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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 June 2025 (26.06.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/137259 A1

(51) International Patent Classification:
C12N 15/11 (2006.01) *C12N 5/10* (2006.01)
C12N 15/09 (2006.01) *A61K 35/76* (2015.01)
C12N 15/864 (2006.01)

(21) International Application Number:
PCT/US2024/061007

(22) International Filing Date:
19 December 2024 (19.12.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/612,909 20 December 2023 (20.12.2023) US

(71) Applicant: **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; Office of the General Counsel, Bldg. 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, CA 94305 (US).

(72) Inventors: **SUCHY, Fabian, Patrik**; Office of the General Counsel, Bldg 170, 3rd Flr, Main Quad, Stanford, CA 94305-2038 (US). **KARIGANE, Daiki**; Office of the General Counsel, Bldg 170, 3rd Flr, Main Quad, Stanford, CA 94305-2038 (US). **BHADURY, Joydeep**; Office of the General Counsel, Bldg 170, 3rd Flr, Main Quad, Stanford, CA 94305-2038 (US). **MAJETI, Ravindra**; Office of the General Counsel, Bldg 170, 3rd Flr, Main Quad, Stanford, CA 94305-2038 (US). **NAKAUCHI, Hiromitsu**; Office of the General Counsel, Bldg 170, 3rd Flr, Main Quad, Stanford, CA 94305-2038 (US).

(74) Agent: **NOLAND, Greta, E.**; Marshall, Gerstein & Borun LLP, 233 S. Wacker Drive, 6300 Willis Tower, Chicago, IL 60606-6357 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))*

WO 2025/137259 A1

(54) Title: DETECTION AND PREVENTION OF UNINTENDED CRISPR/AAV-MEDIATED CONCATEMERIC KNOCKINS

(57) Abstract: The disclosure provides methods for preventing unintended concatemeric insertions arising from CRISPR/AAV-mediated genetic modifications to increase editing efficiency, decrease off-target effects, and improve on-target fidelity. Also provided are digital droplet PCR-based methods for evaluating complex genotypes and detecting unintended concatemeric insertions.

DETECTION AND PREVENTION OF UNINTENDED CRISPR/AAV-MEDIATED CONCATEMERIC KNOCKINS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. provisional application No. 63/612,909 filed December 20, 2023, which is incorporated herein by reference in its entirety.

INCORPORATION BY REFERENCE OF THE SEQUENCE LISTING

[0002] This application contains, as a separate part of disclosure, a Sequence Listing in computer-readable form (Filename: 70014_SeqListing.xml; 7,259 bytes – XML file dated December 18, 2024) which is incorporated by reference herein in its entirety.

FIELD

[0003] The disclosure provides methods for preventing unintended concatemeric insertions arising from CRISPR/AAV-mediated genetic modifications to increase editing efficiency, decrease off-target effects, and improve on-target fidelity. Also provided are digital droplet PCR-based methods for evaluating complex genotypes and detecting unintended concatemeric insertions.

BACKGROUND

[0004] The combination of CRISPR paired with adeno-associated virus (AAV) has proven to be highly efficient for producing site-specific genetic modifications. Ribonucleoprotein (RNP) is electroporated into cells where it cleaves its genomic target site, creating a double-stranded break. The AAV serves as a delivery vehicle for a single-stranded DNA repair template which is utilized by the cell's endogenous homology-directed repair (HDR) machinery to repair the RNP-induced break. This approach has been used to make both small changes and large insertions into the genome of cells *in vitro* and *in vivo*. Importantly, the use of AAV as a nucleic acid delivery vehicle is well-studied and its safety established in numerous clinical trials worldwide for multiple purposes including gene correction and enzyme replacement therapy, see for example NCT02702115 and NCT03041324.

[0005] However, anomalous insertions of AAV into CRISPR cut-sites can obscure preclinical conclusions and hinder reproducibility. In particular, concatemeric insertions, also referred to as "knockins", may inadvertently produce genetic knockouts or other unexpected phenotypes, resulting in unintended and potentially dangerous clinical outcomes.

[0006] Strategies for detecting and preventing anomalous insertions from CRISPR/AAV-mediated genetic modifications are needed.

BRIEF SUMMARY

[0007] The present invention provides compositions and methods for detecting and preventing unintended concatemer formation arising from site directed nuclease/AAV-mediated genetic modifications.

[0008] Provided are methods for preventing unintended concatemeric insertions of a viral genome at an insert site in a cellular genomic DNA of a cell, the method comprising introducing into a population of cells a first and second ribonucleoprotein ("RNP") and a recombinant adeno-associated virus (AAV) comprising a repair template and incubating the cells under conditions sufficient for cleavage of a target sequence of cellular genomic DNA by the first RNP and cleavage of a viral inverted terminal repeat (ITR) by the second RNP. In aspects, each RNP includes a complex of a guide RNA (gRNA) and a clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) endonuclease. In aspects, the gRNA of the first RNP includes a guide sequence complementary to the target sequence of cellular genomic DNA and the gRNA of the second RNP includes a guide sequence complementary to a sequence of the viral ITR. The methods may be used, for example, to remove viral ITRs after transduction of a population of human hematopoietic stem cells (HSCs) and increase engraftment of the modified HSCs when transplanted into patients.

[0009] Also provided are methods for preventing unintended concatemeric insertions of a viral genome at an insert site in a genomic DNA of a cell, the method comprising introducing into a population of cells a ribonucleoprotein ("RNP") and a recombinant adeno-associated virus (AAV) comprising a repair template, and incubating the cells under conditions sufficient for cleavage of a target sequence of cellular genomic DNA by the RNP, where the recombinant AAV includes two polynucleotide insertions flanking its homology arms, each insertion comprising the target sequence.

[0010] Also provided are methods for preventing unintended concatemeric insertions of a viral genome at an insert site in a genomic DNA of a cell, the method comprising introducing into a population of cells a ribonucleoprotein ("RNP"), a modified adeno-associated virus (AAV) comprising a repair template, and a PI3K/mTOR kinase inhibitor, and incubating the cells under conditions sufficient for cleavage of a target sequence of cellular genomic DNA by

the RNP, where the PI3K/mTOR kinase inhibitor is introduced before, simultaneously with, or after the RNP.

[0011] The methods may also include where the RNP comprises a complex of a guide RNA (gRNA) and a clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) endonuclease. The methods may also include where the PI3K/mTOR kinase inhibitor is selected from apitolisib (GDC-0980), bimiralisib (PQR309), dactolisib (NVP-BEZ235), gedatolisib (PKI-587; PF-05212384), paxalisib (GDC-0084), or samotolisib (LY3023414). In aspects, the PI3K/mTOR kinase inhibitor is selected from apitolisib, bimiralisib, dactolisib, gedatolisib, omipalisib, paxalisib, samotolisib, voxtalisib, CMG 002, GNE-477, GSK1059615, MCX-83, NSC765844, PF-04691502, PF-04979064, PI-108, PI-103BE, PKI-402, SN202, or VS-5584. The methods may also include where the PI3K/mTOR kinase inhibitor is dactolisib (NVP-BEZ235). The method may also include where the RNP is introduced to the population of cells by electroporation. The methods may also include where the RNP is introduced to the population of cells by injection or by intravenous infusion. The methods may also include where the AAV is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, or AAV12. The methods may also include where the AAV is AVV2 or AAV6. The methods may also include where the Cas endonuclease is a Cas9 or a Cas12a endonuclease. The methods may also include where the population of cells includes any vertebrate or mammalian cell type. The methods may also include where the population of cells includes pluripotent stem cells, induced pluripotent stem cells, embryonic stem cells, or hematopoietic stem and progenitor cells.

[0012] Also provided are methods for detecting concatemeric insertions of a viral genome at an insert site in a genomic DNA of a cell, the methods comprising digital droplet PCR genotyping of the cellular DNA in combination with restriction endonuclease analysis.

[0013] In aspects, the method includes a copy number variation analysis which includes performing a first ddPCR reaction to produce a first set of amplicons, the reaction comprising a target cellular genomic DNA, a first primer/probe set comprising a first reference primer directed to a first genomic reference site in *cis* or *trans* with the insert site in the cellular genomic DNA, an optional second primer/probe set comprising a second reference primer directed to a second genomic reference site in *cis* or *trans* with the insert site, a third primer/probe set comprising a primer directed to a region of the viral genome, and a fourth primer/probe set comprising a primer directed to a cellular genomic region spanning the insert

site such that the primers and probe are generally at least 25 bp from the insert site, performing a second ddPCR reaction to produce a second set of amplicons, the second ddPCR reaction comprising the target cellular genomic DNA, the first primer/probe set, optionally the second primer/probe set, the third primer/probe set, the fourth primer/probe set, and at least one restriction endonuclease which cleaves within the insert, and comparing the first and second sets of amplicons to determine whether concatemeric insertions of the viral genome are present. If concatemeric insertions are present, the number of detected inserts will increase in the second ddPCR reaction with the restriction enzyme.

[0014] In aspects, the method includes a linkage analysis which includes performing a ddPCR reaction to produce a set of amplicons, the ddPCR reaction comprising a target cellular genomic DNA, a first primer/probe set comprising a first reference primer directed to a first genomic reference site in *cis* with the insert site in the cellular genomic DNA, an optional second primer/probe set includes a second reference primer directed to a second genomic reference site in *cis* with the insert site, a third primer/probe set includes a primer directed to a region of the viral genome, and a fourth primer/probe set includes a primer directed to a region of the viral genome not intended to be inserted; and comparing the linkage of amplicons of the primer/probe sets to determine a ratio of viral inserts linked to the target site, where detection of the fourth amplicon linked to a reference amplicon indicates concatemeric or non-HDR mediated insertions have occurred. In aspects, method may also include where the fourth primer/probe set comprises a primer directed to a viral ITR region or a region of the viral genome outside of the viral homology arms. The method may also include performing a second ddPCR reaction comprising the target cellular genomic DNA, the first primer/probe set, optionally the second primer/probe set, the third primer/probe set, the fourth primer/probe set, and at least one restriction endonuclease which cleaves at a target site within the viral genome; and comparing the linkage of amplicons of each of the primer/probe sets in the ddPCR reactions performed with and without the restriction endonuclease in order to determine whether the insertions are concatemeric or non-HDR mediated insertions. For example, where concatemeric insertions occurred, a restriction enzyme that cuts the ITR should increase the number of viral inserts detected, while the linkage between reference amplicons and insert amplicons should be unchanged. In contrast, if non-HDR mediated inserts occurred, the linkage between references and insert amplicons should decrease because the ITR region is between the two primer/probe sets.

[0015] In aspects, method includes a nuclease mediated loss of linkage analysis which includes performing a first ddPCR reaction to produce a first set of amplicons, the reaction comprising a target cellular genomic DNA, a first primer/probe set comprising a first reference primer directed to a first genomic reference site in *cis* with an insert site in the cellular genomic DNA, a second primer/probe set comprising a second reference primer directed to a second genomic reference site in *cis* with the insert site; performing a second ddPCR reaction to produce a second set of amplicons, the reaction comprising the target cellular genomic DNA, the first primer/probe set, the second primer/probe set, and at least one restriction endonuclease which cleaves a site between the first and second cellular genomic reference sites, and comparing the first and second sets of amplicons to detect loss of linkage between the first and second cellular genomic reference sites, where loss of linkage indicates a concatemeric insertions of a viral genome. If the target of the endonuclease exists between the two amplicons, linkage should decrease in the second ddPCR reaction. As such, the endonuclease can be used to detect intended and unintended insertions (e.g., ITRs) at the target site.

[0016] Other technical features may be readily apparent to one skilled in the art from the following figures, descriptions, and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] FIG. 1A is a schematic illustrating ddPCR allele counting strategy for genotyping. Four ddPCR targets are multiplexed in a single well for each sample. Primers indicated by arrows, and probes indicated by small rectangles flanked by primers. Ref-1 and Ref-2 indicate trans references used to determine overall cell number. RE indicates restriction enzymes. HEX and FAM indicate probe color. High and Low indicate concentration of probe (used for ddPCR amplitude multiplexing), resulting in high or low clusters shown in FIG. 1B

[0018] FIG. 1B shows a representative two-dimensional ddPCR plot of the reaction shown in FIG. 1A. Ref-1 and Ref-2 indicate trans references used to determine overall cell number; KI indicates an allele at the target locus (e.g., *CD14*, *AAVS1*, etc.) that contains at least one insert; No KI indicates an allele at the target locus that does not contain any inserts.

[0019] FIG. 1C is a schematic indicating concatemeric knockin. Without restriction enzyme, concatemeric inserts will be linked.

[0020] FIG. 1D is a schematic indicating ddPCR droplet partitioning. Without restriction enzyme, linked concatemeric inserts will partition in the same droplet and be counted only once.

[0021] FIG. 2A is a schematic illustrating ddPCR linkage analysis strategy for measuring knockin (KI) frequency in bulk (non-subcloned) samples. Cis reference sites are shown as small light-gray and dark-gray squares. Ubc-GFP amplicon site indicated by small square above the Ubc-GFP insert. Ubc-mCherry amplicon site indicated by small square above the Ubc-mCh insert. Dashed-lines indicate linkage, which can be measured by ddPCR.

[0022] FIG. 2B is a schematic illustrating ddPCR linkage analysis strategy for counting concatemers in double knockin subclones using ddPCR. Scissors indicate restriction enzymes. X indicates linkage lost, such that concatemeric inserts will be separated.

[0023] FIG. 2C is a schematic illustrating ddPCR linkage analysis strategy for measuring KI frequency in bulk non-subcloned samples. Cis reference sites shown as small light-gray and dark-gray squares. UBC-GFP amplicon site indicated by small square above Ubc-GFP insert. Dashed-lines indicate linkage, which can be measured by ddPCR. ID1 and ID2 are small (<100 bp) unique DNA sequences added outside the homology arms in the viral genome as shown at the top.

[0024] FIG. 2D is a schematic illustrating ddPCR linkage analysis strategy for measuring concatemeric-KI frequency in bulk non-subcloned samples. Cis reference sites shown as small light-gray and dark-gray squares. ddPCR amplicons for ID1 and ID2 shown as dark squares above ID1 and ID2.

[0025] FIG. 3A Digital PCR genotype of select subclones after knocking in CRE in the CD14 locus in human PSCs. Two copies per cell are expected (left axis). Extra copies shown on right axis. Additional insertions of +1 are present for the samples represented by bars 4 and 5 from left; additional insertions of +2 and +3, respectively, are present for the samples represented by bars 6 and 7. Extra insert indicates number of insertions greater than 1 or 2 for monoallelic and biallelic knockins, respectively. Interpreted genotype at bottom. No KI indicates no insert; KI indicates insert is present.

[0026] FIG. 3B ddPCR genotyping of 36 subclones expanded after knocking in CRE into *CD14* locus in PT-iPSCs. Interpreted genotype indicated by circles at bottom: white = WT; gray = monoallelic; black = biallelic; red = cannot be determined. +/-RE indicates analysis is performed with (+) or without (-) restriction enzymes that cut sites to separate concatemers. KI

indicates an allele at the target locus that contains at least one insert; No KI indicates an allele at the target locus that does not contain any inserts. Average of the two references set to 2 copies/cell. Numbers correspond to sample ID's.

[0027] FIG. 3C shows Southern analysis of the subclones shown in FIG. 3A. Bands are indicated on left as corresponding to no insert (no KI), one copy of the insert (KI), two copies of the insert (KI+1), and three copies of the insert (KI+2). The number of extra copies of the insert are shown above the respective lanes as +1, +1, +2, and +3.

[0028] FIG. 3D Comparison of PCR-L, PCR-R, PCR-F, ddPCR, and Southern blot genotype after knocking in CRE into the CD14 locus in human PSCs. Each row indicates an individual sample; sample # indicated on left. Red indicates unexpected result due to incorrect band size, missing information (*e.g.*, no band), or additional insertions.

[0029] FIG. 4A shows gel electrophoresis analysis from PCR genotyping. 36 subclones expanded after knocking in Ubc-CRE into *AAVS1* locus in PT-iPSCs. Ladder size shown in kb. Sample ID indicated at top and bottom of gels; L = ladder. Interpreted genotype indicated by circle at top of each gel: white = WT; gray = monoallelic; black = biallelic; red = cannot be determined. Top, middle, and bottom row of gel indicate PCR-L, PCR-R, and PCR-F genotyping strategies, respectively.

[0030] FIG. 4B shows Ref-1 and Ref-2 from ddPCR. Average of the two references set to 2 copies/cell. Sample ID shown at bottom.

[0031] FIG. 4C ddPCR genotyping results. Normalized to references shown in FIG. 4B. The reaction was run +/- restriction enzyme (RE). Interpreted genotype indicated by circle at bottom graph: white = WT; gray = monoallelic; black = biallelic; red = additional insertions or cannot be determined.

[0032] FIG. 5A Plot of Southern blot band sizes from 11 select subclones expanded after knocking in CRE into CD14 locus in PT-iPSCs. Dashed lines indicate the average size of a group of similar bands. Red circle indicates the additional band in subclone #12 that did not have similar size to adjacent groups.

[0033] FIG. 5B is a schematic showing viral genome size and theoretical arrangement of concatemeric knockin. Arrows indicate primers used for junction PCR. Below the schematic is shown Souther analysis of junction PCR of the samples. Sample ID shown on bottom; L = ladder. Bold indicates sample had an off-target insertion as indicated by ddPCR and Southern blot. ddPCR-based genotype listed on top. WT = wildtype (no knockins). M = monoallelic

knockin. B = biallelic knockin. + indicates the number of additional concatemeric insertions. +Off indicates off-target insertion(s).

[0034] FIG. 5C Plot comparing the copies of concatemer segments as measured by ddPCR. 11 subclones analyzed (same as FIG. 5A) after restriction enzyme digestion. X-axis indicates additional insertions of CRE per cell. Y-axis indicates ITRs per cell (diamonds), additional left homology arms per cell (LHA, squares), and additional right homology arms per cell (RHA, circles).

[0035] FIG. 6A Ref-1 and Ref-2 from ddPCR analysis of 36 subclones expanded after knocking in CRE into *CD14* locus in PT-iPSCs. Average of the two references set to 2 copies/cell. Sample ID shown at bottom. References are in cis: ref 1 is upstream (5') of LHA and ref 2 is downstream (3') of RHA.

[0036] FIG. 6B ddPCR results of counting LHA and RHA. Analyzed +/- restriction enzyme (RE) used to separate concatemeric regions. Normalized to references shown in FIG. 6A.

[0037] FIG. 6C qPCR measurement of AAV6 ITR when analyzed +/- AhdI RE. Y-axis is fluorescent intensity; x-axis if cycle number. Left graph shows DNA extracted from sample #10 (concatemeric monoallelic KI). Right graph shows DNA extracted from sample #14 (monoallelic KI by end-joining). Δ indicates change in cycle threshold when run with and without AhdI.

[0038] FIG. 6D ddPCR results of counting ITR insertions in the 36 subclones +/- RE (+RE is AhdI and MseI). Normalized to cis-references.

[0039] FIG. 6E is a schematic showing ddPCR linkage analysis of ITR and cis-references. Left indicates concatemeric knockin with chimeric ITR inserted. Right indicates end-joining-mediated knockin. Legend at bottom. Dashed lines indicate amplicons are linked. Red X indicates linkage is lost when adding RE.

[0040] FIG. 6F Linkage heat map between cis-reference sites and ITRs measured by ddPCR (analyzed with addition of AhdI). Sample ID shown at top. Amplicon sites and legend shown in panel j. Top row indicates the % of ref-1 sites bound to an ITR; middle row indicates linkage between ref-2 sites bound to an ITR; bottom row indicates the % of ref-1 sites bound to ref-2 sites.

[0041] FIG. 7A is a schematic showing dual-knockin and double selection.

[0042] FIG. 7B Line graph showing correlation of GFP mean fluorescent intensity (MFI) and copies of Ubc-GFP per cell. MFI measured from uniformly processed images. Ubc-GFP copies/cell measured as shown in FIG. 2B. Dashed line indicates linear regression equation, $R^2 = 0.952$.

[0043] FIG. 7C Line graph showing correlation of mCherry mean fluorescent intensity (MFI) and copies of Ubc-mCherry per cell. MFI measured from uniformly processed images. Ubc-mCherry copies/cell measured as shown in FIG. 2B. Dashed line indicates linear regression equation, $R^2 = 0.954$

[0044] FIG. 8A is a schematic showing linked (upper) and delinked (lower) AAV viral genomes.

[0045] FIG. 8B is a schematic illustrating the IC and DC methods for removing viral ITRs after transduction using Cas9 RNP.

[0046] FIG. 8C shows the different loci and cell lines tested. n indicates number of subclones analyzed in each group

[0047] FIG. 8D is a bar graph showing the results of clonal analysis of concatemer frequency after double knockin, as determined by ddPCR. Comparison of NC, IC, and DC methods. Loci and cell lines were as in FIG. 8C. 16-28 colonies were analyzed per datapoint; n = 458 total. Mean indicated above bar in bold. Error bars indicate 1 standard deviation. *** indicates $p < 0.0005$. p values are from paired t-test.

[0048] FIG. 8E is a line graph showing bulk analysis of UBC-GFP knocked into HBB locus in H9 ESCs at various viral dilutions. GFP-positive cells (determined by flow cytometry) indicated by the gray line. The % of KI alleles that contain a concatemeric insertion (determined by ddPCR) indicated by the red line.

[0049] FIG. 9 is a line graph showing bulk ddPCR concatemer analysis of recovered human cells 4 months after HSPC transplantation. Prior to transplantation, HSPCs had UBC-GFP knocked into HBB locus. Error bars indicate 1 standard deviation. *** indicates $p < 0.0005$. p values are from paired t-test.

[0050] FIG. 10 PI3K/mTOR inhibitors is a bar graph showing results of a Ubc-GFP knockin into the *HBB* locus in PT-iPSCs under eight different conditions, NC, IC, DC, and IC+DC either with or without treatment with 300 nM PI3K/mTOR kinase inhibitor (NVP-BEZ235) for 24 hours after electroporation. About 5 days later, GFP positive cells were sorted and expanded

for a week, after which the percentage of edited alleles containing a concatemer were quantified using the same methodology as illustrated in FIG. 2C and FIG. 2D.

[0051] FIG. 11A is a schematic showing three PCR genotyping strategies. PCR-L = 3 primer in-out PCR performed spanning the left homology arm. PCR-R = 3 primer in-out PCR performed spanning the right homology arm. PCR-F = 2 primer PCR with primers located outside of both homology arms.

[0052] FIG. 11B is a schematic showing expected gel-electrophoresis banding patterns and corresponding genotypes from PCR reactions in FIG. 11A. KI = band corresponding to knockin allele. No KI = band corresponding to allele without a knockin.

[0053] FIG. 12 is a graph showing percent human cells in the mouse bone marrow relative to mouse cells (*i.e.*, percent engraftment) four months after transplantation. NC indicates "no cut" condition. DC indicates "distal cut" condition.

DETAILED DESCRIPTION

[0054] CRISPR paired with adeno-associated virus (AAV) is an efficient platform for producing targeted genetic insertions. However, this tool may also produce a concealed gene-editing error that results in additional, concatemeric insertions of viral genome at the target site. This can result in adverse phenotypes and is antithetical to the intended purpose of genome engineering and precision gene therapy. Remarkably, the present inventors found these hidden anomalies in the majority of edited cells, yet such anomalies have not been previously reported. The lack of reporting may be due to the difficulty in detecting these errors by common analytical methodologies. The present invention provides compositions and methods for detecting and preventing unintended concatemer formation arising from CRISPR/AAV-mediated genetic modifications. Incorporation of the methods described here into established gene-editing pipelines will enable safer and more reliable genetic modifications that translate into more robust preclinical research, greater reproducibility, and increased clinical safety.

[0055] Accordingly, in one aspect the disclosure provides methods for preventing unintended concatemeric insertions arising from CRISPR/AAV-mediated genetic modifications. By reducing unintended concatemeric insertions the methods increase editing efficiency, decrease off-target effects, and improve on-target fidelity.

[0056] In another aspect, the disclosure provides methods for evaluating complex genotypes to detect unintended concatemeric insertions.

[0057] The term "CRISPR" refers to a system for genetic modification utilizing a class of enzymes, the clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) endonucleases. Cas proteins contain an endonuclease domain for nucleic acid cleavage and at least one RNA binding domain that interacts with a guide RNA. In aspects of the compositions and methods described here, any targeted nuclease may be used, including but limited to Cas9, Cas12, etc. In aspects, the Cas enzyme is Cas9. In other aspects, the Cas enzyme is Cas12a, also referred to as Cpf1.

[0058] CRISPR gene editing technology utilizes ribonucleoprotein complexes of a Cas endonuclease and a synthetic guide RNA (gRNA), referred to as an "RNP". The gRNA of the RNP comprises an enzyme-specific region, which binds to the Cas endonuclease, and a region complementary to a target nucleic acid, which may be referred to as the "recognition sequence" or the "guide sequence" of the gRNA. A "target sequence" refers to the sequence of a target nucleic acid that is complementary to the guide sequence of a gRNA. In the context of gene editing, the target sequence may be, for example, a sequence of a genomic DNA.

[0059] Thus, a "guide RNA" or "gRNA" refers to an RNA molecule that binds to a Cas protein and targets the Cas protein to a target sequence, *e.g.*, within a genomic DNA. Some gRNAs contain two separate RNA molecules, referred to as an "activator-RNA" and a "targeter-RNA", which may also be referred to as a tracrRNA and a crRNA, respectively. Other gRNAs contain the crRNA and tracrRNA associated as a single RNA molecule and may be referred to as a "single-guide RNA" or an "sgRNA." In aspects, an sgRNA comprises a crRNA fused to a tracrRNA, optionally via a linker polynucleotide.

[0060] In the context of CRISPR/AAV-mediated genetic modifications, AAV serves as a nucleic acid delivery vehicle for a DNA template to be utilized by the cell's endogenous homology-directed repair (HDR) machinery to repair the RNP-induced break. HDR includes "homologous recombination", a cellular process in which a double-stranded DNA-break is repaired using a homologous DNA sequence as the repair template. In the context of gene editing, the homologous DNA sequence is an exogenous nucleic acid delivered to the cell by the AAV, and may be referred to as the DNA template or the DNA repair template. The term "homology" in this context refers to sequence similarity rather than phylogenetic relatedness.

[0061] Adeno-associated viruses (AAV) are small non-enveloped single-stranded DNA viruses that are non-pathogenic in humans and depend on helper viruses for replication. In aspects of the compositions and methods described here, the AAV may be any suitable

serotype. In aspects, the AAV is serotype 6, referred to as "AAV6". In other aspects, the AAV may be AAV1, AAV2, AAV3, AAV4, AAV5, AAV7, AAV8, AAV9, AAV10, AAV11, or AAV12. In aspects, the AAV is a recombinant AAV. In aspects, the AAV is a recombinant AAV2 or a recombinant AAV6.

[0062] In aspects, the AAV is a recombinant AAV comprising cellular genomic DNA sequences flanking the repair template, where the flanking sequences are homologous to the regions of genomic DNA flanking the cellular target site. The two AAV flanking sequences may be referred to herein as "homology arms".

[0063] The present disclosure provides methods for preventing concatemeric insertions of a viral genome at a target site where a CRISPR system is utilized in combination with adeno-associated virus (AAV) delivery of a single-stranded DNA repair template for producing site-specific genetic modifications in cells or tissues *in vitro*, *ex vivo*, or *in vivo*.

[0064] In one aspect, provided is a method for preventing concatemeric insertions of a viral genome at a target genomic DNA site in a cell, the method comprising introducing into a population of cells a first and second ribonucleoprotein ("RNP") and an AAV comprising a repair template, each RNP comprising a complex of a guide RNA (gRNA) and a clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) enzyme, where the gRNA of the first RNP includes a guide sequence complementary to a target sequence of cellular DNA and the gRNA of the second RNP includes a guide sequence complementary to an endogenous sequence within a viral inverted terminal repeat (ITR) of the AAV genome. Under such conditions, cleavage of the target DNA sequence occurs by the first RNP, repair of the RNP-induced break occurs by the cells' endogenous homology-directed repair (HDR) machinery using the AAV repair template, and cleavage of the viral ITR by the second RNP. The method may be used, for example, to remove viral ITRs after transduction of a population of human hematopoietic stem cells (HSCs) and increase engraftment of the modified HSCs when transplanted into patients.

[0065] In one aspect, provided is a method for preventing concatemeric insertions of a viral genome at a target genomic DNA site in a cell, the method comprising introducing into a population of cells a ribonucleoprotein ("RNP") comprising a complex of Cas enzyme and gRNA containing a guide sequence complementary to a target cellular DNA sequence, and a modified AAV comprising a repair template, wherein the modified AAV comprises two insertion sequences flanking its homology arms, each insertion sequence comprising the target

cellular DNA sequence recognized by the RNP. Under such conditions, cleavage of the target cellular DNA occurs by the RNP, repair of the RNP-induced break occurs by the cells' endogenous homology-directed repair (HDR) machinery, and cleavage of the distal regions of the viral DNA occurs by the same RNP that targets the cellular DNA.

[0066] In one aspect, provided is a method for preventing concatemeric insertions of a viral genome at a target genomic DNA site in a cell, the method comprising introducing into a population of cells a ribonucleoprotein ("RNP"), an AAV comprising a DNA repair template, and a PI3K/mTOR kinase inhibitor, and incubating the cells under conditions sufficient to allow for cleavage of a target DNA by the RNP and repair by the cells' endogenous homology-directed repair (HDR) machinery. In aspects, the kinase inhibitor is introduced before, simultaneously with, or after the RNP. In aspects, the PI3K/mTOR kinase inhibitor is selected from apitolisib (GDC-0980), bimiralisib (PQR309), dactolisib (NVP-BEZ235), gedatolisib (PKI-587; PF-05212384), paxalisib (GDC-0084), or samotolisib (LY3023414). In aspects, the PI3K/mTOR kinase inhibitor is selected from apitolisib, bimiralisib, dactolisib, gedatolisib, omipalisib, paxalisib, samotolisib, voxtalisib, CMG 002, GNE-477, GSK1059615, MCX-83, NSC765844, PF-04691502, PF-04979064, PI-108, PI-103BE, PKI-402, SN202, or VS-5584. In aspects, the PI3K/mTOR kinase inhibitor is dactolisib (NVP-BEZ235).

[0067] In aspects of any of the foregoing methods, the RNP may be introduced into the cells by any suitable method. For example, in aspects the RNP is introduced into the cells by electroporation. In other aspects the RNP may be introduced by viral transduction, lipid nanoparticles, etc.

[0068] In aspects of any of the foregoing methods, the AAV is introduced to the cells immediately after electroporation, or within at least about 5 minutes following electroporation. In aspects, the AAV is introduced into cells between about 5 and 25 minutes following electroporation.

[0069] In aspects of any of the foregoing methods, the Cas protein is a Cas9 protein.

[0070] In aspects of any of the foregoing methods, the AAV is AAV6.

[0071] In aspects of any of the foregoing methods, the cells may be any type of vertebrate cell, preferably any type of mammalian cell. In aspects, the cells are pluripotent stem cells ("PSCs"), including induced pluripotent stem cells ("iPSCs"), or embryonic stem cells ("ESCs"). In aspects, the cells may be hematopoietic stem and progenitor cells, collectively

"HSPCs". The cells may be hematopoietic stem cells ("HSCs"). In aspects, the cells are human cells. In aspects, the cells are murine cells.

[0072] In aspects, of any of the foregoing methods, the cells may be *in vitro*, *ex vivo*, or *in vivo*.

[0073] Also provided by the present disclosure are methods for detecting concatemeric insertions of a viral genome at a target site where a CRISPR system is utilized in combination with adeno-associated virus (AAV) delivery of a single-stranded DNA repair template for producing site-specific genetic modifications in cells or tissues *in vitro*, *ex vivo*, or *in vivo*. The methods combine digital droplet PCR with restriction analysis to determine complex genotypes such as those arising from concatameric insertions.

[0074] Three methods of digital droplet PCR (hereinafter "ddPCR") genotyping are described herein. Each method can detect the basic genotype, *i.e.*, wildtype, monoallelic, or biallelic in subcloned cells, whereas each has varying capabilities to detect more complex genotypes including concatemeric insertions. Throughout the present disclosure "insertions" may also be referred to interchangeably as "knockins".

[0075] Each of the three ddPCR genotyping methods utilize relatively short amplicons, less than 150 basepairs (<150 bp), which allows efficient multiplexing and shorter run times. Up to four primer/probe sets are multiplexed per reaction. Since only two channels are used (FAM and HEX), amplitude multiplexing is necessary. *See* Whale, A.S. *et al.*, *Fundamentals of multiplexing with digital PCR.* Biomolecular detection and quantification, 2016. **10**:15-23 for details of ddPCR multiplexing. The results are rendered in two dimensional plots. In accordance with the methods provided here, samples are analyzed multiple times with different restriction enzymes added directly to the ddPCR reaction, as discussed in more detail below.

Allele counting

[0076] In one aspect, the ddPCR genotyping method includes a copy number variation (CNV) analysis, which measures the concentration of various genetic segments and compares their ratios to determine copies per cell and the corresponding genotype. The method is illustrated by the schematic in FIG. 1A. Since only two channels are used (FAM and HEX), amplitude multiplexing is necessary, as noted above. FIG. 1B shows an exemplary two-dimensional plot of the multiplexed data.

[0077] As illustrated in the figures, four primer/probe sets are multiplexed. The first two primer/probe sets, *Ref1* and *Ref2,* target two different genomic reference sites, which may be in *cis* or in *trans* with the knockin site, also referred to herein as the "target" site or "cut site" with reference to the sequence of cellular genomic DNA cleaved by the RNP. These serve both as a reference for overall cell number and a quantitation control. Concentrations of the reference sets are maintained at a 2:2 ratio due to the diploid character of most genes in mammalian cells. Variation of more than 10% in the ratio may indicate abnormal karyotypes, DNA degradation, and/or improper gating, such that the experiment should be repeated.

[0078] The third "*KI*" is designed to target the insertion sequence, homology arm, or any other part of the viral genome. This primer/probe set can be used to determine if an insertion is concatemeric, off-target, or non-homologous end joining (NHEJ) -mediated.

[0079] The fourth primer/probe set, "*NoKI*" spans the target site such that the primers and probe are at least 25 bp from the cut. This primer/probe set will detect the wild-type (WT) allele or alleles with indels, but will fail if there is a large insertion due to the increased distance between the forward and reverse primers. This primer/probe set is used to determine if a knockin is wildtype, monoallelic, or biallelic. It is important to ensure that larger Cas9-induced deletions (>25 bp) are infrequent at the measured loci prior to employing this method.

[0080] The concentration of each target and corresponding copies/cell can be calculated as shown below.

$$C = (\ln(n_{total}) - \ln(n_{neg})) / 0.00085$$

$$C' = C / (((C_{Ref1} / a_{Ref1}) + (C_{Ref2} / a_{Ref2})) / 2)$$

where

C is volumetric concentration (copies/ul);

C' is cellular concentration (copies/cell);

n_{-total} is total number of ddPCR droplets;

n_{-neg} is number of ddPCR droplets that are negative for a given measured target; and

a is alleles per cell for a given locus (usually 1 or 2).

[0081] The constant 0.00085 indicates the average droplet volume in microliters (μl) used in the ddPCR equipment and may be different for a particular ddPCR machine. The number of alleles per cell for a given locus, a, will be 2 for most experiments or 1 for reactions that utilize IL2RG references in male cells.

[0082] To determine the total number of insertions after performing a knockin, this assay must be utilized in combination with restriction enzymes that separate concatemeric knockins. This is illustrated schematically in FIG. 1C. When using a primer/probe set to target the KI gene segment, extra insertions are indicated by copies/cell greater than 0, 1, or 2 for wildtype, monoallelic, or biallelic genotypes, respectively. However, if reanalyzed without restriction enzymes, the number of measured insertions/cell should decrease if concatemeric insertions exist, as illustrated in FIG. 1D. This assay can be used in this way to measure the number of concatemeric inserts per cell in subcloned samples.

[0083] Compared to the *linkage analysis* assay discussed below, the *allele counting* assay has two advantages. First, the *linkage analysis* assay requires at least two runs, *i.e.*, with and without restriction enzymes, to determine genotype and the total number of insertions. In contrast, the *allele counting* assay only requires a single run, *i.e.*, with restriction enzyme. Thus, it works well as a quick screen to select subcloned cell lines with intended genotypes. Second, the references designed for this assay work in trans, and can be located anywhere on the genome. Therefore, they can be recycled for other assays.

[0084] Accordingly, provided is a method for detecting concatemeric insertions of a viral genome in a target cell, the method comprising performing a first ddPCR reaction, the reaction comprising a target cellular genomic DNA, a first primer/probe set comprising a first reference primer directed to a first genomic reference sites in *cis* or in *trans* with a target site in the cellular genomic DNA; a second primer/probe set comprising a second reference primer directed to a second genomic reference site in *cis* or in *trans* with the target site (this may in some cases be optional); a third primer/probe set comprising a primer directed to a region of the viral genome; and a fourth primer/probe set comprising a primer directed to a cellular genomic region spanning the target site such that the primers and probe are generally at least 25 bp from the target site; performing a second ddPCR reaction comprising the same components as the first ddPCR reaction and at least one restriction endonuclease which cleaves the insert; and comparing the amplicons of each of the primer/probe sets in the ddPCR reactions performed with and without

the restriction endonuclease in order to determine whether concatemeric insertions of the viral genome are present the target cell.

Linkage analysis

[0085] In another aspect, the ddPCR genotyping method includes comparing the linkage of two or more gene segments to determine genotypes. Linkage can be determined by assessing the number of digital PCR droplets that are positive for two or more gene segments targeted by different primer/probe sets. The method utilizes three or four multiplexed primer/probe sets, as illustrated in FIG. 2A.

[0086] The first two primer/probe sets, *Ref1* and *Ref2,* target two different cis reference sites that flank both sides of the intended genomic knockin site, also referred to as the genomic insert site. These primer/probe sets amplify short regions located 0-1000 bp outside the homology arms and should be linked if DNA degradation has not occurred. Similar to the allele counting strategy described above, these primer/probe sets serve as references for overall cell numbers and quantitation controls. Since these primer/probe sets are in cis and are nearby the editing site, these references also serve two additional purposes. First, digital PCR linkage analysis can be used to determine if these references are linked to knockin gene segments to determine genotypes. Second, the two references serve as linkage controls to ensure the DNA is not significantly fragmented or degraded after purification, which would diminish the measured linkage of nearby gene segments. We find this works best if the two references have >75% linkage.

[0087] A third primer/probe set, *KI,* is designed to target knockin sequences or other parts of the viral genome. Linkage analysis can be used to determine the percentage of references that are linked to *KI* and infer complex genotypes.

[0088] Linkage analyses are not bidirectionally equivalent. For example, in the case of a monoallelic knockin of GFP into a human autosome, 100% of GFP will be linked to *Ref1* (written as GFP⇒Ref1 = 100%), but only 50% of Ref1 will be linked to GFP (written as Ref1⇒GFP = 50%). Additionally, since Ref2⇒GFP should be similar to Ref1⇒GFP, the average is usually used to determine Ref⇒GFP.

[0089] To determine the basic genotype, this assay must not be run with restriction enzymes that cut within the knockin sequence or homology arms. Restriction enzymes that cut outside of the reference sites are acceptable, and they may be necessary to homogenize DNA and minimize digital PCR rain. However, this assay should be run a second time for all samples with restriction

enzymes that separate potential concatemers to determine the total number of insertions as determined by allele counting CNV, illustrated schematically in FIG. 2B.

[0090] The calculation for determining the percent linkage of Targ1⇒Targ2 is:

$$\%L_{Targ1 \Rightarrow Targ2} = (1 - ((C_{Targ1\_or\_Targ2} - C_{Targ2}) / C_{Targ1})) * 100$$

where

%L is percent linkage;

C is volumetric concentration (copies/ul); and

$C_{Targ1\_or\_Targ2}$ is determined by treating any droplet positive for Targ1 and/or Targ2 as positive.

[0091] Linkage analysis has multiple advantages over allele counting. First, it is easier to design because there is more latitude for the cis reference primer/probe sets than there is for the *No_KI* primer/probe sets. Second, since the *No_KI* primer/probe set is not needed, it can be replaced by another primer/probe set. This is particularly useful when analyzing double knockins, such as GFP and mCherry because both targets and references can be multiplexed in one reaction. Third, large deletions that would disrupt the *No_KI* primer/probe set would be less disruptive because the references are located hundreds of base pairs away. Fourth, episomal DNA or off-target insertions can be distinguished from on-target insertions by measuring linkage between cis references and insert sequences. This is a unique property of this assay and is particularly useful if long-range PCR or Southern blots are not practical.

[0092] Linkage analysis can also be used to detect ITRs or other viral regions linked to genomic loci, enabling the quantification of %-alleles-with-concatemeric-KI. Therefore, linkage analysis can measure the allelic abundance of concatemers in bulk edited samples without subcloning. This would not work well with ITR CNV because the CNV alone does not distinguish between on- and off-target knockins. Finally, when coupled with restriction enzymes that cut only one side of inserts, linkage analysis can determine orientations and/or sequences of multiple knockins.

[0093] Accordingly, provided is a method for detecting concatemeric insertions of a viral genome in a target cell, the method comprising performing a ddPCR reaction, the reaction comprising a target cellular genomic DNA, a first primer/probe set comprising a first reference

primer directed to a first genomic reference site in *cis* with a target site in the cellular genomic DNA; a second primer/probe set comprising a second reference primer directed to a second genomic reference site in *cis* with the target site (this may be optional in some cases); a third primer/probe set comprising a primer directed to a region of the viral genome; and a fourth primer/probe set comprising a primer directed to a region of the viral genome that may not be intended to be inserted into the target genomic site such as the ITR or regions outside the homology arms; and comparing the linkage of amplicons of the primer/probe sets to determine the ratio of viral inserts knocked in (or linked) to the target site, as well as potentially unintended insertions such as non-HDR mediated insertions or concatemeric inserts. Viral inserts linked to the genomic references indicate a knockin has occurred. ITRs (or other viral regions that are outside the homology arms) that are linked to the genomic references indicated that a non-HDR mediated insertions or concatemeric insertions have occurred. A second ddPCR reaction can be run with restriction enzymes that cut specific regions within the viral genome. By comparing linkage before and after restriction enzyme usage, distinguishment can be made between concatemeric insertions and non-HDR mediated insertions.

Nuclease-mediated loss of linkage

[0094] In another aspect, the ddPCR genotyping method includes utilizing nucleases including restriction endonucleases or a Cas9 endonuclease, or similar, to recognize and quantify gene segments. Cis reference primer/probe sets are designed for linkage analysis as described above. The ddPCR reaction is then run with and without nucleases that target unique sites within knockins or other parts of the viral genome. If the recognition site exists between the reference primer/probe sets, linkage between the two reference sites will be lost. This can be quantified by the equation below:

$$\%\text{alleles}_{\text{Target KI}} = 1 - (\%L_{\text{with nuclease}} / \%L_{\text{without nuclease}})$$

[0095] This strategy has several unique advantages. First, only two primer/probe sets are used, lowering cost and complexity. Instead of *KI* or *No_KI* primer/probe sets, nucleases such as restriction enzymes are used, which are readily available. PCR-resistant sequences, such as sequences with inhibiting secondary structures, do not have to be efficiently amplified by primer/probe sets to be detected. Finally, the assay is agnostic to episomal viral DNA and can therefore detect knockins or concatemers shortly after editing without waiting for episomal DNA to dissipate. Although linkage analysis can also distinguish between episomal and on-target

sequences, ddPCR reactions can easily be overloaded and pushed outside of their dynamic range when too many episomal copies of DNA exist, such as shortly after editing.

[0096] Accordingly, provided is a method for detecting concatemeric insertions of a viral genome in a target cell, the method comprising performing a ddPCR reaction, the reaction comprising a target cellular genomic DNA, a first primer/probe set comprising a first reference primer directed to a first genomic reference site in *cis* with a target site in the cellular genomic DNA; a second primer/probe set comprising a second reference primer directed to a second genomic reference site in *cis* with the target site; and comparing the linkage of amplicons of the primer/probe sets to when run with and without endonucleases such as restriction enzymes. The enzymes can usually be directly added to the ddPCR reaction during preparation, minimizing workload. The restriction enzymes will target DNA sites that are between the two *cis* reference amplicons. The unique sequence detected by a restriction enzyme can then bet detected by loss of linkage between the two reference sites. As such, inserts can be quantified without the need to PCR amplify across the interrogated insert.

EXAMPLES

[0097] The experiments below describe the development of strategies to detect and thoroughly characterize unintended concatemeric viral insertions in CRISPR/Cas9-AAV systems, highlighting their abundance in multiple genes and clinically important stem cell types. Also described is the development of methods that disable concatemer formation without disrupting already established gene-editing pipelines.

Additional insertions detected by digital PCR

[0098] Since Cas9 RNP electroporation followed by AAV6 transduction enables high efficiency editing, knockins can be created without selection, and purified populations can be isolated by subcloning and genotyping. Employing this method, CRE was knocked-in to human iPSCs at the 3' end of the *CD14* locus, and 36 subclones were expanded. DNA was extracted from the subclones and analyzed by three common PCR methods to determine if the knockins were biallelic, monoallelic, or wildtype. 3-primer in-out PCR consistently reported the same genotype when analyzing either the left (PCR-L) or right (PCR-R) side of the editing site. However, 2-primer-PCR, which is designed to amplify the full region on both sides of the editing site (PCR-F), was not concordant with PCR-L and PCR-R. In 8/36 samples, PCR-F was missing the upper (*KI*: knockin) band shown in PCR-L and PCR-R. To further elucidate the genotype, an alternate

method was developed, which uses droplet digital PCR (ddPCR) to count wildtype alleles (*no KI*) and CRE (*KI*) alleles, as discussed above and illustrated in FIG. 1A and FIG. 1B. Since each cell contains two copies of the *CD14* locus, the sum of *no KI* and *KI* alleles should equal two per cell. Surprisingly, as shown in FIG. 3A, this strategy found that more than half of the edited cells contained additional copies of CRE at integer intervals. Additional copies were detected in all samples with discordant PCR analyses.

Additional insertions are concatemers

[0099] As is commonplace, the ddPCR analysis in FIG. 3A was performed with restriction enzymes to fragment the genomic DNA in a controlled manner. When rerunning the reaction without fragmentation, the number of detected additional insertions was greatly diminished in most of the subclones, as shown in FIG. 3B. Since ddPCR is similar to limiting dilution assays, linked regions of the genome partition together in the same droplet when the DNA is not fragmented. Therefore, concatemeric knockins would only be counted once in non-fragmented DNA, as illustrated in FIG. 1C and FIG. 1D. This suggests that the additional insertions are concatemeric in all subclones except #12.

[0100] Southern blots can measure the size of large regions of DNA without PCR amplification. Since the concatemers were not readily detected by PCR, 11 subclones with various ddPCR genotypes were further expanded and analyzed by Southern blot. As shown in FIG. 3C, the Southern blot revealed that many samples had larger knockin-bands at regular intervals. As expected, most samples had only one or two bands, however sample #12 displayed a third band of irregular size, which corresponds to the off-target insertion detected by ddPCR. Importantly, the Southern blot analysis was 100% concordant with the ddPCR allele counting strategy. Collectively, these data suggest that concatemeric knockins frequently occur when using Cas9/AAV-mediated genome editing.

[0101] These concatemers are difficult to detect by classic PCR genotyping strategies, but as shown in FIG. 3D, they are revealed by Southern blots and ddPCR after controlled genomic fragmentation.

[0102] Since Southern blots generally require multiple days of preparation and consume >1000 fold more starting material than ddPCR, ddPCR analyses were used for subsequent genotyping. Repeat experiments were performed at the *AAVS1* locus and the genotype analyzed by PCR and ddPCR. Results similar to those at the *CD14* locus were seen at the *AAVS1* locus

with concatemeric insertions occurring in 39% of *CD14* clones and 36% of *AAVS1* clones. FIG. 4A shows gel electrophoresis from PCR genotyping. 36 subclones were expanded after knocking in Ubc-CRE into *AAVS1* locus in PT-iPSCs. Ladder size shown in kb. Sample ID indicated at top and bottom of gels; L = ladder. Interpreted genotype indicated by circle at top of each gel: white = WT; gray = monoallelic; black = biallelic; red = cannot be determined. Top, middle, and bottom row of gel indicate PCR-L, PCR-R, and PCR-F genotyping strategies, respectively.

[0103] FIG. 4B shows Ref-1 and Ref-2 from ddPCR. Average of the two references set to 2 copies/cell. Sample ID shown at bottom.

[0104] FIG. 4C shows ddPCR genotyping results. Normalized to references shown in FIG. 4B. The reaction was run +/- restriction enzyme (RE). Interpreted genotype indicated by circle at bottom graph: white = WT; gray = monoallelic; black = biallelic; red = additional insertions or cannot be determined.

Concatemers contain ordered repeats of full viral genome

[0105] The band sizes on the *CD14* Southern blot were measured to determine the length of the concatemers, shown in FIG. 5A. Bands were spaced at concise intervals, with an average distance of 2.6 kb between groups of concatemeric bands.

[0106] Since this is the same size as the full AAV6 viral genome used to deliver the knockin template, multiple insertions could result from linked viral genomes as shown in FIG. 5B. PCR was attempted across the potential head-to-tail concatemeric junction. Indeed, bands only appeared in samples which contained concatemeric knockins. However, the PCR was inefficient and resulted in two weak bands. Sanger sequencing revealed that the bottom band was missing the ITRs entirely, containing a perfect linkage of opposing left and right homology arms. Sanger sequencing failed on the upper band, so it was sequenced with nanopore technology instead. This revealed a unique chimeric-ITR composed of two back-to-back ITRs missing their distal ends. Reamplification of gel-extracted upper and lower bands resulted in only the lower bands. This suggests that the lower band is an artifact of PCR that occurs with an increasing number of PCR cycles. No bands were seen when using primer combinations to detect head-to-head or tail-to-tail concatemers. Collectively, these data suggest that (1) concatemeric knockins are joined by the viral ITRs, which creates a distinct chimeric ITR, (2)

PCR amplification and subsequent sequencing across the chimeric ITR is difficult, and (3) the concatemers primarily exist in a head-to-tail orientation.

[0107] We next sought to test the regularity of the concatemeric knockins. Since the full length concatemeric KI cannot be PCR amplified and directly sequenced, ddPCR was used to detect the stoichiometry of DNA segments corresponding to regions of the viral genome. For each additional CRE, there was also an additional left homology arm and right homology arm detected at a 1:1 ratio, shown in FIG. 5C. This was true for all samples except #12, which had a partial off-target insertion verified by Southern Blot, and sample #14, which had an end-joining mediated on-target knockin verified by PCR and sequencing. Interestingly, ddPCR failed to detect the ITR in all samples except #14 (see FIG. 6D, ITR -RE), suggesting that the chimeric ITR formed in concatemeric knockins is more PCR-resistant than the ITR found in end-joining mediated knockins that have been previously reported.

[0108] Since PCR inefficiencies can be due to nearby DNA secondary structures, ITR quantification was repeated with restriction enzymes that cut within the chimeric ITR, adjacent to the amplicon. This strategy led to more efficient qPCR and ddPCR detection of the ITR (see FIG. 6C and FIG. 6D), revealing that two ITR regions exist for each additional insertion of CRE (FIG. 5C), which agrees with the schematic shown in FIG. 5B. To ensure that the detected ITRs were not episomal, a ddPCR linkage analysis was performed to determine whether the genomic locus was linked to the ITRs (FIG. 6E). All loci with concatemers had ITRs linked to one or both CD14 alleles (FIG. 6F). These data prove that the concatemeric knockins have ITR sequences that are integrated into the genomic locus, and these integrated chimeric ITRs are hidden without careful restriction enzyme digestion.

Concatemers affect gene expression

[0109] If ITRs are the drivers of concatemerization, then concatemeric knockins should occur between two viral particles carrying different DNA sequences. To test this, a dual-knockin of mCherry and GFP was attempted in iPSCs at the *TET2* locus, as illustrated in FIG. 7A. Five days after editing, the cells were analyzed and sorted by flow cytometry for subcloning. The single positive quadrants (*i.e.*, mCherry only, GFP only) appeared bimodal, which likely corresponds to a monoallelic or biallelic knockin of a single color. Interestingly, the double positive quadrant had a large, polydisperse region. This heterogenous gene expression could be due to concatemeric knockins. 24 colonies were sorted and expanded from the dimmer, more homogeneous cluster, and 48 colonies were sorted and expanded from the brighter, polydisperse cluster. ddPCR linkage

analysis was used to analyze the genotype, revealing that 46/48 *high gate* subclones contained concatemers. In contrast, only 4/24 *low gate* subclones contained concatemers. Fluorescence imaging revealed that *high gate* subclones were brighter and had more well-to-well fluorescence variability than the *low gate* subclones, which was made more apparent with uniform image processing of the entire set of subclones. The mean fluorescence intensity correlated strongly with the total number of insertions, as shown in FIG. 7B and FIG. 7C. Linkage analysis revealed that most of the *low gate* subclones with concatemers had a monoallelic knockin genotype, such that both mCherry and GFP were inserted into a single allele, while the other allele did not contain a knockin. This is particularly problematic for gene-editing strategies that employ dual-knockins to achieve a selectable biallelic genotype because some cells will still contain an unedited allele. Overall, 50/72 subclones had concatemeric knockins, of which 21/50 had at least one mCherry and one GFP linked in the same allele. These data show that concatemeric knockins can also occur between viral particles carrying different DNA. Importantly, concatemeric knockins may greatly change the level of gene expression.

ITR removal prevents concatemers in PSCs

[0110] If the ITRs are driving the concatemeric knockins, then removal of ITRs should result in more monomeric knockins, as illustrated in FIG. 8A. Two strategies were designed that use Cas9 RNPs to remove the viral ITRs post-transduction with minimal modification to gene-editing pipelines.

[0111] For the first strategy, a sgRNA was designed to make an internal cut (IC) within the ITR at an endogenous PAM site, as illustrated in FIG. 8B. This sgRNA is complexed with Cas9 and electroporated into cells at the same time as the RNP that cuts the target gene. For the second strategy, the 23 bp target sequence recognized by the RNP that cuts the genomic locus was inserted into the viral genome on both ends flanking the viral homology arms, such that the gene-targeting RNP would also make two distal cuts (DC) in the viral genome and remove the ITRs. IC and DC methods were compared to the original method (no cut, NC) in three PSC lines and four loci, including the clinically relevant *HBB* locus. For all conditions, double fluorescent knockins were selected as previously shown in FIG. 7A. Flow cytometry revealed a clear decrease in polydispersity when using IC or DC methodology, particularly in the double positive quadrant. The overall knockin rate (i.e., %fluorescent cells) for NC ranged from 27% to 79%, which is similar to previously published values. IC and DC treatments resulted in minor decreases in average overall editing efficiencies by 19% and 9%, respectively; however,

the reduction in the double positive cells was more pronounced (42% and 23% decreases, respectively.

[0112] From each condition, illustrated in FIG. 8C, 16 to 28 subclones (458 total) were sorted and expanded from the double positive population.

[0113] ddPCR genotyping revealed that an average of 59% of the subclones in the NC group had concatemeric knockins, as shown in FIG. 8D. This decreased to 13.2% and 4.8% for the IC and DC groups, respectively. These data show that concatemeric knockins occur at high frequency in human PSCs, regardless of the cell line or locus. Importantly, post-transduction removal of the viral ITRs significantly reduces the frequency of concatemeric knockins.

Rate of concatemers unaffected by MOI

[0114] It is known that the multiplicity of infection (MOI) will affect knockin rate, but how MOI affects the rate of concatemeric knockins is unknown. For this analysis, UBC-GFP was knocked in to the *HBB* locus at various MOI dilutions ranging from ~3000 to 10. To avoid the high sampling error that can be prevalent in small-population-size clonal analyses, a ddPCR linkage-analysis strategy was developed to detect concatemeric alleles in a bulk, mixed population.

[0115] As expected, and as illustrated in FIG. 8E, flow-cytometry revealed that the percent of positive cells decreased from 56% to 2% as MOI was reduced. The positive cells were then sorted and expanded in each MOI group without subcloning. Bulk allele analysis showed that ~35% of the knockin alleles were concatemeric at all MOIs. This was repeated in male PSCs at the *IL2RG* locus, located on X chromosome, which yielded similar results. In both cases, the concatemeric insertions contained an average of 3-4 repeats at all MOIs. Collectively, these data suggest that the concatemeric knockin rate is constant in human PSCs, regardless of locus, cell-line, MOI, or number of genomic target sites. If an allele has a knockin, there is a 35% chance that it is concatemeric. Thus, the theoretical frequency of at least one concatemeric knockin occurring in a biallelically edited PSC is 58%. This matches our empirical result from FIG. 8D.

ITR removal does not increase NHEJ knockin

[0116] ITR removal changes the structure of the viral genome, potentially altering its dynamics regarding the frequency of non-homologous end joining (NHEJ)-mediated insertions. This could cause imprecise knockins. To analyze this, in-out PCR was used to check for NHEJ-mediated knockins at the *TET2* locus in iPSC subclones with seemingly correct ddPCR genotypes (i.e.,

biallelic knockins without concatemers). All samples except for one subclone from the IC group had normal band sizes. This indicates that NHEJ-mediated knockin events are uncommon in all groups, and subcloning is not an efficient method to measure their frequencies. As an alternative, flow-cytometry was used to measure NHEJ-mediated insertions by repeating *HBB* and *IL2RG* knockin experiments while intentionally mismatching the Cas9 RNP genomic target sites and viral homology arms. When making a cut at the *HBB* locus, a virus that carried UBC-GFP with homology arms targeting the *IL2RG* locus was added. The reciprocal experiment was also performed. When cutting the *HBB* locus, insertions occurred in 0.53% and 0.50% of the cells in NC and IC groups, respectively. When cutting the *IL2RG* locus, insertions occurred in 0.15% and 0.08% of the cells in the NC and IC groups, respectively. These data suggest that ITR removal does not increase the frequency of unwanted NHEJ-mediated knockins.

Concatemers also occur in HSPCs

[0117] CD34+ blood cells contain both hematopoietic stem and progenitor cells (HSPCs). Although it is the stem cell population (HSCs) that has the capacity to repopulate all linages of the blood, HSCs are difficult to enrich from HSPCs. As such, HSPCs are the targets of multiple gene therapies that ultimately aim to cure all genetic diseases of the blood.

[0118] To determine if concatemeric knockins occur in HSPCs, Cas9/AAV6-mediated double knockin experiments were repeated in human cord-blood-derived CD34+ HSPCs. Since clonal HSPC expansion is inefficient and the DC method appeared more effective than the IC method, only NC and DC groups were compared. When performing knockins at the *TET2* and *HBB* loci, results were similar to those of PSCs, with an average of 58% and 7.0% of the subclones having concatemeric knockins in the NC and DC groups, respectively. The experiment was repeated with male cord blood at the *IL2RG* locus (X chromosome), resulting in 38% and 1.4% of colonies containing concatemeric knockins in the NC and DC groups, respectively.

[0119] Only a subset of the CD34+ HSPCs can enable long-term reconstitution of the blood after bone marrow transplantation. Therefore, it is necessary to determine whether concatemeric insertions occur in the functional stem cells or just the progenitor cells. To test this, UBC-GFP was knocked into human CD34+ cells at the *HBB* locus, and GFP+ cells were transplanted into immunodeficient mice a few days later. After 4 months, bone marrow was collected from the mice, and the human cells were isolated by flow cytometry. Since transplanted blood progenitor cells generally exhaust after 12 weeks in mice, the remaining human cells were likely derived from engrafted human HSCs. As shown in FIG. 9, bulk ddPCR analysis revealed that an average

of 32% of the edited human alleles had concatemeric knockins in the 4 mice from the NC group. This dropped to below 5% for the DC group. Notably, the average engraftment rate of edited cells was 0.22% and 2.5% for NC and DC groups, respectively, suggesting that the DC approach did not impair HSC engraftment. Collectively, these data demonstrate that Cas9/AAV6-mediated concatemeric knockins also occur at high frequency in human HSCs. Like in iPSCs, this can be minimized by post-transduction ITR removal.

[0120] Surprisingly, inclusion of a PI3K/mTOR kinase inhibitor substantially reduced the presence of concatemeric knockins without diminishing the total number of cells carrying the knockin sequence. Ubc-GFP was knocked into the *HBB* locus in PT-iPSCs under eight different conditions including treatment with 300 nM PI3K/mTOR kinase inhibitor (NVP-BEZ235) for 24 hours after electroporation. About 5 days later, GFP positive cells were sorted and expanded for a week, after which the percentage of edited alleles containing a concatemer were quantified. As shown in FIG. 10, inclusion of inhibitor decreased concatemer formation by about 2.7 fold. Combining the inhibitor with either of the ITR removal strategies described herein, IC, DC, or a combination of both IC and DC, resulted in a greater reduction in concatemeric insertions compared to any method alone.

ITR removal decreases unwanted concatemeric knockins and increases HSC engraftment

[0121] Gene modified human hematopoietic stem cells (HSCs) can be transplanted into patients to cure diseases. However, while making genetic knockins in HSCs with Cas9 and adeno-associated virus (AAV) vectors, engraftment efficiency drops. Thus, it is not always possible to get enough gene-edited cells to achieve desired results.

[0122] Experiments were performed to compare engraftment efficiency after removing the AAV inverted terminal repeat (ITR), a genetic region in the AAV genome contemplated herein to be responsible for causing genotoxic effects and thus decreasing HSC engraftment.

[0123] Human HSPC were electroporated/transduced with Cas9 and AAV targeting the *HBB* locus. AAV and Cas9 both travel to the cell nucleus. There, Cas9 makes a cut in the host genomic DNA and the AAV provides the single-stranded DNA repair template. In the DC condition, the Cas9 recognition site is inserted next to the AAV ITRs, such that Cas9 makes a cut in the host genomic DNA and also removes the viral vector ITRs. Three days after electroporation/transduction, GFP positive edited $CD34^+$ cells were sorted (2e5 to 3e5 cells) and transplanted into one femur of sub-lethally irradiated mice (200 rad, 2 to 24 hours before

transplant). 4 months after transplantation, the human cell-engrafted mice were sacrificed, and all bone marrow was harvested by crushing the bones. Non-specific antibody binding was blocked (human and mouse Fc block, BD, USA) and stained (30 min, 4 °C, dark) with PE-Cy7-conjugated anti mouse CD45.1 antibody (A20, Invitrogen, USA), PE-Cy5-conjugated anti mouse TER-119 antibody (TER-119, Invitrogen, USA), V450-conjugated anti human CD45 antibody (HI30, BD Horizon), and PE-conjugated anti HLA-ABC antibody (W6/32, eBioscience), and analyzed by flow cytometry.

[0124] Figure 12 shows the removal of the ITRs after transduction (DC condition) increased engraftment by approximately 10x over the control [no cut (NC) condition—ITR left in place]. Therefore, ITR removal not only decreases unwanted concatemeric knockins, it also increases HSC engraftment.

Experimental Methods Detail

Digital droplet PCR genotyping

[0125] Each ddPCR reaction was prepared and analyzed with the QX200 ddPCR system (BioRad, USA) and ddPCR™ Supermix for Probes (No dUTP) per BioRad's standard recommendations unless otherwise stated. All reactions were mixed to 25 ul and contained up to 1-5 ul of DNA. 2.5-5 U of restriction enzyme was frequently added, and digestion was performed immediately before droplet generation at 37 °C x 5 minutes. For droplet generation, 20 ul were loaded into the droplet generator cassette in groups of eight per BioRad's protocol. Thermocycler conditions: 95 °C x 10 minutes; 50 cycles of 94 °C x 30 sec and 60 °C x 60 sec; 98 °C x 10 minutes; hold at 4°C. Cluster identification was performed manually with QuantaSoft Analysis Pro version 1.0.596 (BioRad, USA). Cluster information was exported into Microsoft excel for further analysis. Forward primer, reverse primer, and probe were at a 3.6:3.6:1 ratio.

[0126] To analyze genotype after knocking in CRE at the *CD14* locus in human iPSCs (results shown in FIG. 3A, FIG. 3B, and FIG. 3D), the ddPCR *allele counting* strategy was used. Two trans references were multiplexed with a *CD14_No_KI* primer/probe set and a *CRE* primer/probe set. The reaction was run with and without a restriction enzyme that separates concatemers. $C'_{CD14\_No\_KI}$ rounding to 0, 1, or 2 was scored as biallelic, monoallelic, and wild-type, respectively. Extra insertions were determined by the equation:

$$C'_{Extra\_CRE} = C'_{CRE\_+RE} - (2 - C'_{CD14\_No\_KI})$$

[0127] To count the number of LHA and RHA gene segments after knocking in *CRE* at the *CD14* locus in human iPSCs (results shown in FIG. 5C, FIG. 6A, and FIG. 6B), the ddPCR *allele counting* strategy was used. Two *CD14* cis references were multiplexed with primer/probe sets targeting LHA and RHA. The reaction was run +/-RE. Extra LHA and RHA insertions were accounted for by subtracting 2 from their measured copies/cell in the +RE reaction. Viral ITRs inserted into the genome were also quantified. These reactions were run with the same two CD14 cis references, a CRE primer/probe set, and a primer/probe set targeting the viral ITR region. The reaction was run three times (with no RE, with AhdI, and with MseI and AhdI), and analyzed by allele counting (FIG. 5C and FIG. 6D). To determine if concatemeric knockins occurred in one or two alleles, the dataset was analyzed by *linkage analysis*. Specifically, CD14_Ref1⇒ITR and CD14_Ref2⇒ITR were measured on the samples run with AhdI (FIG. 6F, upper and middle row). Genomic ITR insertions were further validated with *nucleus mediated loss of linkage*, by analyzing CD14_Ref1⇒CD14_Ref2 on the same +AhdI dataset (FIG. 6F, bottom row). This method showed that sample #14 had and at least one ITR inserted into the target site, but its orientation (i.e., lack of linkage to references when adding AhdI) suggests it was an NHEJ mediated insertion. This is further supported by *allele counting*, which shows the samples has an addition LHA and RHA but only one copy of CRE.

[0128] All subclones (both PSCs and HSPCs) with knockins of UBC-mCherry and/or UBC-GFP were genotyped by ddPCR *linkage analysis*. Each loci had two cis references designed as described above, which were multiplexed with UBC-mCherry and UBC-GFP primer/probe sets. Each reaction was run with and without restriction enzymes that separate concatemeric knockins. For basic genotyping calculations, linkage analysis was performed by pooling UBC-mCherry and UBC-GFP clusters into a single *insert* population. Since linkage analysis is unidirectional, the average linkages were determined by:

$$\%L_{Avg\_RefàRef} = (\%L_{Ref1àRef2} + \%L_{Ref2àRef1}) / 2$$

$$\%L_{Avg\_RefàIns} = (\%L_{Ref1àIns} + \%L_{Ref2àIns}) / 2$$

[0129] A correction can be made for DNA degradation:

$$\%L_{Corr-\_Avg\_RefàIns} = \%L_{Avg\_RefàIns} / \%L_{Avg\_RefàRef} * 100$$

[0130] KI and No KI alleles were calculated from the digital PCR run -RE:

$C'_{KI} = \%L_{Corr\_Avg\_RefàIns} * a$

$C'_{No\_KI} = a - C'_{KI}$

[0131] Basic genotypes were determined by rounding C'_{No_KI} to the nearest integer. 0, 1, and 2 indicated biallelic, monoallelic, and WT genotypes respectively.

[0132] All runs were reanalyzed with +RE, and the total copies/cell of GFP and mCherry were calculated as described in *allele counting*. Runs that did not have the following quality parameters were excluded: total droplet number > 8000; the concentration of ref1 is no more than 10% different than the concentration of ref2; average concentration of ref1 and ref2 > 12 copies/ul; $\%L_{Avg\_RefàRef}$ > 75% without fragmenting restriction enzyme and < 15% with fragmenting restriction enzyme; and the copies/ul of GFP and mCherry without fragmentation is not 20% greater than with restriction enzyme.

[0133] If both runs (+/-RE) passed quality control, more complex genotypes were determined. Off-target and concatemeric insertions were determined by rounding each value in the following equations to the nearest integer:

$C'_{Off\text{-}targ\_Ins} = C'_{Ins\_\text{-}RE} - C'_{KI}$

$C'_{Concat\_Ins} = C'_{Ins\_+RE} - C'_{KI} - C'_{Off\text{-}targ\_Ins}$

[0134] Linkage between GFP and mCherry was determined to be true if the sum of $\%L_{GFPàmCh}$ and $\%L_{mChàGFP}$ was > 20% when analyzed -RE.

[0135] For bulk genotyping, GFP was knocked into the target locus, and DNA was extracted from a group of GFP positive cells for analysis. This is distinct from previous genotyping analyses, which had all been performed on DNA extracted from a pure, clonal population. Bulk genotyping was performed on the viral dilution series in PSCs (FIG. 8E) and the human cells recovered after HSPC xenotransplantation (FIG. 9). For these experiments, the viral backbone was modified such that two unique DNA sequences <100 bp were inserted flanking the homology arms, as shown in FIG. 2C. Primer/probes sets were designed to detect these two sequences termed ID1 and ID2. The bulk samples were first analyzed -RE with cis reference primer/probe sets and the GFP primer/probe set, then reanalyzed with the No_KI primer/probe set. These datasets were used to calculate percent KI alleles using both *linkage analysis* and *allele counting*:

$\%KI = \%L_{Corr\_Avg\_RefàGFP}$

$$\%KI = (1 - (C_{No\_KI} / ((C_{Ref1} + C_{Ref2}) / 2))) * 100$$

[0136] To detect on-target concatemeric knockins, the samples were analyzed -RE with the cis references and ID primer/probe sets. The percent of KI alleles with concatemeric insertions was determined by:

$$\%L_{Avg\_RefàID} = (\%L_{Ref1àID1\_or\_ID2} + \%L_{Ref2àID1\_or\_ID2}) / 2$$

$$\%L_{Corr\_Avg\_RefàID} = \%L_{Avg\_RefàID} / \%L_{Avg\_RefàRef}$$

$$\%KI_{Concat} = \%L_{Corr\_Avg\_RefàID} / \%KI * 100$$

[0137] To determine the average size of concatemers, the samples were run with the cis references and GFP primer/probe sets +RE. The average concatemer size (including the first insert) was calculated by:

$$Avg\_concat\_size = ((C'_{GFP\_+RE} - C'_{GFP\_-RE}) / \%KI_{Concat}) + 1.$$

PSC lines and maintenance

[0138] Four human PSC lines were used in this study. Patient-derived iPSC (PT-iPSC, iSU223n)[33, 34], H9 ESC (WiCell, USA), H1 ESC (WiCell, USA), and wild-type iPSC (WT-iPSC, SUN004.1)[35] and their derivative subclones were routinely propagated feeder-free in StemFit® Basic02 medium supplemented with 20 ng/ml bFGF or Basic04 complete type (Ajinomoto, Japan) on cell culture plastics with iMatrix-511 (Nippi, Japan) basement membrane matrix, with single-cell dissociation by using TrypLE™ Express Enzyme (Gibco, Thermo Fisher Scientific, USA). Y-27632 dihydrochloride (TOCRIS, UK) was added for 24 hours after passage unless otherwise indicated. PT-iPSCs were used unless another PSC line is indicated in the text or description of the figures.

AAV vector production

[0139] AAV vector plasmids were cloned in the pAAV-MCS plasmid containing ITRs from AAV serotype 2 (AAV2)[21]. CD14 vector contained P2A and Cre. AAVS1 vectors contained a UBC promoter and CreERT or fluorescence reporter genes such as GFP or mCherry. HBB, TET2, RUNX1, and IL2RG vectors contained UBC promoters and fluorescence reporter genes such as GFP or mCherry. The homology arms for CD14, AAVS1, TET2, IL2RG, ASXL1, and CCR5 were approximately 400 bp (379-414 bp), and the left and right homology arms for HBB were 537 bp and 420 bp, respectively. Vectors with distal cut sites (DC) for ITR removal contained

the same sequences as genomic sgRNA target sequences, with the external PAM (ExP) facing inward between the ITR and homology arm. AAV6 particles were produced in 293FT cells transfected using standard PEI transfection with ITR-containing plasmids and pDGM6 containing the AAV6 cap genes, AAV2 rep genes, and adenovirus five helper genes. Particles were harvested after 72 hours, purified using the AAVpro Purification Kit (Takara Bio, Japan) according to manufacturer instructions, and then stored at -80 °C until further use. Viral titer measured as vector genomes/cell was determined by ddPCR using primers and probes that target the viral ITR in a similar manner as previously described by Aurnhammer, C., *et al.*, Human Gene Therapy, Part B: Methods, 2012. **23**(1): p. 18-28.

Electroporation and transduction of cells

[0140] All synthetic sgRNAs and Cas9 protein were purchased from Integrated DNA Technologies (Coralville, IA, USA). The genomic sgRNA target sequences with PAM were as follows:

HBB, 5'-CTTGCCCCACAGGGCAGTAACGG-3' (SEQ ID NO: 1);

IL2RG, 5'-TGGTAATGATGGCTTCAACATGG-3' (SEQ ID NO: 2);

RUNX1, 5'-TACCTTGAAAGCGATGGGCAGGG-3' (SEQ ID NO: 3);

AAVS1, 5'-GGGGCCACTAGGGACAGGATTGG-3' (SEQ ID NO: 4);

TET2, 5'-TCATGGAGCATGTACTACAATGG-3' (SEQ ID NO: 5); and

CD14, 5'-CTAGCGCTCCGAGATGCATGTGG-3' (SEQ ID NO: 6).

[0141] Cas9 RNP was made by incubating protein with sgRNA at a molar ratio of 1:2.5 at 25 °C for 10 min immediately prior to electroporation into $CD34^+$ HSPCs or iPSC/ESC, with a final concentration of Cas9 at 150 ng/μl during electroporation. For experiments employing the IC method of ITR removal, sgRNA of ITR sequence (5'-GCGCGCTCGCTCGCTCACTGAGG-3'; SEQ ID NO: 7) was separately incubated with Cas9 and added at the same final concentration in addition to the other RNP. Both $CD34^+$ HSPCs and iPSC/ESC were electroporated using the Lonza Nucleofector 4D (program DZ-100 for $CD34^+$ HSPCs and program CB-150 for iPSC/ESC). Immediately after electroporation, AAV6 donor vectors were added at an MOI (vector genomes/cell) of 10,000 for $CD34^+$ HSPCs, 5,000 for PT-iPSC, and 1,000 for WT-

iPSC/ESC, unless indicated otherwise in figures. After overnight incubation, cells were washed. For iPSC/ESC, Y-27632 dihydrochloride was added to media 48 hours after electroporation.

Flow cytometry and subcloning

[0142] Cells were washed and stained with propidium iodide (PI) at a final concentration of 1 μg/ml or DAPI at a final concentration of 0.1 μg/ml right before analysis or sorting. Cell analysis and cell sorting were performed on a FACSAria II (BD, IN, USA). To subclone edited PSCs, five days after electroporation, DAPI negative and fluorescence positive cells (just DAPI negative cells for no-marker vector in CD14 P2A-Cre and AAVS1 UBC-CreERT2) were sorted into 96-well plates. After 7 days, the subclones were further expanded in 12-well plates. Edited PSCs that were not subcloned were sorted five days after electroporation into 12-well plates, followed by a second sort ~2 weeks later to ensure purity. For edited $CD34^+$ cells, three days after electroporation, DAPI negative and fluorescence positive cells were sorted into 150 μl myeloid differentiation media consisting of MyeloCult H5100 (Stemcell Technologies, Canada) supplemented with SCF (20 ng/ml), TPO (20 ng/ml), Flt3-Ligand (20 ng/ml), IL-6 (20 ng/ml), IL-3 (20 ng/ml), GM-CSF (20 ng/ml) and G-CSF (20 ng/ml). ~3 weeks later, large fluorescent colonies were used for subsequent analysis.

DNA extraction

[0143] For ddPCR, qPCR, and standard PCR analysis, DNA was extracted using a crude lysis buffer containing 0.1% SDS, 5 mM EDTA, 0.2 mg/ml proteinase K (Thermo Fisher Scientific, Waltham, MA), 15 mM tris (pH 8.2), and 100-200 mM NaCl. For PSCs, 100 ul lysis buffer was added to approximately 1e5 to 1e6 pelleted cells, followed by incubation at 55 °C for 10 minutes and heat inactivation at 80 °C for 10 minutes. For HSPCs, 35 ul lysis buffer was added to approximately 5e3 to 1e5 cells, followed by the same heating protocol. All samples were homogenized by pipet trituration until smooth, and debris was pelleted through centrifugation at 6000 g for 3 min.

[0144] For Southern blots, genomic DNA was purified from approximately 1e7 cells using multiple columns from the DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany) and following the recommended protocol with RNase A.

Cord blood processing

[0145] Frozen mononuclear cells derived cord blood (CB) were purchased from New York Blood Center. $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) were isolated using a human CD34 MicroBead Kit (Miltenyi Biotec, San Diego, CA, USA). $CD34^+$ HSPCs were cultured in HSPC expansion media consisting of StemSpan SFEM II (Stemcell Technologies, Canada) supplemented with SCF (20 ng/ml), TPO (20 ng/ml), Flt3-Ligand (20 ng/ml), IL-6 (20 ng/ml) and UM171 (35 nM). Cells were cultured at 37°C, 5% CO_2, and 5% O_2.

HSPC transplantation

[0146] Three days after electroporation/transduction, GFP positive edited $CD34^+$ cells were sorted (2e5 to 3e5 cells) and transplanted into one femur of sub-lethally irradiated mice (200 rad, 2 to 24 hours before transplant). 4 months after transplantation, the human cell-engrafted mice were sacrificed, and all bone marrow was harvested by crushing the bones. Non-specific antibody binding was blocked and stained (30 min, 4 °C, dark) with PE-Cy7-conjugated anti mouse CD45.1 antibody (A20, eBioscience), PE-Cy5-conjugated anti mouse TER-119 antibody (TER-119, eBioscience), V450-conjugated anti human CD45 antibody (HI30, BD Horizon), and PE-conjugated anti HLA-ABC antibody (W6/32, eBioscience), and analyzed by flow cytometry. For cell sorting, the engrafted human cells were isolated using human CD45 MicroBeads (Miltenyi Biotec, USA). The enriched human cells were stained (30 min, 4 °C, dark) with V450-conjugated anti human CD45 antibody and PE-conjugated anti HLA-ABC antibody, isolated by flow cytometry, and analyzed by digital PCR.

Mice and animal care

[0147] All mouse experiments were conducted according to a protocol approved by the Institutional Animal Care and Use Committee (Stanford Administrative Panel on Laboratory Animal Care #22264) and adhering to the U.S. National Institutes of Health's Guide for the Care and Use of Laboratory Animals. 6 to 8-week-old male or female NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG) mice were used (Jackson laboratory, USA). All mice were housed in a pathogen-free animal facility in microisolator cages and the experimental protocol was approved by Stanford University's Administrative Panel on Lab Animal Care (APLAC #22264).

Southern blot genotyping

[0148] Southern blot analysis was performed to genotype the *CD14* locus after knocking in CRE in PT-iPSCs. Standard methods for Southern blot with genomic DNA were applied. Briefly,

10 ug of each DNA sample were subjected to restriction digest with 200 units NdeI (New England Biolabs, USA) and 200 units BlpI (New England Biolabs, USA). Swiss-Webster albino mouse genomic DNA (Promega) was used as a negative control. For preparation of size standards, $6.6x10^6$ copies of a 22 kb plasmid that contained the right homology arm CD14 sequence was spiked into 1.5 ug mouse gDNA and digested in a volume of 100 ul with various combinations of enzymes to yield the desired fragment size. A total of 7 different size standard digests were performed separately and pooled only after addition of phenol-chloroform-isoamyl alcohol (Invitrogen). Fragments were phenol-chloroform purified following the overnight restriction digest and precipitated using one volume 2-propanol after the addition of 1/10 volume of 3 M sodium acetate and 5 ul GlycoBlue (15mg/ml, Invitrogen). Following an overnight incubation at -20 °C, fragments were pelleted by centrifugation at 4 °C, washed with 70% ethanol, air dried, and resuspended in 40 ul tris-EDTA buffer with gel loading buffer (Blue Juice, Invitrogen). Samples were loaded onto a 10 x 6 inch 0.8% agarose gel using TAE as running buffer. The gel was run at 28 V for 24-hrs, incubated with denaturating buffer (3 M NaCl, 0.4 M NaOH) twice for 30 min with agitation, and incubated in transfer buffer (3 M NaCl, 8 mM EDTA) for 15 min with agitation. The gel was then blotted overnight onto a positively charged nylon membrane (Roche, Germany) using a piece of Whatman paper serving as a wick to transfer the buffer and thus the DNA from the gel onto the membrane by the means of capillary action. After crosslinking (GS GeneLinker, BioRad, USA), the membrane was pre-hybridized with 10 ug/ml salmon sperm DNA in PerfectHyb Plus Hybridization buffer (Sigma, USA) for a minimum of 2 hrs at 60 °C with rotation. The probe was generated by amplification of the 400 bp long right homology arm sequence from the rAAV vector using primers RHA-F (5'-GCGTGGTCCCAGCCTGTGC-3') and RHA-R (5'-GCAGCCCTAGCCAGGAGTC-3'). This fragment encompasses part of the CD14 coding sequence, the 3' UTR as well as some intergenic sequence downstream of the CD14 gene. The amplicon was gel-purified and 10 ng were labeled with [a-^{32}P] dCTP using the *Bca*BEST Labeling Kit (Takara Bio, Japan) according to the manufacturer's instructions. Unincorporated nucleotides were removed with an Illustra Microspin G-25 column and the probe was added to the pre-hybridized membrane. Hybridization was allowed to occur for 2-3 days at 60 °C with rotation. The membrane was washed twice under low-stringent condition (2× SSC, 20 min at room temperature), followed by one wash under high-stringent conditions (2× SSC with 0.1% SDS, 30 min at 60 °C). The membrane was exposed onto a phosphoimager screen and visualized using the Personal Molecular Imager (BioRad, USA).

Image analysis was performed using QuantityOne software. Band sizes were determined by interpolation between the log of adjacent size standards.

PCR genotyping

[0149] Two types of PCR genotyping strategies were used: 3-primer in-out PCR (see Bak *et al.*, *CRISPR/Cas9 genome editing in human hematopoietic stem cells. Nature protocols*, 2018. **13**(2):358-376) and 2-primer PCR. Approximate primer locations are shown in FIG. 11A. 3-primer in-out PCR can be used to analyze either the left side (PCR-L) or the right side (PCR-R) of the knockin site. 2-primer PCR spans the full knockin site (PCR-Full). Both methods result in an expected banding pattern shown in FIG. 11B. SeqAmp DNA polymerase (Takara Bio, Japan) was used for PCR amplification. All PCR reactions were run following manufacturers recommendations at a final volume of 20 ul with 0.5-1 ul of extracted DNA and 0.25 uM of each primer. All reactions were loaded on a 1% agarose gel in TAE mixed with 1:10,000 SYBR Safe (APExBio). 2 ul of Trackit 1 kb Plus (Invitrogen) or GeneRuler 1 kb (Thermo Scientific) were used for ladders. PCR product was diluted 1:6 with loading buffer, and 5-10 ul were loaded in each lane. In-out PCR was also used to analyze GFP and mCherry knockins at the TET2 locus. In this case, only two primers were used for PCR-L and PCR-R because all samples had a biallelic concatemer-free genotype.

Junction-spanning PCR and sequencing

[0150] PCR was used to amplify a product spanning the concatemer junction using similar PCR conditions as described above. Samples were the same as those analyzed by Southern blot. 10 μl of PCR product was analyzed by gel electrophoresis. For sequencing, upper and lower bands were cut from the gel and purified with the NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel, Germany). The bands were submitted for Sanger sequencing (Elim Biopharm, USA) and Primordium Labs' Nanopore-based sequencing (Primordium, USA).

qPCR

[0151] ITR qPCR was performed on the QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems, USA) using TaqMan™ Fast Advanced Master Mix (Applied Biosystems, USA); final reaction volume was 10 ul in a 384-well plate. All samples were analyzed with ITR and CD14-LHA primer/probe sets at a final concentration of 0.45 and 0.125 uM for primers and probes, respectively. The analyzed human genomic DNA was extracted from PT-iPSCs with CRE

knocked into *CD14* (sample #10 and #14 in FIG. 3B). Samples were diluted to similar molar concentrations (final concentration of ~300 copies of CD14-LHA per µl as determined by ddPCR without restriction enzymes). Both samples were analyzed +/- 2.5 units of the restriction enzyme AhdI (New England Biolabs, USA) added directly to the qPCR mixture. Thermocycler conditions were run in accordance with manufacturer recommendations with the addition of an initial 37 °C x 5-minute incubation.

Microscopy

[0152] Cells were imaged using the Operetta High Content Imaging System (Perkin Elmer, USA) with GFP and mCherry fluorescent filter sets. ImageJ 1.52p (NIH, USA) was used to subtract background and measure the average GFP and mCherry fluorescent intensity within the cell colonies. Images are displayed using individualized or uniform imaging processing. For individualized image processing, auto-brightness-&-contrast was used to set display thresholds prior to converting and combining images. For uniform processing, display thresholds were set to the same values for all images.

Polydispersity in flow cytometry quadrants

[0153] Polydispersity in flow cytometry plots was measured as relative change in average absolute deviations. In brief, the four centroids of the cell cluster in each quadrant were calculated separately, using GFP as the x-axis and mCherry as the y-axis. The average distance of each cell to the respective centroid was then measured, and IC and DC measurements were normalized to the control (NC). All measurements were performed with linear (not log-transformed) data points.

Statistics and graphs

[0154] ddPCR 1D and 2D plots were generated with QuantaSoft Analysis Pro version 1.0.596 (BioRad, USA). All other graphs were generated in Microsoft Excel and PowerPoint. Linear regressions were measured in Microsoft Excel. Significance was calculated with two-sided t-tests using Microsoft Excel. For paired t-tests, the following function was used:

p-value =T.TEST(*array1*,*array2*,2,1)

[0155] For unpaired t-tests, the following function was used:

p-value =T.TEST(*array1*,*array2*,2,2)

[0156] While the invention herein disclosed has been described by means of specific embodiments and applications thereof, numerous modifications and variations could be made thereto by those skilled in the art without departing from the scope of the invention set forth in the claims.

[0157] It will be appreciated that the present invention is set forth in various levels of detail in this application. In certain instances, details that are not necessary for one of ordinary skill in the art to understand the invention, or that render other details difficult to perceive may have been omitted. The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting beyond the scope of the appended claims. Unless defined otherwise, technical terms used herein are to be understood as commonly understood by one of ordinary skill in the art to which the disclosure belongs.

[0158] The abbreviations used herein have their conventional meaning within the chemical and biological arts.

[0159] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by a person of ordinary skill in the art. *See, e.g.*, Singleton *et al., Dictionary of Microbiology and Molecular Biology*, 2nd ed., J. Wiley & Sons (New York, NY 1994); Green and Sambrook, *Molecular Cloning, A Laboratory Manual*, 4th ed., Cold Springs Harbor Press (Cold Springs Harbor, NY 2012). Any methods, devices and materials similar or equivalent to those described herein can be used in the practice of this disclosure. The following definitions are provided to facilitate understanding of certain terms used frequently herein and are not meant to limit the scope of the present disclosure.

[0160] The terms "nucleic acid", "polynucleotide", and "oligonucleotide" are used interchangeably herein to refer to polymers of deoxyribonucleotides or ribonucleotides in either single-, double- or multiple-stranded form, or complements thereof. The term "nucleotide" refers, in the usual and customary sense, to a single unit of a polynucleotide, *i.e.*, a monomer. Nucleotides can be ribonucleotides, deoxyribonucleotides, or modified versions thereof. Examples of polynucleotides include single and double stranded DNA, single and double stranded RNA, and hybrid molecules having mixtures of single and double stranded DNA and RNA. RNA may include messenger RNA (mRNA), small interference RNA (siRNA), short hairpin RNA (shRNA), micro RNA (miRNA), guide RNA (gRNA), CRISPR RNA (crRNA), and transactivating RNA (tracrRNA). DNA may include plasmid DNA (pDNA), minicircle

DNA, genomic DNA (gNDA), and fragments thereof. The term "duplex" in the context of polynucleotides refers, in the usual and customary sense, to double strandedness.

[0161] "Complementarity" refers to the ability of a nucleotide sequence in one strand of nucleic acid to hydrogen bond with another sequence of an opposing nucleic acid strand via Watson-Crick pairing of the bases in each strand of the resulting duplex. The complementary bases in DNA are typically A / T and C / G; and in RNA are A / U and C / G. Where two nucleic acids share perfect complementarity every base in the duplex is bonded to its complementary base. However, perfect complementarity is not required for formation of a stable duplex. In this context, the complementary is nevertheless sufficient to form a duplex under standard conditions or under a desired set of specific conditions, *e.g.*, of temperature and salt concentration, which may also be referred to as "hybridization" conditions in the context of *e.g.*, a nucleic acid probe hybridizing to a target sequence. Conditions appropriate for hybridization between two nucleic acids depend on the length of the nucleic acids and the degree of complementation.

[0162] Percent complementarity between two nucleic acid sequences or between similar regions within nucleic acids can be determined using methods widely known in the art, including the basic local alignment search tools ("BLAST") available from the US National Library of Medicine. BLAST and similar algorithms may be used to determine a percentage of sequence identity between two nucleic acid sequences. The percent identity calculation is determined by comparing two optimally aligned sequences, meaning the alignment that gives the greatest number of perfectly matched residues over a comparison window. The portion of the sequence within the comparison window may comprise additions or deletions, referred to as "gaps" relative to the reference sequence in order to produce the optimal alignment. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity. Unless otherwise specified, the comparison window is the full length of the shorter of the two sequences being compared.

[0163] "Contacting" is used in accordance with its plain ordinary meaning and refers to the process of allowing at least two distinct species (e.g. chemical compounds including biomolecules or cells) to become sufficiently proximal to react, interact, associate, or

physically touch. It should be appreciated, however, that the resulting reaction product can be produced directly from a reaction between the added reagents or from an intermediate from one or more of the added reagents which can be produced in the reaction mixture. In aspects, contacting includes, for example, allowing a nucleic acid to interact with an endonuclease.

[0164] An "insert" refers to a DNA sequence corresponding to a part of the AAV viral genome that is inserted or "knocked in" to a target site in the cell's genome. In the illustrative embodiments and examples described herein, the inserts are CRE, GFP, mCherry, etc. However, in the context of genetic engineering the insert will typically be a heterologous gene, for example a gene encoding a protein of interest to be inserted into a target site in a cellular genomic DNA. The number of measured insertions may change if analyzed with and without restriction enzymes that separate concatemers, as discussed herein.

[0165] The term "about", unless otherwise specified herein, means a range of values including the specified value, which a person of ordinary skill in the art would consider reasonably similar to the specified value. In aspects, about means within a standard deviation using measurements generally acceptable in the art. In aspects, about means a range extending to +/- 10% of the specified value, and including the specified value.

[0166] The phrases "at least one", "one or more", and "and/or", as used herein, are open-ended expressions that are both conjunctive and disjunctive in operation. The terms "a", "an", "the", "first", "second", etc., do not preclude a plurality. For example, the term "a" or "an" entity, as used herein, refers to one or more of that entity. As such, the terms "a" (or "an"), "one or more" and "at least one" can be used interchangeably herein.

[0167] Also as used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative ("or").

[0168] As used herein, the term "comprises/comprising" does not exclude the presence of other elements, components, features, regions, integers, steps, operations, etc. Additionally, although individual features may be included in different claims, these may possibly advantageously be combined, and the inclusion in different claims does not imply that a combination of features is not feasible and/or advantageous. The transitional phrase "consisting essentially of" (and grammatical variants) is to be interpreted as encompassing the recited materials or steps "and those that do not materially affect the basic and novel characteristic(s)"

of the recited embodiment. Thus, the term "consisting essentially of" as used herein should not be interpreted as equivalent to "comprising."

CLAIMS

What is claimed is:

1. A method for preventing unintended concatemeric insertions of a viral genome at an insert site in a genomic DNA of a cell, the method comprising

introducing into a population of cells a first and second ribonucleoprotein ("RNP") and a recombinant adeno-associated virus (AAV) comprising a repair template; and

incubating the cells under conditions sufficient for cleavage of a target sequence of cellular genomic DNA by the first RNP and cleavage of a viral inverted terminal repeat (ITR) by the second RNP.

2. The method of claim 1, wherein each RNP comprises a complex of a guide RNA (gRNA) and a clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) endonuclease.

3. The method of claim 2, wherein the gRNA of the first RNP comprises a guide sequence complementary to the target sequence of cellular genomic DNA and the gRNA of the second RNP comprises a guide sequence complementary to a sequence of the viral ITR.

4. A method for preventing unintended concatemeric insertions of a viral genome at an insert site in a genomic DNA of a cell, the method comprising

introducing into a population of cells a ribonucleoprotein ("RNP") and a recombinant adeno-associated virus (AAV) comprising a repair template; and

incubating the cells under conditions sufficient for cleavage of a target sequence of cellular genomic DNA by the RNP,

wherein the recombinant AAV comprises two polynucleotide insertions flanking its homology arms, each insertion comprising the target sequence.

5. A method for preventing unintended concatemeric insertions of a viral genome at an insert site in a genomic DNA of a cell, the method comprising

introducing into a population of cells a ribonucleoprotein ("RNP"), a modified adeno-associated virus (AAV) comprising a repair template, and a PI3K/mTOR kinase inhibitor; and

incubating the cells under conditions sufficient for cleavage of a target sequence of cellular genomic DNA by the RNP,

wherein the PI3K/mTOR kinase inhibitor is introduced before, simultaneously with, or after the RNP.

6. The method of claim 4 or 5, wherein the RNP comprises a complex of a guide RNA (gRNA) and a clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) endonuclease.

7. The method of claim 5, wherein the PI3K/mTOR kinase inhibitor is selected from apitolisib (GDC-0980), bimiralisib (PQR309), dactolisib (NVP-BEZ235), gedatolisib (PKI-587; PF-05212384), paxalisib (GDC-0084), or samotolisib (LY3023414). In aspects, the PI3K/mTOR kinase inhibitor is selected from apitolisib, bimiralisib, dactolisib, gedatolisib, omipalisib, paxalisib, samotolisib, voxtalisib, CMG 002, GNE-477, GSK1059615, MCX-83, NSC765844, PF-04691502, PF-04979064, PI-108, PI-103BE, PKI-402, SN202, or VS-5584.

8. The method of claim 7, wherein the PI3K/mTOR kinase inhibitor is dactolisib (NVP-BEZ235).

9. The method of any one of claims 1 to 8, wherein the RNP is introduced to the population of cells by electroporation.

10. The method of any one of claims 1 to 8, wherein the RNP is introduced to the population of cells by injection or by intravenous infusion.

11. The method of any one of claims 1 to 10, wherein the AAV is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, or AAV12.

12. The method of claim 11, wherein the AAV is AVV2 or AAV6.

13. The method of claim 11 or 12, wherein the Cas endonuclease is a Cas9 or a Cas12a endonuclease.

14. The method of any one of claims 11 to 13, wherein the population of cells comprises any vertebrate or mammalian cell type.

15. The method of any one of claims 1 to 13, wherein the population of cells comprises pluripotent stem cells, induced pluripotent stem cells, embryonic stem cells, or hematopoietic stem and progenitor cells.

16. A method for detecting concatemeric insertions of a viral genome at an insert site in a genomic DNA of a cell, the method comprising digital droplet PCR genotyping of the cellular DNA in combination with restriction endonuclease analysis.

17. The method of claim 16, wherein the method comprises a copy number variation analysis, the method comprising

performing a first ddPCR reaction to produce a first set of amplicons, the reaction comprising a target cellular genomic DNA, a first primer/probe set comprising a first reference primer directed to a first genomic reference site in *cis* or *trans* with the insert site in the cellular genomic DNA; an optional second primer/probe set comprising a second reference primer directed to a second genomic reference site in *cis* or *trans* with the insert site; a third primer/probe set comprising a primer directed to a region of the viral genome; and a fourth primer/probe set comprising a primer directed to a cellular genomic region spanning the insert site such that the primers and probe are generally at least 25 bp from the insert site;

performing a second ddPCR reaction to produce a second set of amplicons, the reaction comprising the target cellular genomic DNA, the first primer/probe set, optionally the second primer/probe set, the third primer/probe set, the fourth primer/probe set, and at least one restriction endonuclease which cleaves within the insert; and

comparing the first and second sets of amplicons to determine whether concatemeric insertions of the viral genome are present, where an increase in detected inserts in the second ddPCR reaction indicates the presence of concatemeric insertions.

18. The method of claim 16, wherein the method comprises a linkage analysis, the method comprising performing a ddPCR reaction to produce a set of amplicons, the reaction comprising a target cellular genomic DNA, a first primer/probe set comprising a first reference primer directed to a first genomic reference site in *cis* with an insert site in the cellular genomic DNA; an optional second primer/probe set comprising a second reference primer directed to a second genomic reference site in *cis* with the insert site; a third primer/probe set comprising a primer directed to a region of the viral genome; and a fourth primer/probe set comprising a primer directed to a region of the viral genome not intended to be inserted; and

comparing the linkage of amplicons of the primer/probe sets, where detection of an amplicon produced from the fourth primer/probe set linked to a reference amplicon indicates concatemeric or non-HDR mediated insertions have occurred.

19. The method of claim 18, wherein the fourth primer/probe set comprises a primer directed to a viral ITR region or a region of the viral genome outside of the viral homology arms.

20. The method of claim 18 or 19, wherein the method comprises performing a second ddPCR reaction comprising the target cellular genomic DNA, the first primer/probe set, optionally the second primer/probe set, the third primer/probe set, the fourth primer/probe set, and at least one restriction endonuclease which cleaves at a target site within the viral genome; and

comparing the linkage of amplicons of each of the primer/probe sets in the ddPCR reactions performed with and without the restriction endonuclease in order to determine whether the insertions are concatemeric or non-HDR mediated insertions.

21. The method of claim 16, wherein the method comprises a nuclease mediated loss of linkage analysis, the method comprising performing a first ddPCR reaction to produce a first set of amplicons, the reaction comprising a target cellular genomic DNA, a first primer/probe set comprising a first reference primer directed to a first genomic reference site in *cis* with an insert site in the cellular genomic DNA, a second primer/probe set comprising a second reference primer directed to a second genomic reference site in *cis* with the insert site;

performing a second ddPCR reaction to produce a second set of amplicons, the reaction comprising the target cellular genomic DNA, the first primer/probe set, the second primer/probe set, and at least one restriction endonuclease which cleaves a site between the first and second cellular genomic reference sites; and

comparing the first and second sets of amplicons to detect loss of linkage between the first and second cellular genomic reference sites, wherein loss of linkage indicates a concatemeric insertions of a viral genome.

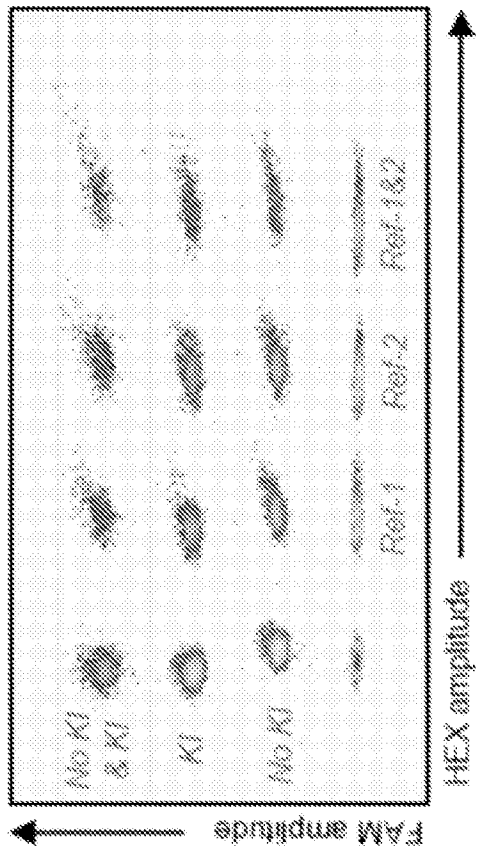

FIG. 1B

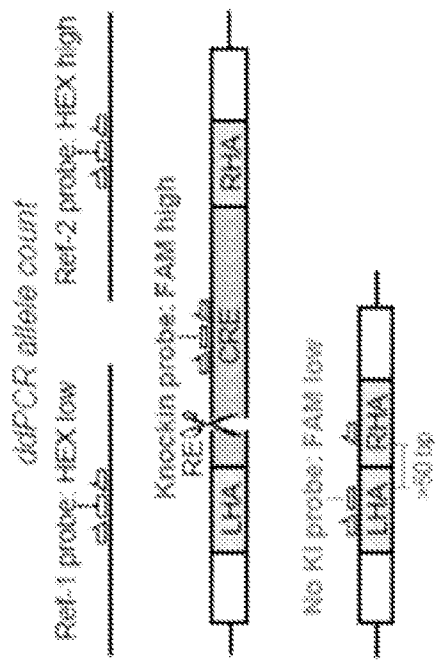

FIG. 1A

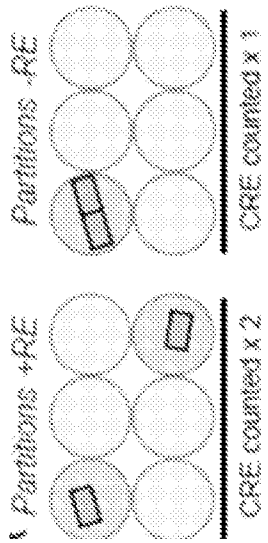

FIG. 1D

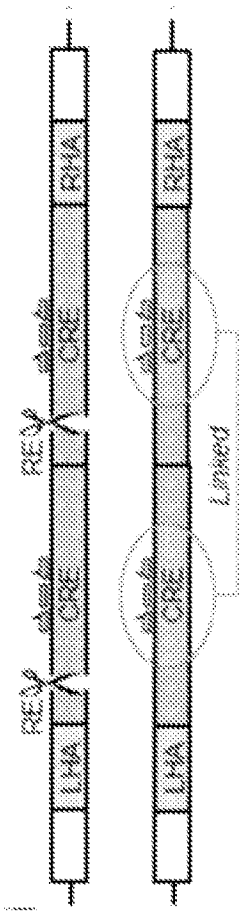

FIG. 1C

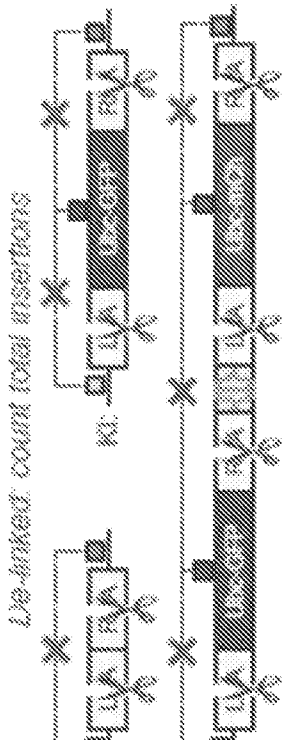

FIG. 2A

FIG. 2B

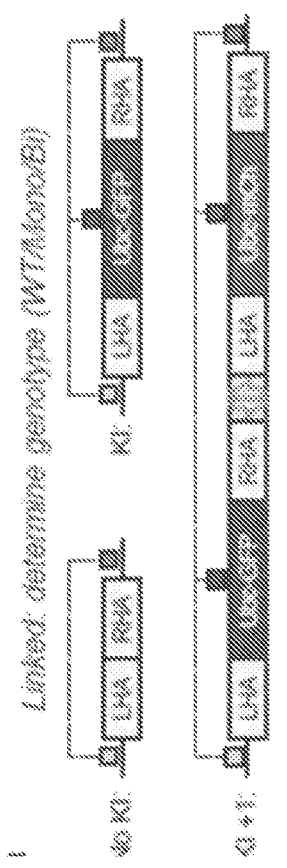

FIG. 2C

FIG. 2D

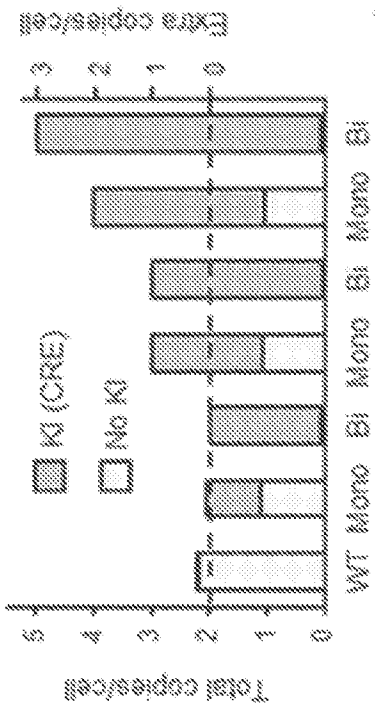
Extra copies/cell
Total copies/cell
KI (CRE)
No KI
WT
Mono
Bi
Mono
Bi
Mono
Bi

FIG. 3A

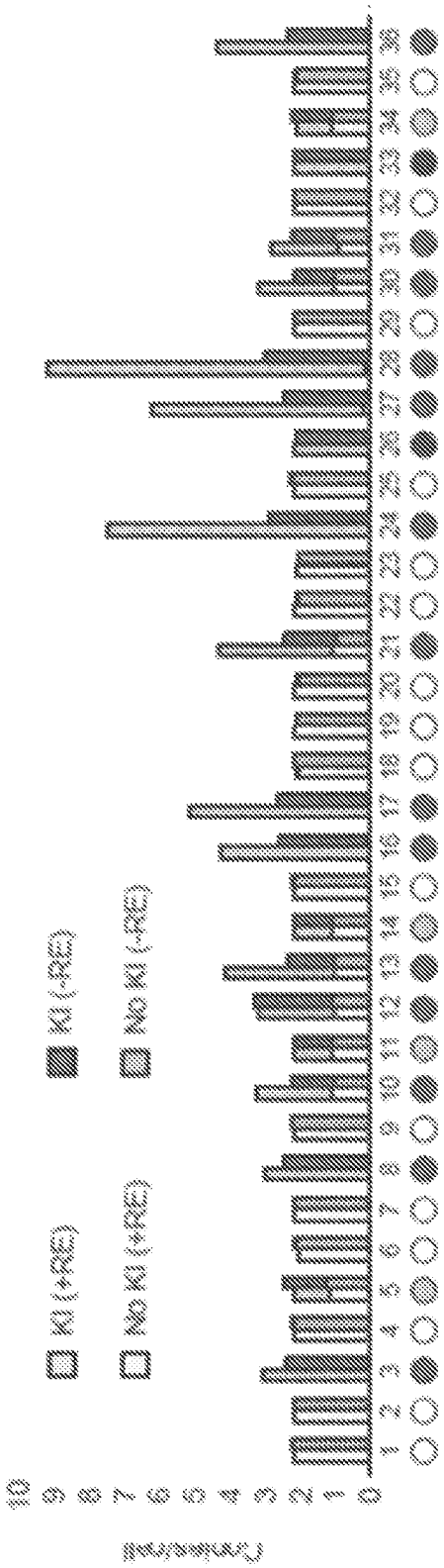
Cells/well
KI (+RE)
No KI (+RE)
KI (-RE)
No KI (-RE)

FIG. 3B

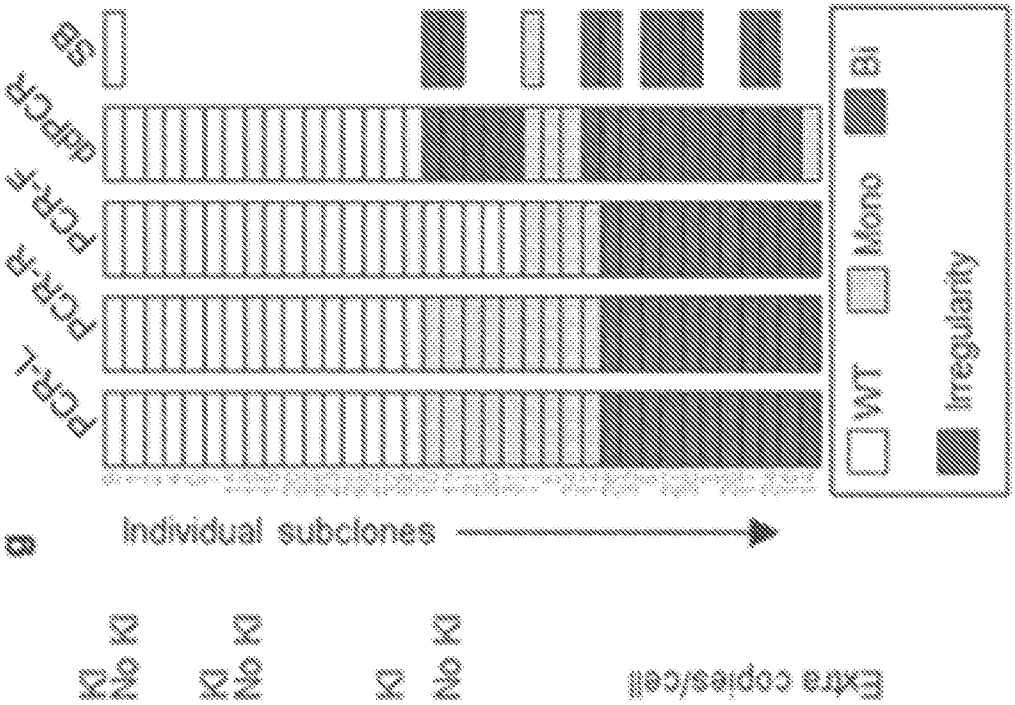
g
Individual subclones
PCR-L
PCR-R
PCR-F
dsPCR
SB
WT
Mono
Bi
Irregularity
Extra copies/cell
KI
No KI

FIG. 3D

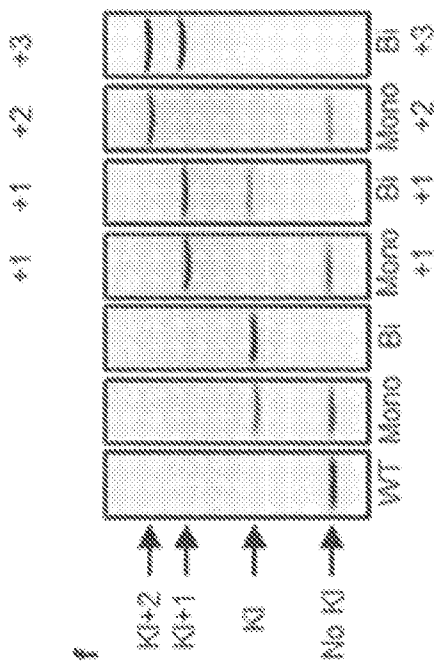
f
KI+2
KI+1
KI
No KI
WT
Mono
Bi
+1
+2
+3

FIG. 3C

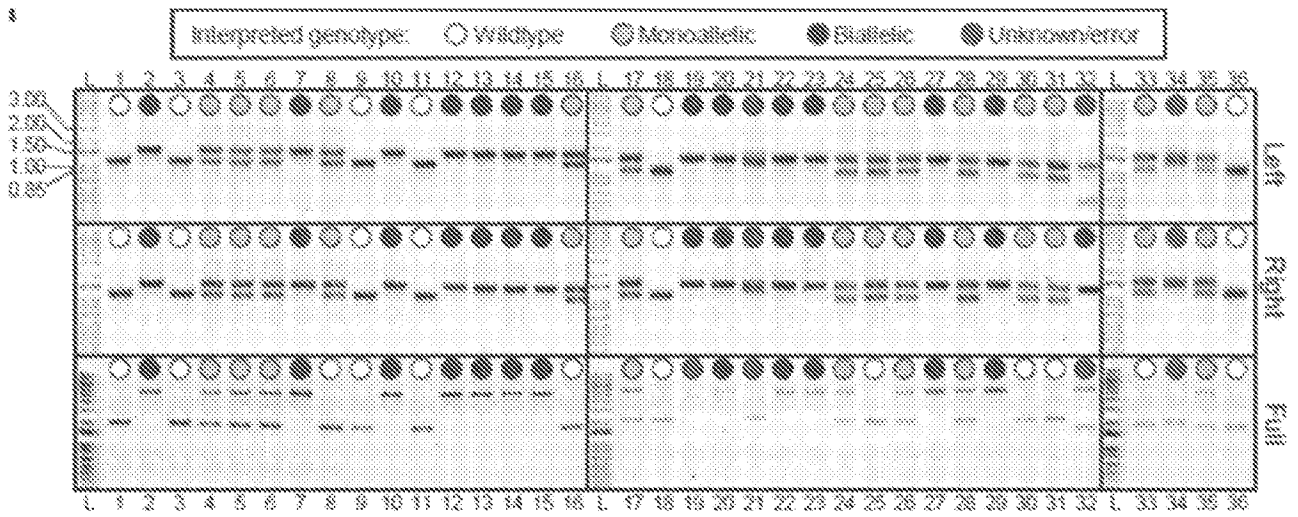

FIG. 4A

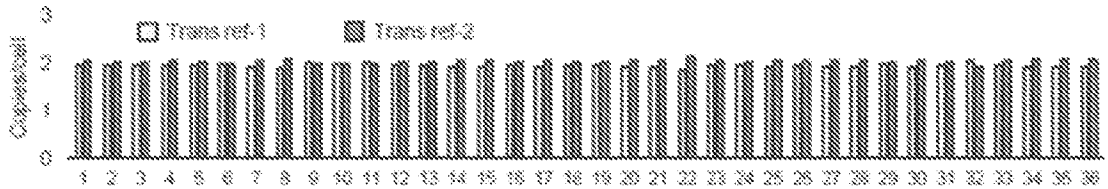

FIG. 4B

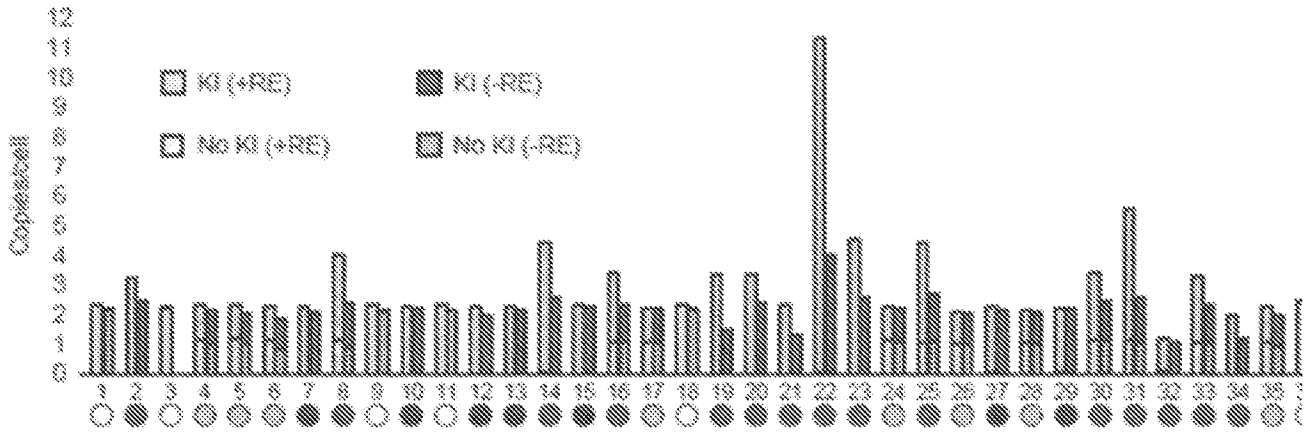

FIG. 4C

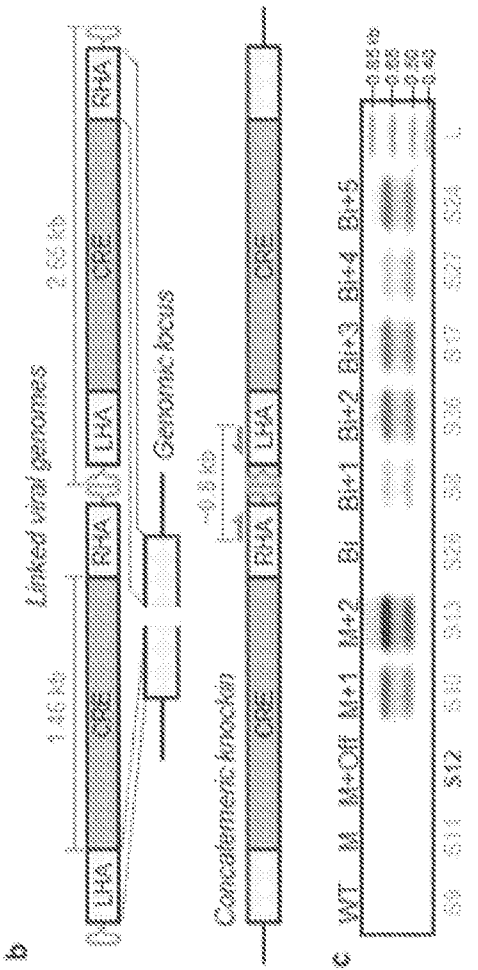

FIG. 5B

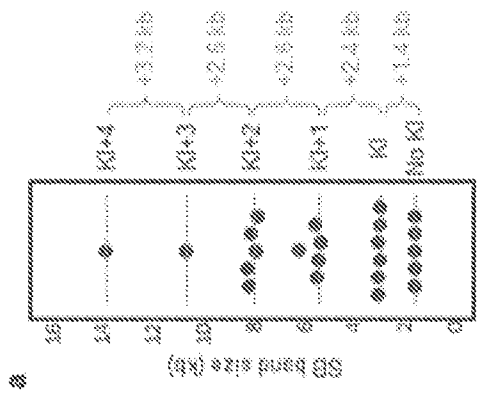

FIG. 5A

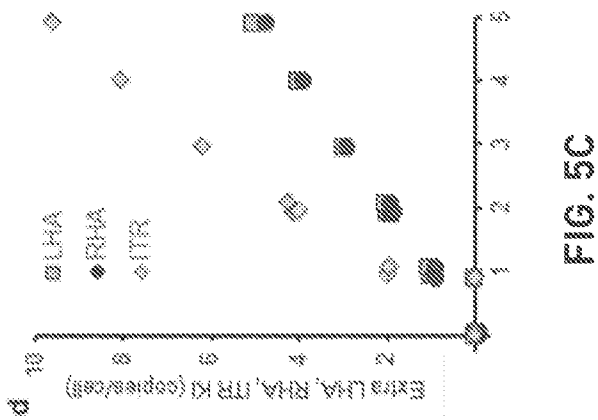

FIG. 5C

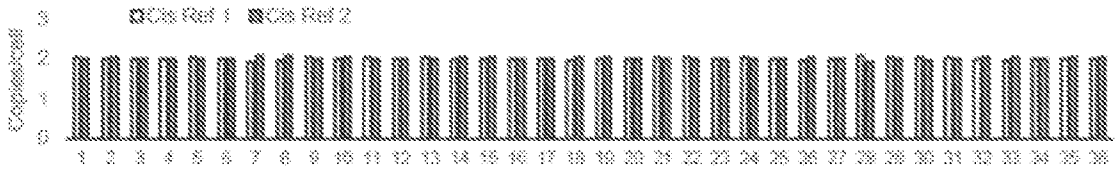

FIG. 6A

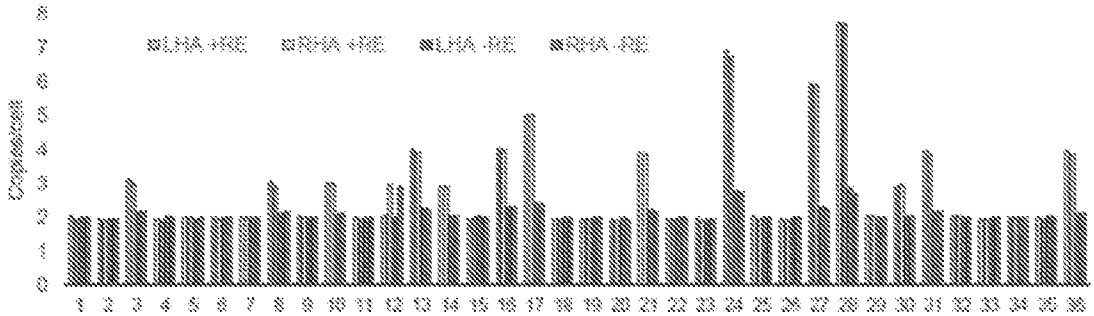

FIG. 6B

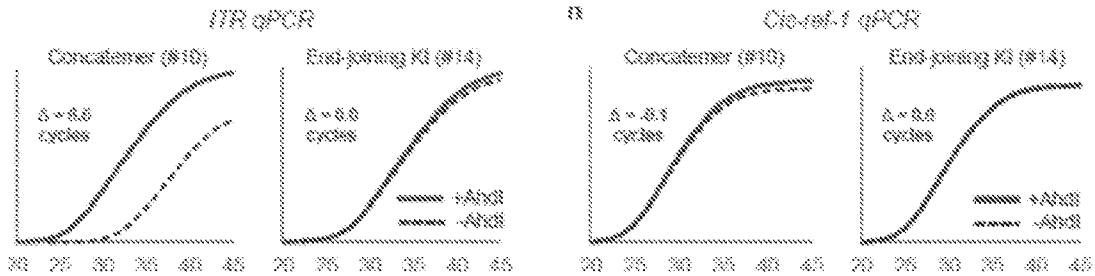

FIG. 6C

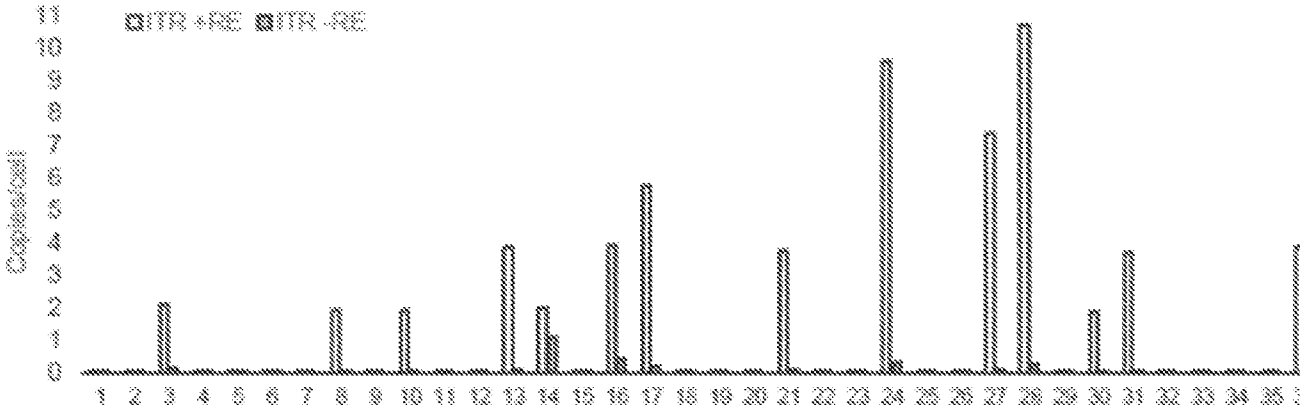

FIG. 6D

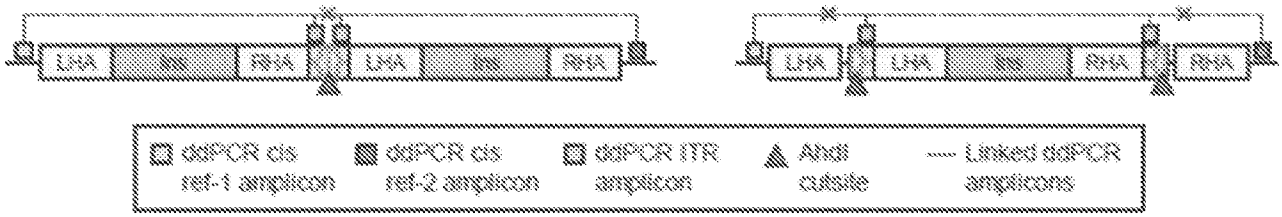

FIG. 6E

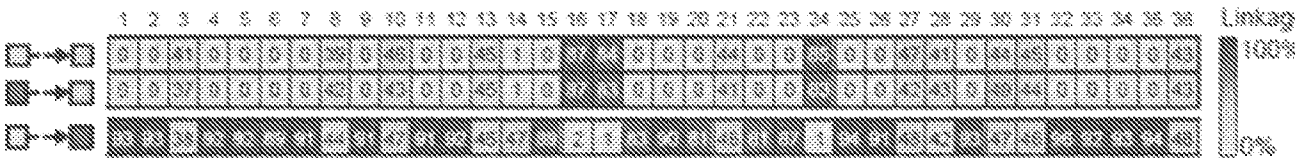

FIG. 6F

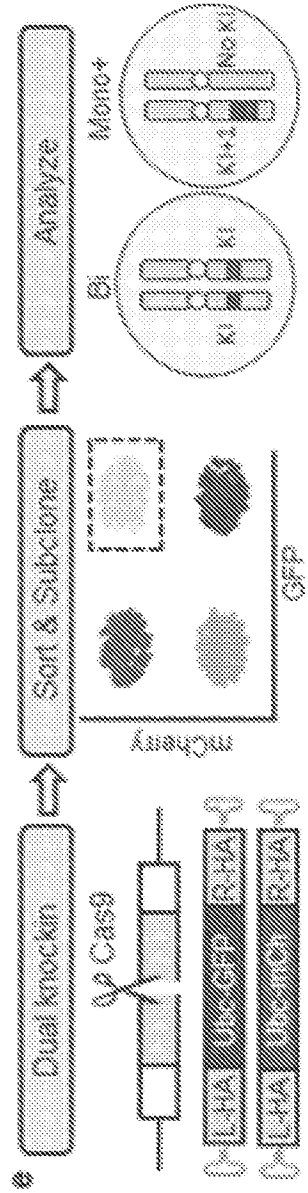

FIG. 7A

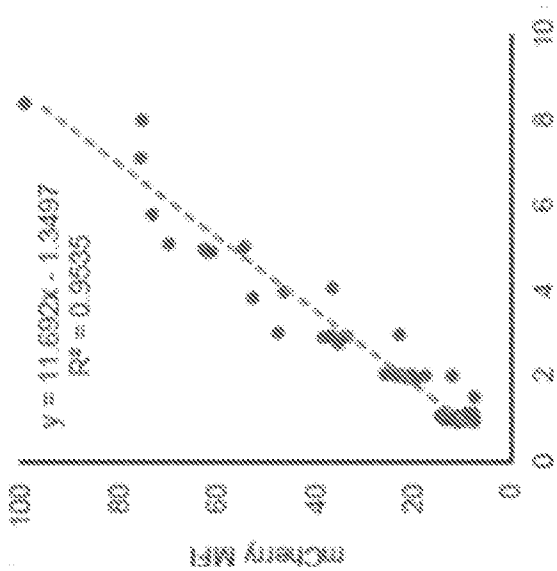

FIG. 7C

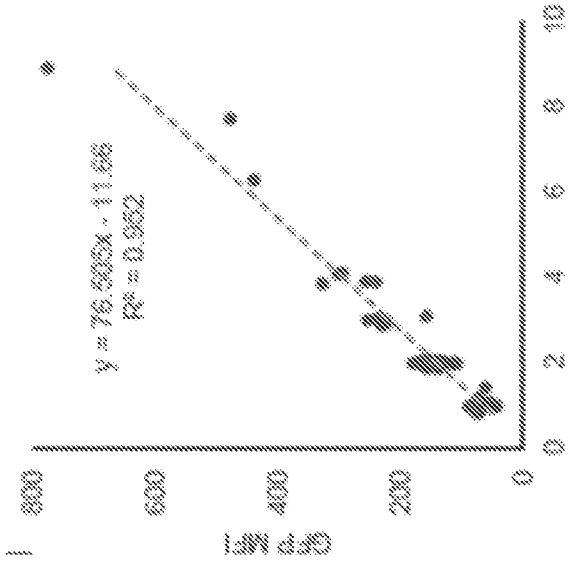

FIG. 7B

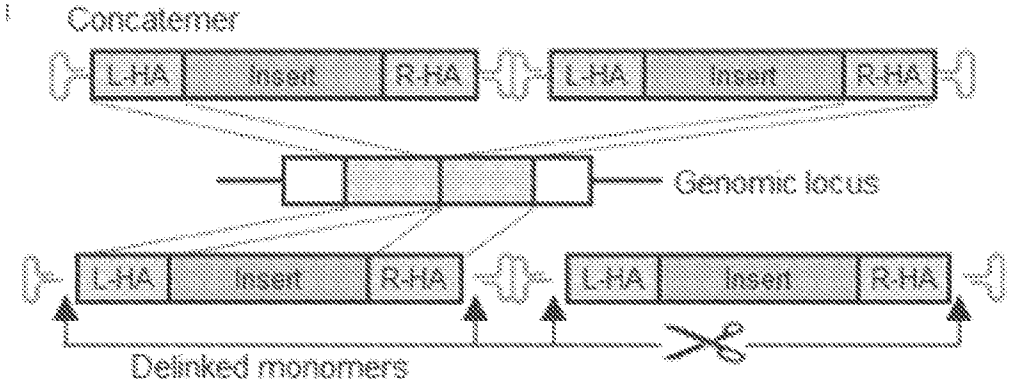
Concatemer
L-HA
Insert
R-HA
L-HA
Insert
R-HA
Genomic locus
L-HA
Insert
R-HA
L-HA
Insert
R-HA
Delinked monomers

FIG. 8A

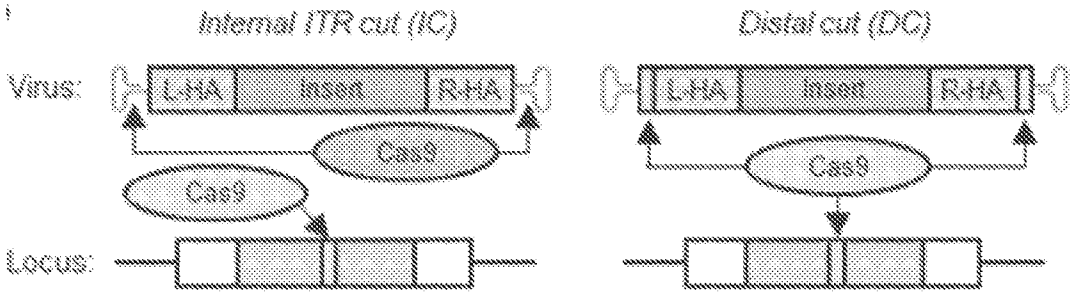
Internal ITR cut (IC)
Distal cut (DC)
Virus:
L-HA
Insert
R-HA
Cas9
Cas9
L-HA
Insert
R-HA
Cas9
Locus:

FIG. 8B

c

Locus	Cell line	n
TET2	PT-iPSC	64
AAVS1	PT-iPSC	79
AAVS1	H9	57
AAVS1	WT-iPSC	58
HBB	PT-iPSC	65
HBB	H9	67
RUNX1	H9	68

FIG. 8C

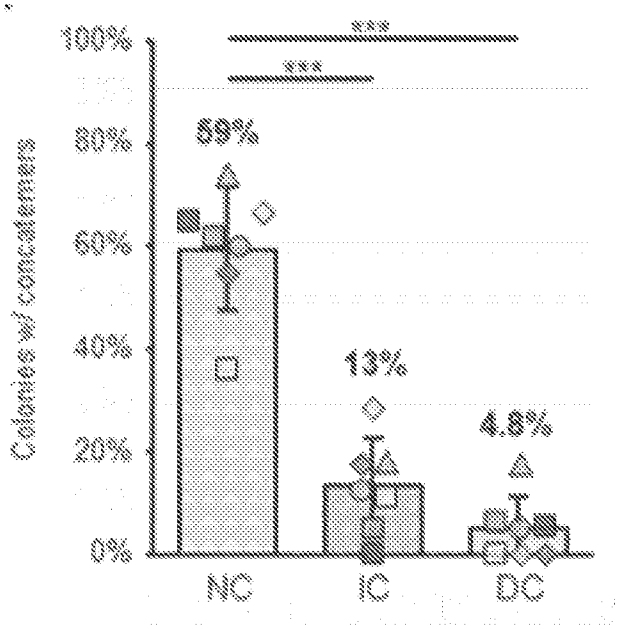

FIG. 8D

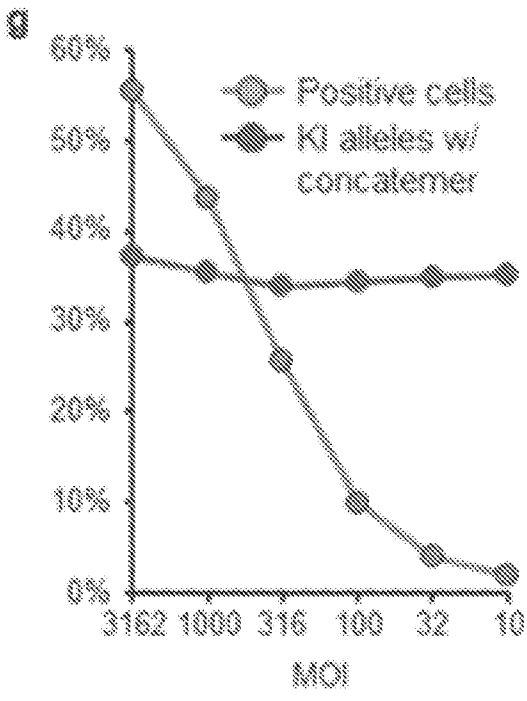

FIG. 8E

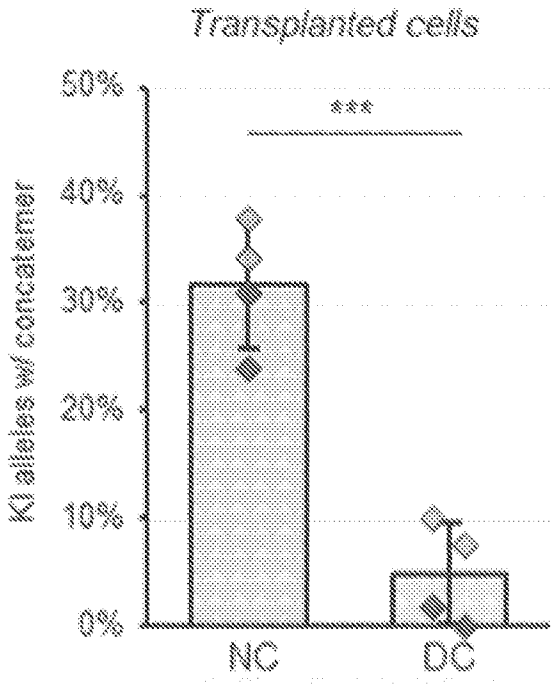
Transplanted cells
50%
40%
30%
20%
10%
0%
***
KI alleles w/ concatamer
NC
DC

FIG. 9

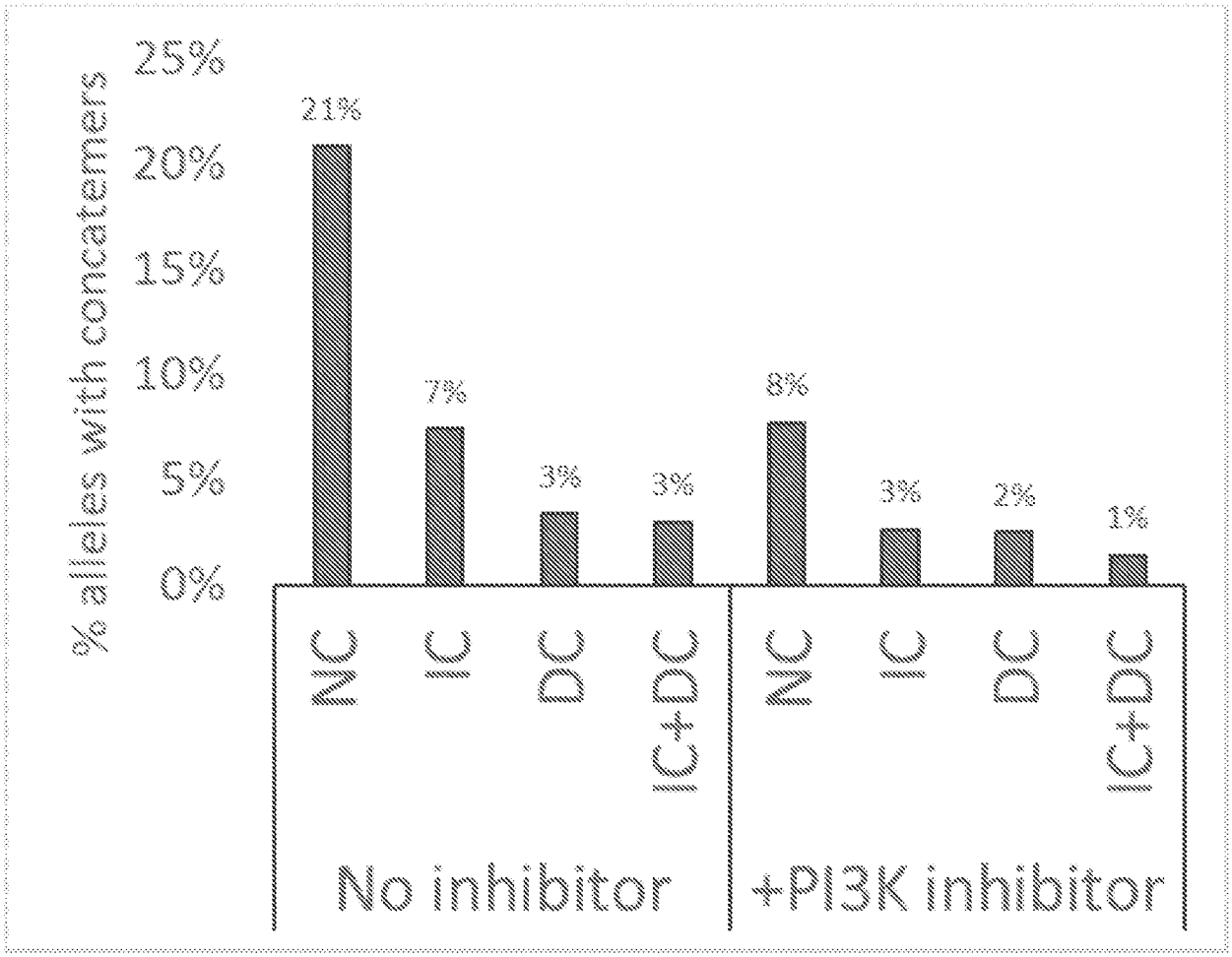
% alleles with concatemers
25%
20%
15%
10%
5%
0%
21%
7%
3%
3%
8%
3%
2%
1%
NC
IC
DC
IC+DC
NC
IC
DC
IC+DC
No inhibitor
+PI3K inhibitor

FIG. 10

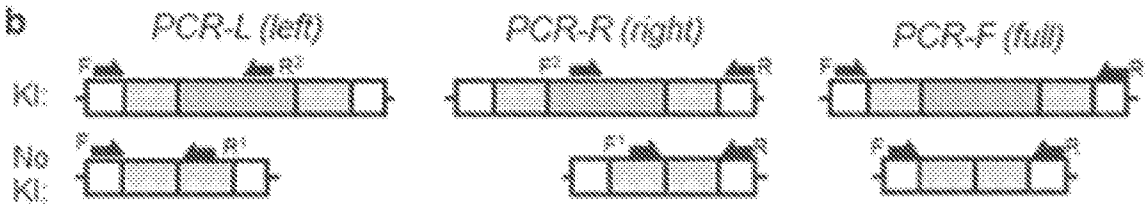

FIG. 11A

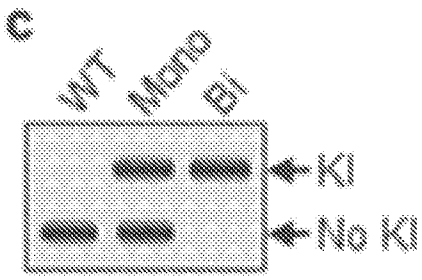

FIG. 11B

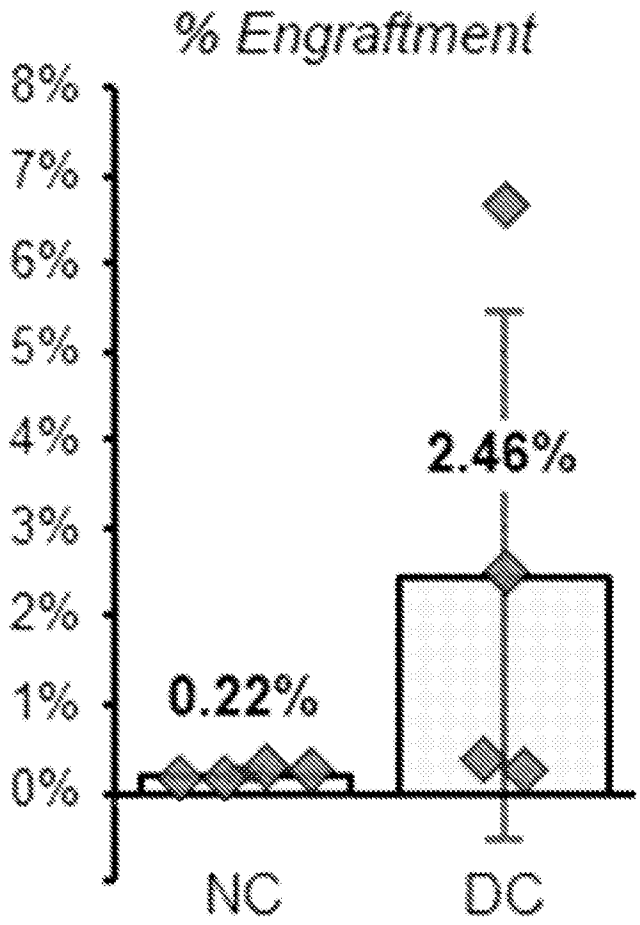

FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/061007

A. CLASSIFICATION OF SUBJECT MATTER

IPC: ***C12N 15/11*** (2025.01); ***C12N 15/09*** (2025.01); ***C12N 15/864*** (2025.01); *C12N 5/10* (2025.01); *A61K 35/76* (2025.01)

CPC: ***C12N 15/11; C12N 15/09; C12N 15/86;*** *C12N 2310/20; C12N 2750/14143*

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)

See Search History Document

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

See Search History Document

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

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2023/0193318 A1 (PRECISION BIOSCIENCENS, INC.) 22 June 2023 (22.06.2023) para [0009]; [0044]; [0048]; [0170]; [0120]	1-8
Y	US 2016/0355796 A1 (THE BROAD INSTITUTE INC.) 08 December 2016 (08.12.2016) para [0055]; [0126]; [0235]	1-4, 6/4
Y	WO 2023/220641 A2 (JUNO THERAPEUTICS, INC.) 16 November 2023 (16.11.2023) para [0601]	7-8
Y	MADDALENA et al. High-Throughput Screening Identifies Kinase Inhibitors That Increase Dual Adeno-Associated Viral Vector Transduction In Vitro and in Mouse Retina. Human Gene Therapy, 01 August 2018, Vol. 29, No. 8, Pgs. 886-901 pg 892, col 1, para 2; pg 892, col 1, para 3; pg 892, col 2, para 1; Table 1	5, 6/5, 7-8

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

Date of the actual completion of the international search	Date of mailing of the international search report
23 January 2025 (23.01.2025)	**01 April 2025 (01.04.2025)**

Name and mailing address of the ISA/US	Authorized officer
COMMISSIONER FOR PATENTS **MAIL STOP PCT, ATTN: ISA/US** **P.O. Box 1450** **Alexandria, VA 22313-1450** **UNITED STATES OF AMERICA**	**KARI RODRIQUEZ**
Facsimile No. **571-273-8300**	Telephone No. **PCT Help Desk: 571-272-4300**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/061007

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed.
   b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)),
      ☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/061007

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.: **9-15**
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:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1.

Group I, claims 1-8, directed to a method for preventing unintended concatemeric insertions of a viral genome at an insert site in a genomic DNA of a cell.

Group II, claims 16-21, directed to a method for detecting concatemeric insertions of a viral genome at an insert site in a genomic DNA of a cell.

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 additional fees, this Authority did not invite payment of additional fees.

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

4. ☑ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: **1-8**

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.