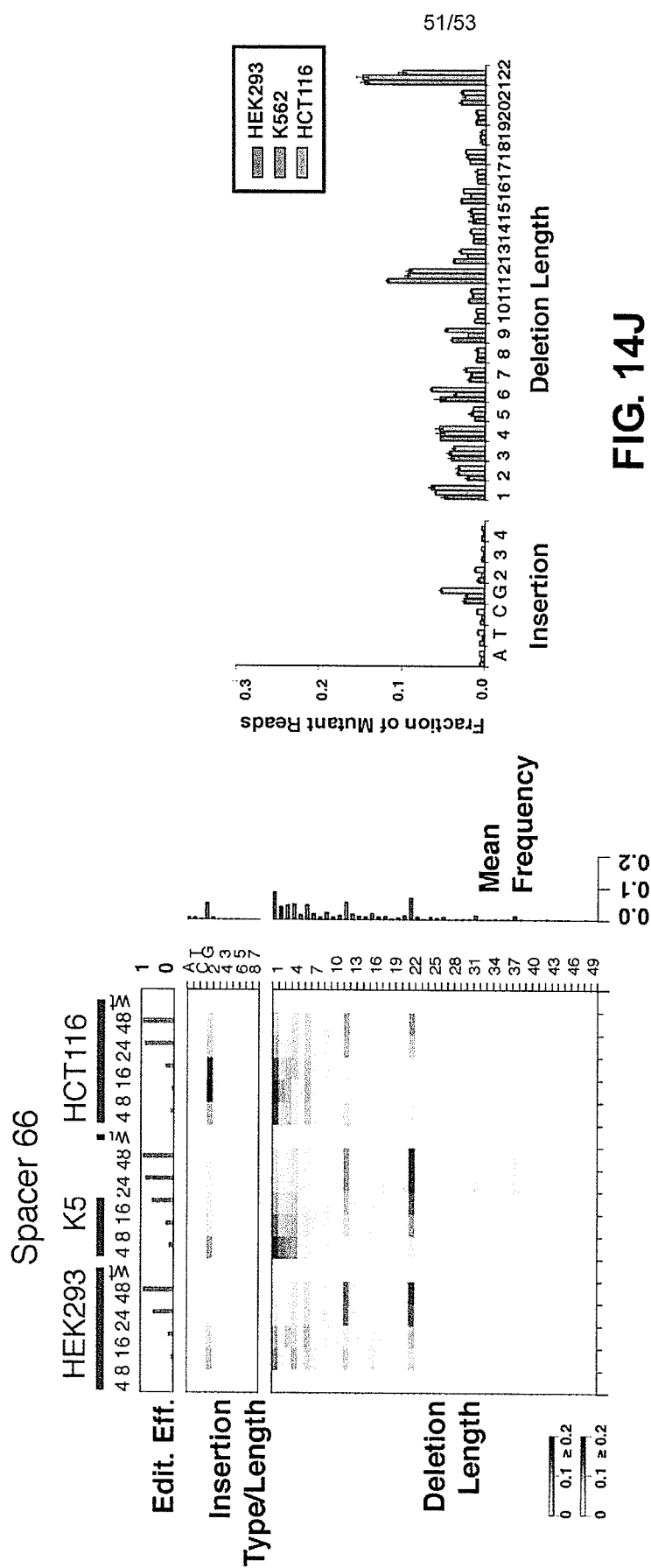
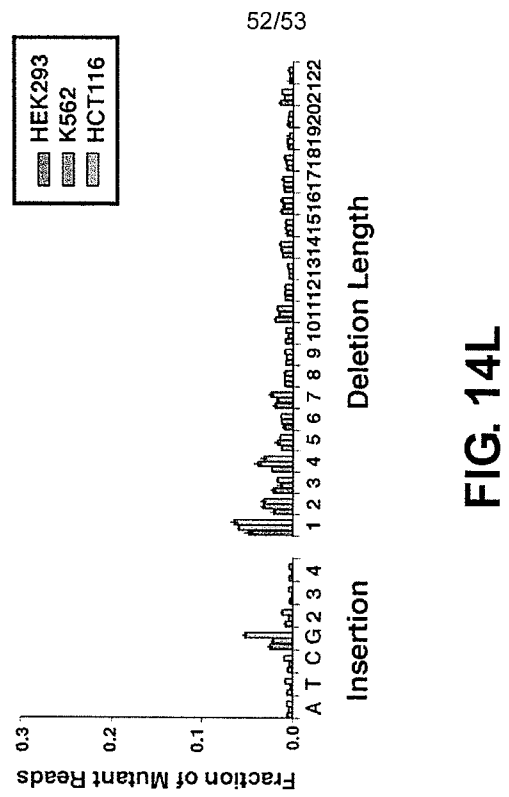
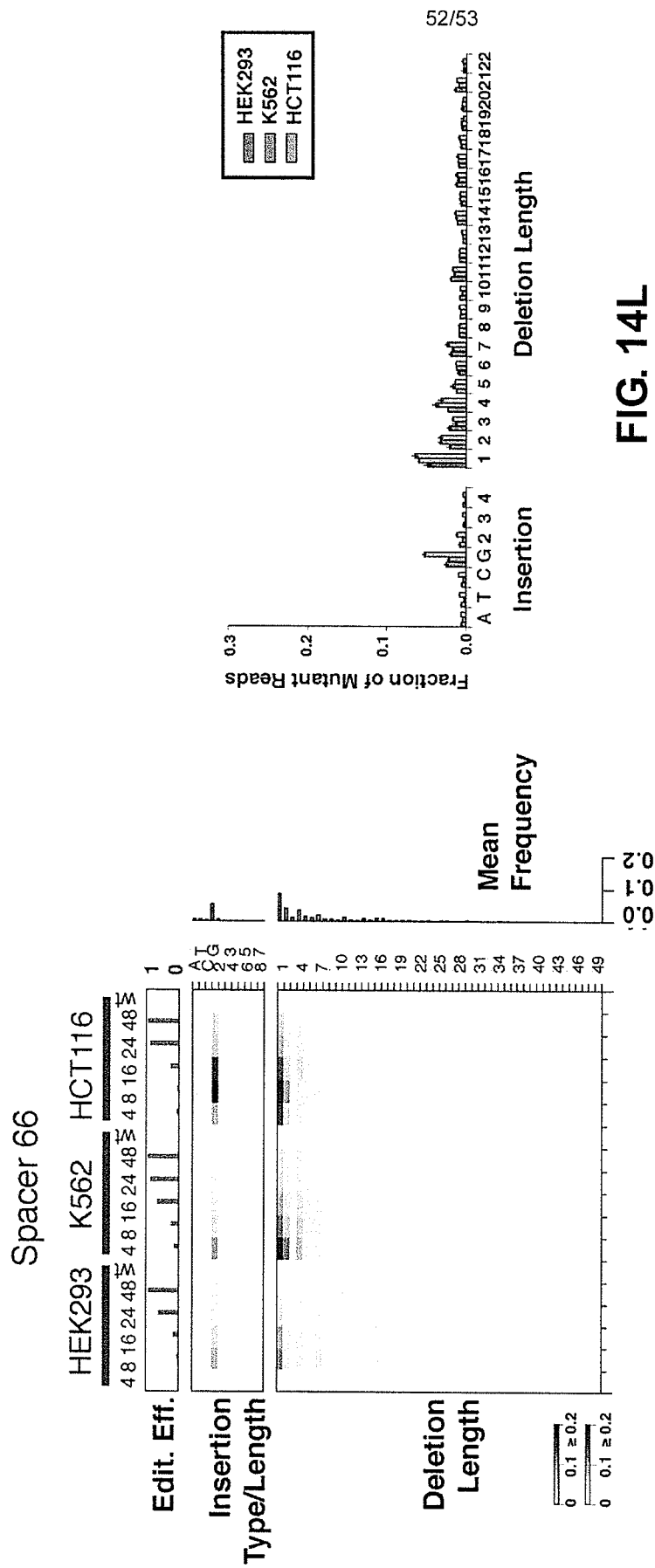


FIG. 14F







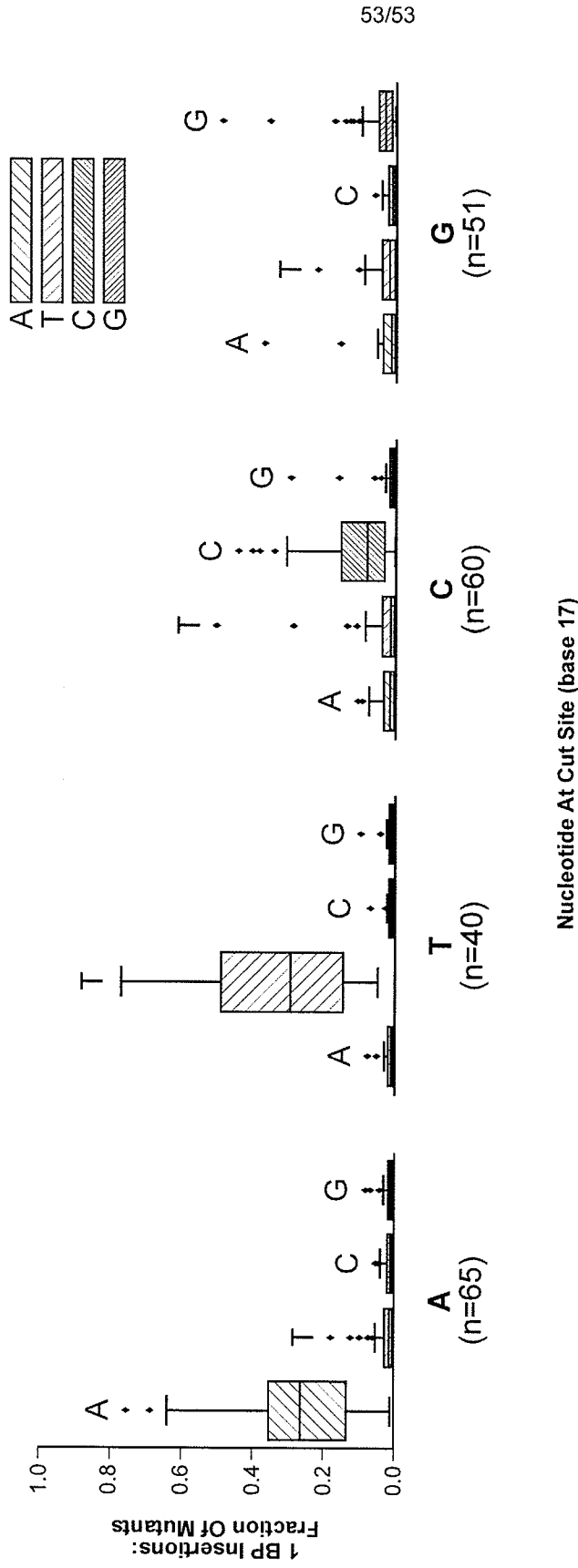


FIG. 15

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/018679

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/10 C12N15/90
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, BIOSIS, CAB Data, Sequence Search, 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, P	WO 2016/073990 A2 (EDITAS MEDICINE INC [US]) 12 May 2016 (2016-05-12) the whole document	1-20
Y	----- WO 01/90404 A1 (IMP CANCER RES TECH [GB]; WEST STEVE CRAIG [GB]; HANAKAHI LESLYN ANN A) 29 November 2001 (2001-11-29) page 42, line 26 - page 43, line 5 page 44, line 4 - page 45, line 3; claim 42	1-20
Y	----- WO 01/68882 A2 (TRANSKARYOTIC THERAPIES INC [US]; IVANOV EVGUENII [US]) 20 September 2001 (2001-09-20) claims 1-75 ----- -/-	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

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

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

13 April 2017

Date of mailing of the international search report

12/05/2017

Name and mailing address of the ISA/

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

Authorized officer

Hornig, Horst

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/018679

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2007/013979 A2 (INVITROGEN CORP [US]; UNIV CALIFORNIA [US]; BERTOLINI LUCIANA R [US];) 1 February 2007 (2007-02-01) claims 1-39 -----	1-20
Y	WO 2011/078662 A1 (KEYGENE NV [NL]; BUNDOCK PAUL [NL]; DE BOTH MICHIEL THEODOOR JAN [NL]) 30 June 2011 (2011-06-30) claims 1-19 -----	1-20
Y	WO 2011/033375 A2 (SELEXIS SA [CH]; MERMOD NICOLAS [CH]; GIROD PIERRE-ALAIN [CH]; GRANDJE) 24 March 2011 (2011-03-24) claims 1-70 -----	1-20
Y	WO 2012/168307 A2 (HELMHOLTZ ZENTRUM MUENCHEN [DE]; KUEHN RALF [DE]; WURST WOLFGANG [DE]) 13 December 2012 (2012-12-13) page 9, line 28 - page 11, line 23; claims 1-15 page 30, line 21 - page 33, line 12 -----	1-20
Y	SIMSEK; JASIN: "Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation", NATURE STRUCTURAL & MOLECULAR BIOLOGY, vol. 17, no. 4, 1 April 2010 (2010-04-01), pages 410-416 +1pg, XP002769166, ISSN: 1545-9993, DOI: 10.1038/nsmb.1773 cited in the application the whole document -----	1-20
Y	SHIBATA ATSUSHI ET AL: "DNA Double-Strand Break Repair Pathway Choice Is Directed by Distinct MRE11 Nuclease Activities", MOLECULAR CELL, vol. 53, no. 1, 5 December 2013 (2013-12-05), pages 7-18, XP028808721, CELL PRESS, CAMBRIDGE, MA, US ISSN: 1097-2765, DOI: 10.1016/J.MOLCEL.2013.11.003 cited in the application the whole document -----	1-20
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/018679

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ROSITSA DUEVA AND GEORGE ILIAKIS: "Alternative pathways of non homologous end joining (NHEJ) in genomic instability and cancer", TRANSL CANCER RES, vol. 2, no. 3, 1 January 2014 (2014-01-01) , pages 163-177, XP002769167, Pioneer Bioscience Publishing Company, DOI: 10.3978/j.issn.2218-676X.2013.05.02 the whole document	1-20
Y	----- PERRAULT RONEL ET AL: "Backup pathways of NHEJ are suppressed by DNA-PK.", JOURNAL OF CELLULAR BIOCHEMISTRY, vol. 92, no. 4, 1 July 2004 (2004-07-01), pages 781-794, XP002769168, ISSN: 0730-2312 the whole document	1-20
Y	----- AGNEL SFEIR ET AL: "Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?", TRENDS IN BIOCHEMICAL SCIENCES, vol. 40, no. 11, 1 November 2015 (2015-11-01), pages 701-714, XP055363900, AMSTERDAM, NL ISSN: 0968-0004, DOI: 10.1016/j.tibs.2015.08.006 cited in the application the whole document	1-20
Y	----- ROBERT FRANCIS ET AL: "Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing.", GENOME MEDICINE, vol. 7, no. 1, 93, 27 August 2015 (2015-08-27), pages 1-11, XP002769169, ISSN: 1756-994X, DOI: 10.1186/s13073-015-0215-6 cited in the application the whole document	1-20
Y	----- MARUYAMA TAKESHI ET AL: "Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining.", NATURE BIOTECHNOLOGY, vol. 33, no. 5, May 2015 (2015-05), pages 538-542, XP055290186, ISSN: 1546-1696 cited in the application the whole document	1-20
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/018679

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	VAN TRUNG CHU ET AL: "Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells", NATURE BIOTECHNOLOGY, vol. 33, no. 5, 24 March 2015 (2015-03-24) , pages 543-548, XP055290254, US ISSN: 1087-0156, DOI: 10.1038/nbt.3198 the whole document	1-20
Y	----- SUPRIYA V. VARTAK ET AL: "Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing", FEBS JOURNAL, vol. 282, no. 22, 9 September 2015 (2015-09-09), pages 4289-4294, XP055342300, GB ISSN: 1742-464X, DOI: 10.1111/febs.13416 the whole document	1-20
A	----- CRISTIAN BOBOILA ET AL: "Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4", THE JOURNAL OF EXPERIMENTAL MEDICINE, vol. 207, no. 2, 8 February 2010 (2010-02-08), pages 417-427, XP055363700, US ISSN: 0022-1007, DOI: 10.1084/jem.20092449 the whole document	1-20
A	----- BOBOILA CRISTIAN ET AL: "Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70.", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, vol. 107, no. 7, 16 February 2010 (2010-02-16), pages 3034-3039, XP055363718, ISSN: 1091-6490 the whole document ----- -/--	1-20

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/018679

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	FRANK-VAILLANT M ET AL: "NHEJ regulation by mating type is exercised through a novel protein, Lif2p, essential to the ligase IV pathway.", GENES & DEVELOPMENT, vol. 15, no. 22, 15 November 2001 (2001-11-15), pages 3005-3012, XP002769171, ISSN: 0890-9369 the whole document	1-20
A	----- TERASAWA MASAHIRO ET AL: "Canonical non-homologous end joining in mitosis induces genome instability and is suppressed by M-phase-specific phosphorylation of XRCC4.", PLOS GENETICS, vol. 10, no. 8, E1004563, 28 August 2014 (2014-08-28), pages 1-15, XP002769172, ISSN: 1553-7404, DOI: 10.1371/journal.pgen.1004563 the whole document	1-20
A	----- MAASSEN N ET AL: "Nonhomologous end joining and homologous recombination DNA repair pathways in integration mutagenesis in the xylose-fermenting yeast Pichia stipitis", FEMS YEAST RESEARCH, vol. 8, no. 5, 1 August 2008 (2008-08-01), pages 735-743, XP002589845, WILEY-BLACKWELL PUBLISHING LTD, GB, NL ISSN: 1567-1356, DOI: 10.1111/J.1567-1364.2008.00383.X [retrieved on 2008-04-23] the whole document	1-20
A	----- LI ANGELA Y J ET AL: "Suppression of nonhomologous end joining repair by overexpression of HMGA2.", CANCER RESEARCH, vol. 69, no. 14, 15 July 2009 (2009-07-15), pages 5699-5706, XP002769173, ISSN: 1538-7445 the whole document	1-20
A	----- ADAMO ADELE ET AL: "Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia.", MOLECULAR CELL, vol. 39, no. 1, 9 July 2010 (2010-07-09), pages 25-35, XP002769174, ISSN: 1097-4164 the whole document	1-20
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/018679

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HANAKAHI L A ET AL: "BINDING OF INOSITOL PHOSPHATE TO DNA-PK AND STIMULATION OF DOUBLE-STRAND BREAK REPAIR", CELL, vol. 102, no. 6, 15 September 2000 (2000-09-15), pages 721-729, XP001032987, CELL PRESS, US ISSN: 0092-8674, DOI: 10.1016/S0092-8674(00)00061-1 the whole document -----	1-20
A	ALLEN C ET AL: "Interactive competition between homologous recombination and non-homologous end joining", , AMERICAN ASSOCIATION FOR CANCER RESEARCH, US, vol. 1, no. 12, 1 October 2003 (2003-10-01), pages 913-920, XP002616893, AMERICAN ASSOCIATION FOR CANCER RESEARCH, US ISSN: 1541-7786 the whole document -----	1-20
A	A. VASILEVA: "Homologous recombination is required for AAV-mediated gene targeting", NUCLEIC ACIDS RESEARCH, vol. 34, no. 11, 28 June 2006 (2006-06-28) , pages 3345-3360, XP055019023, ISSN: 0305-1048, DOI: 10.1093/nar/gkl455 the whole document -----	1-20
A	WO 2014/172458 A1 (UNIV WASHINGTON CT COMMERCIALI [US]) 23 October 2014 (2014-10-23) claims 1-48 -----	1-20
A	WO 2005/012524 A1 (UNIV SHEFFIELD [GB]; HELLEDAY THOMAS [SE]) 10 February 2005 (2005-02-10) claims 1-32 -----	1-20

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/018679

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.:

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.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. claims: 1-20

A method of modulating DNA repair outcomes;

1.1. claims: 1-20

Invention 1: A method of modulating DNA repair outcomes comprising: (1) contacting a selected target nucleic acid comprising a region to be modified with (a) an agent that suppresses non-homologous end-joining (NHEJ), thereby favoring alternative end-joining (alt-EJ)/microhomology mediated end-joining (MMEJ), and (b) a DNA binding molecule that targets said selected target nucleic acid; and (2) producing one or more double-strand breaks in the targeted region using a programmable endonuclease, thereby triggering DNA repair pathways to repair the breaks, whereby repair of the cleaved target nucleic acid proceeds substantially by MMEJ and is done in the absence of a donor polynucleotide, thereby modulating the DNA repair outcome.

Invention 2: A method of modulating DNA repair outcomes comprising: (1) contacting a selected target nucleic acid comprising a region to be modified with (a) an agent that suppresses alternative end-joining (alt-EJ)/microhomology mediated end-joining (MMEJ), thereby favoring non-homologous end-joining (NHEJ), and (b) a DNA binding molecule that targets said selected target nucleic acid; and (2) producing one or more double-strand breaks in the targeted region using a programmable endonuclease, thereby triggering DNA repair pathways to repair the breaks, whereby repair of the cleaved target nucleic acid proceeds substantially by NHEJ and is done in the absence of a donor polynucleotide, thereby modulating the DNA repair outcome.

Invention 3: A method of modulating DNA repair outcomes comprising: (1) contacting a selected target nucleic acid comprising a region to be modified with (a) a first agent that suppresses alternative end-joining (alt-EJ)/microhomology mediated end-joining (MMEJ), and a second agent that suppresses non-homologous end-joining (NHEJ), thereby favoring homology directed repair (HDR), and (b) a DNA binding molecule that targets said selected target nucleic acid; (2) producing one or more double-strand breaks in the targeted region using a programmable endonuclease; and (3) inserting at least a portion of a donor polynucleotide into said target nucleic acid at the double-strand break by an HDR DNA repair pathway, thereby modulating the DNA repair outcome.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2017/018679

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2016073990	A2	12-05-2016	NONE
WO 0190404	A1	29-11-2001	AU 5653001 A 03-12-2001 CA 2408749 A1 29-11-2001 EP 1283903 A1 19-02-2003 JP 2004500849 A 15-01-2004 WO 0190404 A1 29-11-2001
WO 0168882	A2	20-09-2001	AT 525472 T 15-10-2011 AU 4563701 A 24-09-2001 CA 2403113 A1 20-09-2001 DK 1263972 T3 02-01-2012 EP 1263972 A2 11-12-2002 JP 2003526376 A 09-09-2003 MX PA02009044 A 25-04-2003 TR 200202196 T2 21-01-2003 US 6569681 B1 27-05-2003 WO 0168882 A2 20-09-2001
WO 2007013979	A2	01-02-2007	US 2007155014 A1 05-07-2007 WO 2007013979 A2 01-02-2007
WO 2011078662	A1	30-06-2011	NONE
WO 2011033375	A2	24-03-2011	AU 2010297011 A1 12-04-2012 CA 2774470 A1 24-03-2011 CN 102575264 A 11-07-2012 EP 2478108 A2 25-07-2012 JP 2013505013 A 14-02-2013 KR 20120099376 A 10-09-2012 RU 2012107634 A 27-10-2013 SG 178940 A1 27-04-2012 US 2012231449 A1 13-09-2012 WO 2011033375 A2 24-03-2011 ZA 201201504 B 24-04-2013
WO 2012168307	A2	13-12-2012	EP 2718446 A2 16-04-2014 US 2014304847 A1 09-10-2014 WO 2012168307 A2 13-12-2012
WO 2014172458	A1	23-10-2014	EP 2986709 A1 24-02-2016 US 2016040155 A1 11-02-2016 WO 2014172458 A1 23-10-2014
WO 2005012524	A1	10-02-2005	AT 516353 T 15-07-2011 AU 2004261779 A1 10-02-2005 BR PI0412909 A 26-09-2006 CA 2533423 A1 10-02-2005 CN 1856313 A 01-11-2006 CN 1856572 A 01-11-2006 CN 102935230 A 20-02-2013 DK 1649017 T3 17-10-2011 EP 1649017 A1 26-04-2006 ES 2367433 T3 03-11-2011 GB 2419882 A 10-05-2006 HK 1094006 A1 28-03-2014 JP 5519097 B2 11-06-2014 JP 5547164 B2 09-07-2014

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2017/018679

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		JP 5848728 B2	27-01-2016
		JP 2006528618 A	21-12-2006
		JP 2012102099 A	31-05-2012
		JP 2014001210 A	09-01-2014
		KR 20060052877 A	19-05-2006
		NO 338398 B1	15-08-2016
		NZ 545307 A	28-08-2009
		PT 1649017 E	30-09-2011
		US 2007179160 A1	02-08-2007
		US 2015005327 A1	01-01-2015
		WO 2005012524 A1	10-02-2005
		ZA 200601386 B	25-04-2007
