

METHODS FOR SELECTIVE AMPLIFICATION FOR EFFICIENT REARRANGEMENT DETECTION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 63/384,299, filed November 18, 2022, which is incorporated herein by reference in its entirety, including any drawings.

FIELD

[0002] The present disclosure relates to methods for quantitative detection of genomic rearrangements.

BACKGROUND

[0003] For gene editing therapy involving nucleases such as CRISPR/Cas9, cleavage of genomic DNA in live cells results in double strand breaks (DSB), which have the potential of recombining during DSB repair to generate genomic rearrangements. Typically, such rearrangements occur between the target site and frequently edited off-target sites, but they may also occur between two off-target sites, or between the target site or off-target site and a fragile site or any other genomic locus where chromosomal breakage or nicking may occur.

[0004] Several molecular methods, including AMP, HTGTS, UDiTaS, CAST-seq, and combinatorial multiplex PCR using rhAmpseq have been applied to enable detection of rearrangements in cells or mammalian tissues following gene editing. However, most of these methods use a low-efficiency DNA shearing and adaptor ligation approach and require significant amounts of genomic DNA as input. One of the methods, UDiTaS, uses a more efficient tagmentation approach, but this method has no enrichment procedures for rearrangements, and thus suffers from low sensitivity for rearrangement detection due to a high background of non-rearranged molecules.

[0005] The disclosure provided herein provides methods that employ a combination of tagmentation and negative selection to reduce DNA input requirements, eliminate time-consuming sample processing steps, reduce background resulting from spurious products, and improve sensitivity for quantitative rearrangement detection.

BRIEF SUMMARY

[0006] The present disclosure relates generally to the development of methods for quantitative detection of genomic rearrangements. In particular, as described in greater detail below, some embodiments of the disclosure provide methods which involve tagmentation in combination with amplification techniques to reduce sample processing times, reduce background, and increase sensitivity for quantitative rearrangement detection. The foregoing summary is illustrative only and is not intended to be in any way limiting. In addition to the illustrative embodiments and features described herein, further aspects, embodiments, objects and features of the disclosure will become fully apparent from the drawings and the detailed description and the claims.

[0007] Provided herein, among others, includes a method for detecting genome-wide re-arrangements in a nucleic acid genome. The method involves (a) contacting a genomic DNA sample obtained from a cell or tissue contacted with a site specific nuclease with a plurality of transposons comprising a first sequence tag at the 5' end of the transposon, under conditions whereby the plurality of transposons is inserted into the genomic DNA sample and the genomic DNA sample is tagmented into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment. In (b), a first amplification reaction is performed to amplify the re-arranged tagmented nucleic acid fragments using (i) a first target-specific primer comprising a nucleotide sequence complementary to a region adjacent to a target site, (ii) a second primer comprising a nucleotide sequence identical to a portion of the sequence tag located at the 5' ends of the fragment and (iii) a blocking oligonucleotide comprising a nucleotide sequence complementary to a region on the distal side of the target site relative to the first target-specific primer to produce primary amplification products comprising a nucleotide sequence comprising a re-arranged target sequence. In (c), the primary amplification products are then sequenced.

[0008] In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a unique molecular identifier (UMI), a transposase recognition site, an index sequence, or a combination thereof. In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a UMI, and a transposase recognition site. In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a UMI, a transposase recognition site and an index sequence.

[0009] In one embodiment, the blocking oligonucleotide comprises an absent or blocked 3' OH, spacers, inverted nucleotides, or other modifications to block extension of the 3' end. In one embodiment, the sequence tag comprises uracil.

[0010] In one embodiment, the blocking nucleotide comprises one or more phosphorothioate bonds, spacers, or other modifications at the 3' and 5' ends, to block exonuclease digestion at the 3' and 5' ends, LNA, BNA, PNA, RNA, DNA, modified nucleic acids, or a combination thereof.

[0011] In one embodiment, the first and second primers comprise second and third sequence tags.

[0012] In one embodiment, the method disclosed herein further includes prior to (c) performing a second amplification reaction using third and fourth nested primers comprising nucleotide sequences complementary to sequences proximal to the inside ends of the amplified products of (b) and fourth and fifth sequence tags at the 5' ends of the nested primers to produce secondary amplification products comprising a re-arranged target sequence and additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fifth and sixth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

[0013] In one embodiment, the third and/or the fourth primers comprise barcode sequences.

[0014] In one embodiment, the fifth and sixth primers comprise a sequencing tag and/or an index sequence.

[0015] In one embodiment, the method disclosed herein further includes prior to (c) performing a second, hemi-nested amplification reaction using the second primer and a third nested primer wherein the third primer comprises a nucleotide sequence complementary to a sequence proximal to target-specific primer end of the amplified products of (b) and optionally a fourth and fifth sequence tags at the 5' ends of the second and/or third primers to produce secondary amplification products comprising a re-arranged target sequence and one or two additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fourth and fifth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

- [0016] In one embodiment, the fourth and fifth primers comprise a sequencing tag and/or an index sequence.
- [0017] In one embodiment, (b) further includes performing a separate amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using a second target specific primer on the opposite side of the target site.
- [0018] In one embodiment, the method disclosed herein further includes prior to (b) contacting the plurality of tagmented nucleic acid fragments with ddNTP or other 3' modified dNTP and a DNA polymerase or terminal deoxynucleotidyl transferase to block all extendable 3' ends.
- [0019] The present disclosure also includes a method for detecting genome-wide re-arrangements in a nucleic acid genome. The method includes (a) contacting a genomic DNA sample obtained from a cell or tissue contacted with a site-specific nuclease with a plurality of transposons comprising a first sequence tag at the 5' end of the transposon, under conditions whereby the plurality of transposons is inserted into the genomic DNA sample and the genomic DNA sample is tagmented into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment; (b) contacting the plurality of tagmented nucleic acid fragments with a sequence specific cleavage reagent; (c) performing a first amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using (i) a first target-specific primer comprising a nucleotide sequence complementary to a region adjacent to a target site, (ii) a second primer comprising a nucleotide sequence identical to a portion of the sequence tag located at the 5' ends of the fragment and ; (d) sequencing the primary amplification products.
- [0020] In one embodiment, the sequence specific cleavage reagent is an enzymatic reagent, or a chemical cleavage agent. In one embodiment, the enzymatic cleavage agent comprises a CRISPR Cas/gRNA complex.
- [0021] In one embodiment, the sequence tag comprises a sequence orthogonal to the genome, a unique molecular identifier (UMI), a transposase recognition site, an index sequence, or a combination thereof. In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a transposase recognition site, and a UMI. In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a transposase recognition site, a UMI, and an index sequence.
- [0022] In one embodiment, the first and second primers comprise second and third sequence tags.

- [0023] In one embodiment, the method further includes prior to (d) performing a second amplification reaction using third and fourth nested primers comprising nucleotide sequences complementary to sequences proximal to the inside ends of the amplified products of (c) and additional sequence tags at the 5' ends of the nested primers, to produce secondary amplification products comprising a re-arranged target sequence and additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fifth and sixth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.
- [0024] In one embodiment, the third and/or the fourth primers include barcode sequences.
- [0025] In one embodiment, the fifth and sixth primers include a sequencing tag and/or an index sequence.
- [0026] In one embodiment, the method further includes prior to (d) performing a second, hemi-nested amplification reaction using the second primer and a third nested primer wherein the third primer comprises a nucleotide sequence complementary to a sequence proximal to target-specific primer end of the amplified products of (c) and optionally a fourth and fifth sequence tags at the 5' ends of the second and/or third primers to produce secondary amplification products comprising a re-arranged target sequence and one or two additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fourth and fifth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.
- [0027] In one embodiment, the second and/or the third primers comprise barcode sequences.
- [0028] In one embodiment, the fourth and fifth primers comprise a sequencing tag and/or an index sequence.
- [0029] In one embodiment, the method further includes performing a separate amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using a second target specific primer on the opposite side of the target site.
- [0030] The present disclosure is also directed to a method for detecting genome-wide re-arrangements in a nucleic acid genome. The method includes (a) contacting a genomic DNA sample obtained from a cell or tissue contacted with a site specific nuclease with a plurality of transposons comprising a first sequence tag at the 5' end of the transposon, wherein the

sequence tag comprises an RNA promoter sequence, under conditions whereby the plurality of transposons is inserted into the genomic DNA sample and the genomic DNA sample is tagged into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment; (b) transcribing the plurality of tagged nucleic acid fragments into RNA; (c) contacting the RNA with target sequence specific DNA oligonucleotide probes and RNase H; (d) performing a reverse transcription amplification reaction using a (i) a first primer comprising a nucleotide sequence complementary to a portion of the sequence tag located at the 3' or 5' end of the fragment and (ii) a second primer comprising a nucleotide sequence complementary to a region adjacent to a target site to produce primary amplification products; (e) sequencing the primary amplification products.

[0031] In one embodiment, the sequence tag comprises a sequence orthogonal to the genome, a unique molecular identifier (UMI), a transposase recognition site, an index sequence, or a combination thereof. In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a transposase recognition site, and a UMI. In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a transposase recognition site, a UMI, and an index sequence.

[0032] In one embodiment, the first and second primers comprise sequence tags.

[0033] In one embodiment, the method further includes prior to (e) performing a second amplification reaction using third and fourth nested primers comprising nucleotide sequences complementary to sequences proximal to the inside ends of the amplified products of (d) and additional sequence tags, to produce secondary amplification products comprising a re-arranged target sequence and additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fifth and sixth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

[0034] In one embodiment, the third and/or fourth primers include barcode sequences.

[0035] In one embodiment, the fifth and sixth primers comprise an adapter sequence and/or an index sequence.

[0036] In one embodiment, the method further includes prior to (e) performing a second, hemi-nested amplification reaction using the second primer and a third nested primer wherein the

third primer comprises a nucleotide sequence complementary to a sequence proximal to target-specific primer end of the amplified products of (d) and optionally a fourth and fifth sequence tags at the 5' ends of the second and/or third primers to produce secondary amplification products comprising a re-arranged target sequence and one or two additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fourth and fifth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

[0037] In one embodiment, the second and/or the third primers comprise barcode sequences.

[0038] In one embodiment, fourth and fifth primers comprise an adapter sequence and/or an index sequence.

[0039] In one embodiment, the method further includes wherein (d) further comprises performing a separate amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using a second target specific primer on the opposite side of the target site.

BRIEF DESCRIPTION OF THE DRAWINGS

[0040] The features of the present disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

[0041] FIG. 1 shows the workflow of an exemplary method of the present disclosure.

[0042] FIG. 2 shows the workflow of an exemplary method of the present disclosure.

[0043] FIG. 3 shows the workflow of an exemplary method of the present disclosure.

[0044] FIG. 4 shows the workflow of an exemplary method of the present disclosure.

[0045] FIG. 5 shows the workflow of an exemplary method of the present disclosure.

[0046] FIG. 6 shows the workflow of an exemplary method of the present disclosure.

[0047] FIG. 7 shows the workflow of an exemplary method of the present disclosure.

[0048] FIG. 8 shows the workflow of an exemplary method of the present disclosure.

[0049] FIG. 9 shows the workflow of an exemplary method of the present disclosure.

[0050] FIG. 10A and FIG. 10B show the blocking effect of blocker oligo.

[0051] FIG. 11A and FIG. 11B show the numbers of split reads identified in example 2 of an exemplary method of the present disclosure.

[0052] FIG. 12A and FIG. 12B show the validation of an inter-chromosomal translocation identified in the genomic DNA used for all methods of the present disclosure.

[0053] FIG. 13A and FIG. 13B show the numbers of split reads identified in example 4 of an exemplary method of the present disclosure.

[0054] FIG. 14 shows the numbers of split reads identified in example 6 of an exemplary method of the present disclosure.

[0055] FIG. 15A and FIG. 15B show the numbers of split reads identified in example 8 of an exemplary method of the present disclosure.

DETAILED DESCRIPTION

[0056] The present disclosure relates generally to new approaches for detecting genomic rearrangements that are associated with genome editing at specific target sites. These methods address the problem with current detection technologies, due to the requirement of significant amounts of input genomic DNA, low sensitivity, and high background. Provided herein are, *inter alia*, improved methods for detection of rearranged nucleic acid molecules which involve tagmentation and negative selection techniques that enable high detection sensitivity and cost-efficiency when compared to currently available techniques for detecting genomic rearrangements. For example, some methods of the disclosure involve the steps of tagmentation and subsequent amplification utilizing sequence tags that allow individual molecular recombination events to be identified. Other methods of the disclosure involve the steps of tagmentation, sequence specific cleavage, and subsequent amplification utilizing sequence tags. The present disclosure also provides methods that involve tagmentation, RNA transcription, and subsequent reverse transcription utilizing sequencing tags.

[0057] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[0058] Although various features of the disclosure can be described in the context of a single embodiment, the features can also be provided separately or in any suitable combination. Conversely, although the present disclosure can be described herein in the context of separate embodiments for clarity, the present disclosure can also be implemented in a single embodiment.

DEFINITION

[0059] The singular form “a”, “an”, and “the” include plural references unless the context clearly dictates otherwise. “A and/or B” is used herein to include all of the following alternatives:

“A”, “B”, “A or B”, and “A and B.”

[0060] “Barcode,” as used herein, refers to one or more known nucleotide sequences that are used to identify a nucleic acid with which the barcode is associated. In some embodiments, the barcode sequence enables multiplexing of products derived from different target sites in separate reactions or different samples.

[0061] “UMI” (unique molecular identifier) as used herein, refers to one or more randomized or semi-randomized nucleotide sequences that are used to group and deduplicate all products derived from a single starting molecule after sequencing.

[0062] “Cleavage”, as used herein, refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or cohesive ends.

[0063] As used herein, the term “detecting” a nucleic acid molecule or fragment thereof refers to determining the presence of the nucleic acid molecule, typically when the nucleic acid molecule or fragment thereof has been fully or partially separated from other components of a sample or composition.

[0064] As used herein, the term “nuclease” refers to a polypeptide capable of cleaving the phosphodiester bonds between the nucleotide subunits of nucleic acids.

[0065] As used herein, the terms “nucleic acid”, “nucleic acid molecule” or “polynucleotide” are used herein interchangeably. They refer to a polymer of deoxyribonucleotides or ribonucleotides in either single- or double-stranded form, and unless otherwise stated, encompass known analogs of natural nucleotides that can function in a similar manner as naturally occurring nucleotides. The terms encompass nucleic acid-like structures with synthetic backbones, as well as amplification products. DNAs and RNAs are both polynucleotides. The polymer may include natural nucleosides (i.e., adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and

deoxycytidine), nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, C5-propynylcytidine, C5-propynyluridine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-methylcytidine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine), chemically modified bases (2'-O,4'-C-methylene bridged/locked nucleic acid), biologically modified bases (*e.g.*, methylated bases), intercalated bases, spacers, modified sugars (*e.g.*, 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose), or modified phosphate groups (*e.g.*, phosphorothioates and 5'-N-phosphoramidite linkages).

[0066] As used herein, the term “oligonucleotide” refers to a string of nucleotides or analogues thereof. Oligonucleotides may be obtained by a number of methods including, for example, chemical synthesis, restriction enzyme digestion or PCR. As will be appreciated by one skilled in the art, the length of an oligonucleotide (*i.e.*, the number of nucleotides) can vary widely, often depending on the intended function or use of the oligonucleotide. Generally, oligonucleotides comprise between about 5 and about 300 nucleotides, for example, between about 15 and about 200 nucleotides, between about 15 and about 100 nucleotides, or between about 15 and about 50 nucleotides. Throughout the specification, whenever an oligonucleotide is represented by a sequence of letters (chosen from the four base letters: A, C, G, and T, which denote adenosine, cytidine, guanosine, and thymidine, respectively), the nucleotides are presented in the 5' to 3' order from the left to the right. In certain embodiments, the sequence of an oligonucleotide includes one or more degenerate residues described herein.

[0067] As used herein, the terms “amplify”, “amplified”, or “amplifying” as used in reference to a nucleic acid or nucleic acid reactions, refers to *in vitro* methods of making copies of a particular nucleic acid, such as a target nucleic acid, or a tagged nucleic acid produced, for example, by a method described herein.

[0068] A “primer” as used herein means a nucleic acid having a sequence complementary and specific to a known sequence in a target or template nucleic acid, *e.g.*, DNA. This means that they must be sufficiently complementary to hybridize with their respective strands to form the desired hybridized products and then be extendable by a DNA polymerase. In some instances, the primer has exact complementarity to the target or template nucleic acid. However, in many situations, exact complementarity is not possible or likely, and one or

more mismatches may exist which do not prevent hybridization or the formation of primer extension products using the DNA polymerase.

[0069] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the disclosure. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the disclosure, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the disclosure.

[0070] All ranges disclosed herein also encompass any and all possible sub-ranges and combinations of sub-ranges thereof. Any listed range can be recognized as sufficiently describing and enabling the same range being broken down into at least equal halves, thirds, quarters, fifths, tenths, and so forth. As a non-limiting example, each range discussed herein can be readily broken down into a lower third, middle third and upper third, and so forth. As will also be understood by one skilled in the art all language such as “up to,” “at least,” “greater than,” “less than,” and the like include the number recited and refer to ranges which can be subsequently broken down into sub-ranges as discussed above. Finally, as will be understood by one skilled in the art, a range includes each individual member. Thus, for example, a group having 1-3 articles refers to groups having 1, 2, or 3 articles. Similarly, a group having 1-5 articles refers to groups having 1, 2, 3, 4, or 5 articles, and so forth.

[0071] It is appreciated that certain features of the disclosure, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the disclosure, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments pertaining to the disclosure are specifically embraced by the present disclosure and are disclosed herein just as if each and every combination was individually and explicitly disclosed. In addition, all sub-combinations of the various embodiments and elements thereof are also specifically embraced by the present disclosure and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

[0072] Although features of the disclosures may be described in the context of a single embodiment, the features may also be provided separately or in any suitable combination. Conversely, although the disclosures may be described herein in the context of separate embodiments for clarity, the disclosures may also be implemented in a single embodiment. Any published patent applications and any other published references, documents, manuscripts, and scientific literature cited herein are incorporated herein by reference for any purpose. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

METHODS OF THE DISCLOSURE

Method I

[0073] As described in greater detail below, one aspect of the present disclosure provides a method for detecting genome-wide re-arrangements in a nucleic acid genome. The method involves (a) contacting a genomic DNA sample obtained from a cell or tissue contacted with a site specific nuclease with a plurality of transposons comprising a first sequence tag at the 5' end of the transposon, under conditions whereby the plurality of transposons is inserted into the genomic DNA sample and the genomic DNA sample is tagmented into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment. A first amplification reaction is performed to amplify the re-arranged tagmented nucleic acid fragments using (i) a first target-specific primer comprising a nucleotide sequence complementary to a region adjacent to a target site, (ii) a second primer comprising a nucleotide sequence identical to a portion of the sequence tag located at the 5' ends of the fragment and (iii) a blocking oligonucleotide comprising a nucleotide sequence complementary to a region on the distal side of the target site relative to the first target - specific primer to produce primary amplification products comprising a nucleotide sequence comprising a re-arranged target sequence. The primary amplification products are then sequenced.

Genomic DNA Samples

[0074] The cell or tissue used in the methods provided herein can be any eukaryotic cell type, including but not limited to human cells, non-human primate cells, mammalian cell types, vertebrate cell types, yeast, plant cells. These cells can include, *e.g.*, primary cells and/or

tissues, cells or tissues cultured for at least a period of time, or a combination of primary and cultured cells and/or tissues.

[0075] In some embodiments, methods described herein is performed on genomic DNA from a single cell. For example, genomic DNA from a single cell can be amplified before performing the methods described herein. Whole genome amplification methods are known in the art. Any of a variety of protocols and/or commercially available kits may be used. Examples of commercially available kits include, but are not limited to, the REPLI-g Single Cell Kit from QIAGEN, GENOMEPLEX® Single Cell Whole Genome Amplification Kit from Sigma Aldrich, Ampli1™ WGA Kit from Silicon Biosystems, and illustra Single Cell GenomiPhi DNA Amplification Kit from GE Healthcare Life Sciences.

[0076] Alternatively or additionally, methods disclosed herein can be used with genomic DNA samples from eukaryotic cells and/or tissues or from prokaryotic cells. For example, methods of the present disclosure can be performed using genomic DNA from microorganisms and/or from isolates from patients (*e.g.*, patients receiving antibiotics). In some embodiments, genomic DNA from microbial communities and/or one or more microbiomes is used, *e.g.*, for metagenomic mining. (See, for example, Delmont et al, “Metagenomic mining for microbiologists,” ISME J. 2011 December; 5(12):1837-43.)

[0077] Genomic DNA can be prepared using any of a variety of suitable methods, including, for example, certain manipulations to cells and/or tissues described herein. Exemplary, non-limiting manipulations include contacting a cell and/or tissue with a nuclease (*e.g.*, a site-specific nuclease and/or an RNA-guided nuclease) or a genome editing system comprising such a nuclease.

Site Specific Nucleases

[0078] As described above, the genomic DNA sample is obtained from a cell or tissue contacted with a site-specific nuclease. Generally, the nuclease is site-specific in that it is known or expected to cleave only at a specific sequence or set of sequences, referred to herein as the nuclease's “target site”.

[0079] In methods presently disclosed herein, contacting step(s) with the nuclease are generally carried out under conditions favorable for the cleavage by the nuclease. That is, even though a given candidate target site or variant target site might not actually be cleaved by the nuclease, the incubation conditions are such that the nuclease would have cleaved at least a significant portion (*e.g.*, at least 1%, at least 10%, at least 20%, at least 30%, at least 40%, at

least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%) of templates containing its known target site. For known and generally well-characterized nucleases, such conditions are generally known in the art and/or can easily be discovered or optimized. For newly discovered nucleases, such conditions can generally be approximated using information about related nucleases that are better characterized (*e.g.*, homologs and orthologs).

[0080] In some embodiments, the nuclease is an endonuclease. In some embodiments, the nuclease is a site-specific endonuclease (*e.g.*, a restriction endonuclease, a meganuclease, a transcription activator-like effector nucleases (TALEN), a zinc finger nuclease, etc.).

[0081] In some embodiments, the site specificity of a site-specific nuclease is conferred by an accessory molecule. For example, the CRISPR-associated (Cas) nucleases are guided to specific sites by “guide RNAs” or gRNAs as described herein. In some embodiments, the nuclease is an RNA-guided nuclease. In some embodiments, the nuclease is a CRISPR-associated nuclease.

[0082] In some embodiments, the nuclease is a homolog or an ortholog of a previously known nuclease, for example, a newly discovered homolog or ortholog.

[0083] In some embodiments, the nuclease is a base editor.

RNA-Guided Nucleases

[0084] RNA-guided nucleases according to the present disclosure include, but are not limited to, naturally-occurring Class 2 CRISPR nucleases such as Cas9, and Cas12a, as well as other nucleases derived or obtained therefrom. In functional terms, RNA-guided nucleases are defined as those nucleases that: (a) interact with (*e.g.*, complex with) a gRNA; and (b) together with the gRNA, associate with, and optionally cleave or modify, a target region of a DNA that includes (i) a sequence complementary to the targeting domain of the gRNA and, optionally, (ii) an additional sequence referred to as a “protospacer adjacent motif,” or “PAM,” which is described in greater detail below. RNA-guided nucleases can be defined, in broad terms, by their PAM specificity and cleavage activity, even though variations may exist between individual RNA-guided nucleases that share the same PAM specificity or cleavage activity. Skilled artisans will appreciate that some aspects of the present disclosure relate to systems, methods and compositions that can be implemented using any suitable RNA-guided nuclease having a certain PAM specificity and/or cleavage activity. For this reason, unless otherwise specified, the term RNA-guided nuclease should be understood as a generic term,

and not limited to any particular type (*e.g.*, Cas9 vs. Cas12a), species (*e.g.*, *S. pyogenes* vs. *S. aureus*) or variation (*e.g.*, full-length vs. truncated or split; naturally-occurring PAM specificity vs. engineered PAM specificity, etc.) of RNA-guided nuclease.

[0085] The PAM sequence takes its name from its sequential relationship to the “protospacer” sequence that is complementary to gRNA targeting domains (or “spacers”). Together with protospacer sequences, PAM sequences define target regions or sequences for specific RNA-guided nuclease/gRNA combinations.

[0086] Various RNA-guided nucleases may require different sequential relationships between PAMs and protospacers. In general, Cas9s recognize PAM sequences that are 3' of the protospacer as visualized relative to the guide RNA targeting domain. Cas12a, on the other hand, generally recognizes PAM sequences that are 5' of the protospacer.

[0087] In addition to recognizing specific sequential orientations of PAMs and protospacers, RNA-guided nucleases can also recognize specific PAM sequences. *S. aureus* Cas9, for instance, recognizes a PAM sequence of NNGRRT or NNGRRV, wherein the N residues are immediately 3' of the region recognized by the gRNA targeting domain. *S. pyogenes* Cas9 recognizes NGG PAM sequences. And *F. novicida* Cas12a recognizes a TTN PAM sequence. PAM sequences have been identified for a variety of RNA-guided nucleases, and a strategy for identifying novel PAM sequences has been described by Shmakov et al., 2015, Molecular Cell 60, 385-397, Nov. 5, 2015. It should also be noted that engineered RNA-guided nucleases can have PAM specificities that differ from the PAM specificities of reference molecules (for instance, in the case of an engineered RNA-guided nuclease, the reference molecule may be the naturally occurring variant from which the RNA-guided nuclease is derived, or the naturally occurring variant having the greatest amino acid sequence homology to the engineered RNA-guided nuclease).

[0088] In addition to their PAM specificity, RNA-guided nucleases can be characterized by their DNA cleavage activity: naturally-occurring RNA-guided nucleases typically form DSBs in target nucleic acids, but engineered variants have been produced that generate only SSBs (discussed above) Ran & Hsu, et al., Cell 154(6), 1380-1389, Sep. 12, 2013 (“Ran”), incorporated by reference herein), or that that do not cut at all.

Cas9

[0089] Crystal structures have been determined for *S. pyogenes* Cas9 (Jinek et al., Science 343(6176), 1247997, 2014 (“Jinek 2014”), and for *S. aureus* Cas9 in complex with a

unimolecular guide RNA and a target DNA (Nishimasu 2014; Anders et al., Nature. 2014 Sep. 25; 513(7519):569-73 (“Anders 2014”); and Nishimasu 2015).

[0090] A naturally occurring Cas9 protein comprises two lobes: a recognition (REC) lobe and a nuclease (NUC) lobe; each of which comprise particular structural and/or functional domains. The REC lobe comprises an arginine-rich bridge helix (BH) domain, and at least one REC domain (*e.g.*, a REC1 domain and, optionally, a REC2 domain). The REC lobe does not share structural similarity with other known proteins, indicating that it is a unique functional domain. While not wishing to be bound by any theory, mutational analyses suggest specific functional roles for the BH and REC domains: the BH domain appears to play a role in gRNA:DNA recognition, while the REC domain is thought to interact with the repeat:anti-repeat duplex of the gRNA and to mediate the formation of the Cas9/gRNA complex.

[0091] The NUC lobe comprises a RuvC domain, an HNH domain, and a PAM-interacting (PI) domain. The RuvC domain shares structural similarity to retroviral integrase superfamily members and cleaves the non-complementary (*i.e.*, bottom) strand of the target nucleic acid. It may be formed from two or more split RuvC motifs (such as RuvC I, RuvCII, and RuvCIII in *S. pyogenes* and *S. aureus*). The HNH domain, meanwhile, is structurally similar to HNN endonuclease motifs, and cleaves the complementary (*i.e.*, top) strand of the target nucleic acid. The PI domain, as its name suggests, contributes to PAM specificity.

[0092] While certain functions of Cas9 are linked to (but not necessarily fully determined by) the specific domains set forth above, these and other functions may be mediated or influenced by other Cas9 domains, or by multiple domains on either lobe. For instance, in *S. pyogenes* Cas9, as described in Nishimasu 2014, the repeat:antirepeat duplex of the gRNA falls into a groove between the REC and NUC lobes, and nucleotides in the duplex interact with amino acids in the BH, PI, and REC domains. Some nucleotides in the first stem loop structure also interact with amino acids in multiple domains (PI, BH and REC1), as do some nucleotides in the second and third stem loops (RuvC and PI domains).

Cas12a

[0093] The crystal structure of *Acidaminococcus sp.* Cas12a in complex with crRNA and a double-stranded (ds) DNA target including a TTTN PAM sequence has been solved by Yamano et al. (Cell. 2016 May 5; 165(4): 949-962 (“Yamano”), incorporated by reference herein). Cas12a, like Cas9, has two lobes: a REC (recognition) lobe, and a NUC (nuclease) lobe. The REC lobe includes REC1 and REC2 domains, which lack similarity to any known

protein structures. The NUC lobe, meanwhile, includes three RuvC domains (RuvC-I, -II and -III) and a BH domain. However, in contrast to Cas9, the Cas12a REC lobe lacks an HNH domain, and includes other domains that also lack similarity to known protein structures: a structurally unique PI domain, three Wedge (WED) domains (WED-I, -II and -III), and a nuclease (Nuc) domain.

[0094] While Cas9 and Cas12a share similarities in structure and function, it should be appreciated that certain Cas12a activities are mediated by structural domains that are not analogous to any Cas9 domains. For instance, cleavage of the complementary strand of the target DNA appears to be mediated by the Nuc domain, which differs sequentially and spatially from the HNH domain of Cas9. Additionally, the non-targeting portion of Cas12a gRNA (the handle) adopts a pseudoknot structure, rather than a stem loop structure formed by the repeat:antirepeat duplex in Cas9 gRNAs.

Base Editors

[0095] Engineered base editors have been developed that provide base-modifying enzyme domains (e.g., deaminases) or reverse transcriptases along with modified CRISPR associated targeting domains, e.g., Cas9 nickase (Gaudelli et al, Nature, 2017, 24644; Komor et al, Nature, 2016,533: 420–424; Yang et al, Nat Commun. 2016;7: 13330, Anzalone, Nature. 2019, . doi:10.1038/s41586-019-1711-4). Such base editors and prime editors typically introduce a single nick on one strand at the target site and then introduce one or more base modifications that are resolved to base edits after DNA replication. It has generally been thought that such enzymes would not induce double strand breaks and consequent rearrangements. However, it has been shown that single-strand nicks can indeed lead to double strand breaks as the result of replisome disassembly and DNA breakage (Vrtis et al, Mol Cell. 2021;81: 1309–1318.e6.) so homologous recombination and large-scale rearrangements may occur because of DNA breakage at sites where DNA nicking occurs. Therefore, the methods of this invention are also applicable to the detection of rearrangements induced by base editors, prime editors, and other site-specific gene editors that introduce single-stranded nicks.

Nucleic Acids Encoding RNA-Guided Nucleases

[0096] Nucleic acids encoding RNA-guided nucleases, e.g., Cas9, Cas12a or functional fragments thereof, are provided herein. Exemplary nucleic acids encoding RNA-guided nucleases have been described previously (see, e.g., Cong et al., Science. 2013 Feb. 15;

339(6121):819-23 (“Cong 2013”); Wang et al., PLoS One. 2013 Dec. 31; 8(12):e85650 (“Wang 2013”); Mali 2013; Jinek 2012).

[0097] In some cases, a nucleic acid encoding an RNA-guided nuclease can be a synthetic nucleic acid sequence. For example, the synthetic nucleic acid molecule can be chemically modified. In certain embodiments, an mRNA encoding an RNA-guided nuclease will have one or more (*e.g.*, all) of the following properties: it can be capped; polyadenylated; and substituted with 5-methylcytidine and/or pseudouridine.

[0098] Synthetic nucleic acid sequences can also be codon optimized, *e.g.*, at least one non-common codon or less-common codon has been replaced by a common codon. For example, the synthetic nucleic acid can direct the synthesis of an optimized messenger mRNA, *e.g.*, optimized for expression in a mammalian expression system, *e.g.*, described herein. Examples of codon optimized Cas9 coding sequences are presented in WO 2016/073990 (“Cotta-Ramusino”).

[0099] In addition, or alternatively, a nucleic acid encoding an RNA-guided nuclease may comprise a nuclear localization sequence (NLS). Nuclear localization sequences are known in the art.

Guide RNA (gRNA) Molecules

[0100] The terms “guide RNA” and “gRNA” refer to any nucleic acid that promotes the specific association (or “targeting”) of an RNA-guided nuclease such as a Cas9 or a Cas12a to a target sequence such as a genomic or episomal sequence in a cell. gRNAs can be unimolecular (comprising a single RNA molecule, and referred to alternatively as chimeric), or modular (comprising more than one, and typically two, separate RNA molecules, such as a crRNA and a tracrRNA, which are usually associated with one another, for instance by duplexing). gRNAs and their component parts are described throughout the literature, for instance in Briner et al. (Molecular Cell 56(2), 333-339, Oct. 23, 2014 (“Briner”), which is incorporated by reference), and in Cotta-Ramusino. In bacteria and archaea, type II CRISPR systems generally comprise an RNA-guided nuclease protein such as Cas9, a CRISPR RNA (crRNA) that includes a 5′ region that is complementary to a foreign sequence, and a trans-activating crRNA (tracrRNA) that includes a 5′ region that is complementary to, and forms a duplex with, a 3′ region of the crRNA. While not intending to be bound by any theory, it is thought that this duplex facilitates the formation of—and is necessary for the activity of—the Cas9/gRNA complex. As type II CRISPR systems were adapted for use in gene editing, it

was discovered that the crRNA and tracrRNA could be joined into a single unimolecular or chimeric guide RNA, in one non-limiting example, by means of a four nucleotide (*e.g.*, GAAA) “tetraloop” or “linker” sequence bridging complementary regions of the crRNA (at its 3' end) and the tracrRNA (at its 5' end). (Mali et al. *Science*. 2013 Feb. 15; 339(6121): 823-826 (“Mali 2013”); Jiang et al. *Nat Biotechnol*. 2013 March; 31(3): 233-239 (“Jiang”); and Jinek et al., 2012 *Science* August 17; 337(6096): 816-821 (“Jinek 2012”), all of which are incorporated by reference herein.)

[0101] Guide RNAs, whether unimolecular or modular, include a “targeting domain” that is fully or partially complementary to a target domain within a target sequence, such as a DNA sequence in the genome of a cell where editing is desired. Targeting domains are referred to by various names in the literature, including without limitation “guide sequences” (Hsu et al., *Nat Biotechnol*. 2013 September; 31(9): 827-832, (“Hsu”), incorporated by reference herein), “complementarity regions” (Cotta-Ramusino), “spacers” (Briner) and generically as “crRNAs” (Jiang). Irrespective of the names they are given, targeting domains are typically 10-30 nucleotides in length, and in certain embodiments are 16-24 nucleotides in length (for instance, 16, 17, 18, 19, 20, 21, 22, 23 or 24 nucleotides in length), and are at or near the 5' terminus of in the case of a Cas9 gRNA, and at or near the 3' terminus in the case of a Cas12a gRNA.

[0102] In addition to the targeting domains, gRNAs typically (but not necessarily, as discussed below) include a plurality of domains that may influence the formation or activity of gRNA/Cas9 complexes. For instance, as mentioned above, the duplexed structure formed by first and secondary complementarity domains of a gRNA (also referred to as a repeat:anti-repeat duplex) interacts with the recognition (REC) lobe of Cas9 and can mediate the formation of Cas9/gRNA complexes. (Nishimasu et al., *Cell* 156, 935-949, Feb. 27, 2014 (“Nishimasu 2014”) and Nishimasu et al., *Cell* 162, 1113-1126, Aug. 27, 2015 (“Nishimasu 2015”), both incorporated by reference herein). It should be noted that the first and/or second complementarity domains may contain one or more poly-A tracts, which can be recognized by RNA polymerases as a termination signal. The sequence of the first and second complementarity domains are, therefore, optionally modified to eliminate these tracts and promote the complete *in vitro* transcription of gRNAs, for instance through the use of A-G swaps as described in Briner, or A-U swaps. These and other similar modifications to the first and second complementarity domains are within the scope of the present disclosure.

[0103] Along with the first and second complementarity domains, Cas9 gRNAs typically include two or more additional duplexed regions that are involved in nuclease activity in vivo but not necessarily in vitro. (Nishimasu 2015). A first stem-loop one near the 3' portion of the second complementarity domain is referred to variously as the "proximal domain," (Cotta-Ramusino) "stem loop 1" (Nishimasu 2014 and 2015) and the "nexus" (Briner). One or more additional stem loop structures are generally present near the 3' end of the gRNA, with the number varying by species: *S. pyogenes* gRNAs typically include two 3' stem loops (for a total of four stem loop structures including the repeat:anti-repeat duplex), while *S. aureus* and other species have only one (for a total of three stem loop structures). A description of conserved stem loop structures (and gRNA structures more generally) organized by species is provided in Briner.

[0104] While the foregoing description has focused on gRNAs for use with Cas9, it should be appreciated that other RNA-guided nucleases have been (or may in the future be) discovered or invented which utilize gRNAs that differ in some ways from those described to this point. For instance, Cas12a ("CRISPR from *Prevotella* and *Franciscella* 1") is a recently discovered RNA-guided nuclease that does not require a tracrRNA to function. (Zetsche et al., 2015, *Cell* 163, 759-771 Oct. 22, 2015 ("Zetsche I"), incorporated by reference herein). A gRNA for use in a Cas12a genome editing system generally includes a targeting domain and a complementarity domain (alternately referred to as a "handle"). It should also be noted that, in gRNAs for use with Cas12a, the targeting domain is usually present at or near the 3' end, rather than the 5' end as described above in connection with Cas9 gRNAs (the handle is at or near the 5' end of a Cas12a gRNA).

[0105] Those of skill in the art will appreciate, however, that although structural differences may exist between gRNAs from different prokaryotic species, or between Cas12a and Cas9 gRNAs, the principles by which gRNAs operate are generally consistent. Because of this consistency of operation, gRNAs can be defined, in broad terms, by their targeting domain sequences, and skilled artisans will appreciate that a given targeting domain sequence can be incorporated in any suitable gRNA, including a unimolecular or chimeric gRNA, or a gRNA that includes one or more chemical modifications and/or sequential modifications (substitutions, additional nucleotides, truncations, etc.). Thus, for economy of presentation in this disclosure, gRNAs may be described solely in terms of their targeting domain sequences.

[0106] More generally, skilled artisans will appreciate that some aspects of the present disclosure relate to systems, methods and compositions that can be implemented using multiple RNA-guided nucleases. For this reason, unless otherwise specified, the term gRNA should be understood to encompass any suitable gRNA that can be used with any RNA-guided nuclease, and not only those gRNAs that are compatible with a particular species of Cas9 or Cas12a. By way of illustration, the term gRNA can, in certain embodiments, include a gRNA for use with any RNA-guided nuclease occurring in a Class 2 CRISPR system, such as a type II or type V or CRISPR system, or an RNA-guided nuclease derived or adapted therefrom.

Transposons and Tagmentation

[0107] In the first step, a genomic DNA sample obtained from a cell or tissue contacted with a site-specific nuclease is contacted with a plurality of transposons to introduce known DNA sequences onto the ends of genomic DNA fragments. For example, Tn5 transposase catalyzes strand transfer via nucleophilic attack on the target DNA by activated 3-OH groups at the transposon ends, leaving a 9-bp gap at the target site (Vaezslami et al, J Bacteriol, 2007;189: 7436–7441). A transposome is a complex of a transposase enzyme and DNA which comprises a transposon end sequence (also known as "transposase recognition sequence" or "mosaic end" (ME)). When a double-stranded oligonucleotides with a transposon recognition sequence at one end are provided, transposition results in the introduction of free DNA ends into the target molecule (*e.g.*, "tagmentation"; Adey et al, Genome Biol. 2010;11: R119). This system can be adapted using hyperactive transposase enzymes and modified DNA oligonucleotides (sequence tags) comprising MEs to introduce tags on both strands of the tagged products with functional DNA molecules (*e.g.*, primer binding sites). Any transposase enzyme with tagmentation activity, *e.g.*, any transposase enzyme capable of mediating strand transfer and integration of oligonucleotides (*e.g.*, tags) at the ends of the tagged DNA, can be used. In some embodiments, the transposase is any transposase capable of conservative transposition. In some embodiments, the transposase is a cut and paste transposase. Other kinds of transposase are known in the art and are within the scope of this disclosure. For example, suitable transposase enzymes include, without limitation, Tn5, Tn5059, Mos-1, HyperMu™, Hermes, Tn7, or any functional variant or derivative of the previously listed transposase enzymes.

- [0108] For instance, the Tn5 transposase may be produced as purified protein monomers. Tn5 transposase is also commercially available (*e.g.*, manufacturer Illumina, Illumina.com, Catalog No. 15027865, TD Tagment DNA Buffer Catalog No.15027866). These can be subsequently loaded with the oligonucleotides of interest, *e.g.*, ssDNA oligonucleotides containing MEs for Tn5 recognition and additional functional sequences (*e.g.*, primer binding sites, *e.g.*, UMIs) are annealed to form a dsDNA mosaic end oligonucleotide (MEDS) that is recognized by Tn5 during dimer assembly (*e.g.*, transposome dimerization). In some embodiments, a hyperactive Tn5 transposase can be loaded with tags (*e.g.*, oligonucleotides of interest) which can simultaneously fragment and tag a genome with the desired sequences.
- [0109] As described above, the transposons include a first sequence tag at the 5' end of the transposon. As used herein, the first sequence tag refers to non-target nucleic acid component, generally DNA, that provides a means of addressing a nucleic acid fragment to which it is joined. For example, in some embodiments, a sequence tag is or comprises a nucleotide sequence that permits identification, recognition, and/or molecular or biochemical manipulation of the DNA to which the tag is attached (*e.g.*, by providing a site for annealing an oligonucleotide, such as a primer for extension by a DNA polymerase, or an oligonucleotide for capture or for a ligation reaction). In some embodiments, sequencing tags are used for generation of templates for next-generation sequencing for a particular sequencing platform (*e.g.*, sequencing tags for: an Illumina sequencing platform; for a Thermo Fisher Ion Torrent sequencing platform; for a Pacific Biosciences' Sequel sequencing platform; for a MGI sequencing platform; or for any other sequencing platform). In some embodiments, a sequencing tag is a full-length Illumina forward (i5) adapter. In some embodiments, a sequencing tag is a full-length Illumina reverse (i7) adapter. In some embodiments, the sequence tag includes a unique sequence (*e.g.* a sequence orthogonal to the genome), a unique molecular identifier (UMI), a transposase recognition site, an index sequence, or a combination thereof. In some embodiments, the sequence tag includes a sequence orthogonal to the genome, a unique molecular identifier (UMI), and a transposase recognition site. In some embodiments, the sequence tag includes a sequence orthogonal to the genome, a unique molecular identifier (UMI), an index sequence, and a transposase recognition site.
- [0110] In some embodiments, the UMI is a randomly generated sequence. In some embodiments, the UMI is between eight and 20 nucleotides in length, for example, between 10 and 16

nucleotides in length, such as 10, 11, 12, 13, 14, 15, and 16 nucleotides in length. The production and use of UMIs in various contexts are known in the art.

[0111] Tagmentation is a common initial step for all methods described herein. In some embodiments, the step of fragmenting the genomic DNA in cells of the biological sample comprises contacting the biological sample containing the genomic DNA with the transposase enzyme (*e.g.*, a transposome, *e.g.*, a reaction mixture (*e.g.*, solution)) including a transposase), under any suitable conditions. In some embodiments, the transposome is assembled by annealing transposon-end containing oligonucleotides for about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 minutes at about 95, 96, 97, 98, or 99°C, followed by about 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, or 500 cycles of 1 minutes at temperatures of about 80, 85, 90, 95, or 99°C while decreasing temperature 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, or 2°C every cycle. The transposase and annealed oligonucleotides can then be incubated at 15, 20, 25, 30, 35, 40°C for 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120 minutes. In some embodiments, such suitable conditions result in the fragmentation (*e.g.*, tagmentation) of the genomic DNA of cells present in the biological sample. Typical conditions will depend on the transposase enzyme used and can be determined using routine methods known in the art. Therefore, suitable conditions can be conditions (*e.g.*, buffer, salt, concentration, pH, temperature, time conditions) under which the transposase enzyme is functional, *e.g.*, in which the transposase enzyme displays transposase activity, particularly tagmentation activity, in the biological sample. In some embodiments, the tagmentation reaction involves mixing about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, or 500 ng genomic DNA with about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1, 2, 3, 4, 5, 6, 7 or 8 µL of the transposome and incubation at 30, 35, 40, 45, 50, 55, 60, or 65°C for about 5, 10, 15, 20, 23, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, or 120 minutes.

[0112] In one non-limiting example, the reaction mixture comprises a transposase enzyme in a buffered solution (*e.g.*, Tris-acetate) having a pH of about 6.5 to about 8.5, *e.g.*, about 7.0 to about 8.0 such as about 7.5. Additionally or alternatively, the reaction mixture can be used at any suitable temperature based on the temperature optimum of the transposase enzyme, such as about 10° to about 55°C, *e.g.*, about 10° to about 54°, about 11° to about 53°, about 12° to about 52°, about 13° to about 51°, about 14° to about 50°, about 15° to about 49°, about 16° to about 48°, about 17° to about 47°C, *e.g.*, about 10°, about 12°, about 15°, about 18°, about

20°, about 22°, about 25°, about 28°, about 30°, about 33°, about 35°, about or 37°C, preferably about 30° to about 40°C, *e.g.*, about 37°C for Tn5. In some embodiments, the transposase enzyme can be contacted with the biological sample for about 10 minutes to about one hour. In some embodiments, the transposase enzyme can be contacted with the biological sample for about 20, about 30, about 40, or about 50 minutes. In some embodiments, the transposase enzyme can be contacted with the biological sample for about 1 hour to about 4 hours.

[0113] Briefly, in some embodiments, Tn5 tagmentation uses staggered transposon ends with an optional 3' extension block (*e.g.*, a 3' dideoxy nucleotide, inverted nucleotide, or amino modifier) and 5' sequence tags. As described above, in some embodiments, the sequence tag harbors a UMI that typically comprises a random 12-nucleotide sequences that differs between each transposon end. The UMIs allow the products generated from a single starting genomic fragment to be identified, grouped and analyzed together after multiple amplification steps. This allows individual molecular recombination events to be identified, and their frequency of occurrence in the population to be quantified relative to other unique events. It also allows variants associated with each haplotype in the starting genome to be identified. The use of tagmentation also avoids inefficient and time-consuming DNA shearing, end repair and adaptor ligation steps, reducing input DNA requirements, and speeding up the process.

[0114] In the method described herein, once the plurality of transposons is inserted into the genomic DNA sample, the genomic DNA sample is tagmented into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment.

Amplification

[0115] In the method described herein, after tagmentation, a first amplification reaction is performed to amplify the re-arranged tagmented nucleic acid fragments using (i) a first target-specific primer comprising a nucleotide sequence complementary to a region adjacent to a target site, (ii) a second primer comprising a nucleotide sequence identical to a portion of the sequence tag located at the 5' ends of the fragment and (iii) a blocking oligonucleotide comprising a nucleotide sequence complementary to a region on the distal side of the target site relative to the first target-specific primer to produce primary amplification products comprising a nucleotide sequence comprising a re-arranged target sequence.

- [0116] Numerous methods of amplifying nucleic acids are known in the art, and amplification reactions include polymerase chain reactions (PCR), ligase chain reactions, strand displacement amplification reactions, rolling circle amplification reactions, transcription-mediated amplification methods such as NASBA (*e.g.*, U.S. Pat. No. 5,409,818), loop mediated amplification methods (*e.g.*, “LAMP” amplification using loop-forming sequences, *e.g.*, as described in U.S. Pat. No. 6,410,278). A nucleic acid that is amplified can be DNA comprising, consisting of, or derived from DNA or RNA or a mixture of DNA and RNA, including modified DNA and/or RNA. The products resulting from amplification of a nucleic acid molecule or molecules (*i.e.*, “amplification products”), whether the starting nucleic acid is DNA, RNA or both, can be either DNA or RNA, or a mixture of both DNA and RNA nucleosides or nucleotides, or they can comprise modified DNA or RNA nucleosides or nucleotides. A “copy” does not necessarily mean perfect sequence complementarity or identity to the target sequence. For example, copies can include nucleotide analogs such as deoxyinosine or deoxyuridine, intentional sequence alterations (such as sequence alterations introduced through a primer comprising a sequence that is hybridizable, but not complementary, to the target sequence), and/or sequence errors that occur during amplification. In some embodiments, the first amplification reaction is a PCR.
- [0117] Protocols for amplification can include, *e.g.*, one round of amplification or multiple rounds of amplification. For example, a first amplification round can be followed by a second amplification round with or without one or more processing (*e.g.*, cleanup, concentrating, etc.) steps in between the two rounds of amplification. Additional rounds of amplification may be used in some embodiments, with or without one or more processing steps in between.
- [0118] The number of amplification cycles can be varied depending on the embodiment. For example, each round of amplification can comprise at least 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 38, or 40 cycles. In embodiments involving more than one round of amplification, the number of amplification cycles used in each round can be the same or it can be different. In cases in which two rounds of amplification are used and the number of amplification cycles in the two rounds differ, the first round can comprise more cycles or it may comprise fewer cycles than the second round. As an example, a first round of amplification can comprise 12 cycles and a second round of amplification can comprise 15 cycles. As another example, a first round of amplification can comprise 10 cycles and a second round of amplification can comprise 12 cycles.

[0119] The temperature of the amplification reaction will vary depending upon the step of the reaction. In some embodiments, an initial denaturing can be performed at about 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99°C for about 1, 2, 3, 4, or 5 minutes, and a final extension can be performed at about 67, 68, 69, 70, 71, 72, 73, 74, or 75°C for about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 minutes.

Blocking Oligonucleotides

[0120] As described above, the first amplification reaction is performed to amplify the re-arranged tagged nucleic acid fragments using (i) a first target-specific primer comprising a nucleotide sequence complementary to a region adjacent to a target site, (ii) a second primer comprising a nucleotide sequence identical to a portion of the sequence tag located at the 5' ends of the fragment and (iii) a blocking oligonucleotide comprising a nucleotide sequence complementary to a region on the distal side of the target site relative to the first target-specific primer.

[0121] As used herein, a blocking oligonucleotide is an engineered single stranded nucleic acid sequence. The blocking oligonucleotide can be comprised of single stranded DNA, RNA, peptide nucleic acid (PNA), locked nucleic acid (LNA), bridged-nucleic acid (BNA), and/or other modified nucleotides. In some embodiments, it is a DNA oligonucleotide that includes multiple modified bases. The blocking oligonucleotide may comprise 10 to 100 nucleotides, but is preferably between 15 and 50 nucleotides in length.

[0122] The blocking oligonucleotide has strong avidity for the template DNA but cannot be extended by the polymerase used for amplification due to an absent or blocked 3'OH, and modifications at or near the 3' end that confer resistance to any 3'→5' exonuclease activity of the polymerase (*e.g.*, multiple phosphorothioate bonds at the 3' end, spacers, inverted nucleotides, LNA, or BNA residues). The blocking oligonucleotide must be complementary to the same strand as the first target specific primer and must hybridize on the distal side of the gene editor target site relative to the location of the first target specific primer. In one embodiment, the blocking oligonucleotide includes an absent or blocked 3' OH, spacers, inverted nucleotides, or other modifications to block extension of the 3' end.

[0123] In one embodiment, the blocking oligonucleotide comprises one or more phosphorothioate bonds, spacers, or other modifications at the 3' and 5' ends, to block polymerase extension and exonuclease digestion, and a backbone of DNA, RNA, LNA, BNA, PNA, modified nucleic acids, or a combination thereof.

[0124] In some embodiments, the blocking nucleotide includes nucleotides that increase duplex stability and melting temperature relative to native DNA. This enables hybridization of the blocking oligonucleotide at a higher temperature than the first and second primers (and therefore enabling hybridization to the template before the primers are able to hybridize during amplification). High duplex stability also reduces the likelihood of dissociation of the blocking oligonucleotide and read-through by the polymerase during primer extension.

[0125] In some embodiments, the blocking oligonucleotide includes a chemical- or photo-crosslinking moiety to permit efficient covalent attachment to the hybridized DNA strand, preventing polymerase extension and dissociation. Exemplary chemical or photo-crosslinking moieties include, without limitation, psoralen, click chemistry, and 3-cyanovinylcarbazole.

Sequencing

[0126] The use of the blocking oligonucleotide, first target-specific primer, and second primer in the amplification reaction produces primary amplification products comprising a nucleotide sequence comprising a re-arranged target sequence. The primary amplification products are then sequenced.

[0127] As used herein, "sequencing" includes any method of determining the sequence of a nucleic acid. Any method of sequencing can be used in the present methods, including chain terminator (Sanger) sequencing and dye terminator sequencing. In preferred embodiments, Next Generation Sequencing (NGS), a high-throughput sequencing technology that performs thousands or millions of sequencing reactions in parallel, is used. Although different NGS platforms use varying assay chemistries, they all generate sequence data from a large number of sequencing reactions run simultaneously on a large number of templates. Typically, the sequence data is collected using a scanner, and then assembled and analyzed bioinformatically. Thus, the sequencing reactions are performed, read, assembled, and analyzed in parallel. Exemplary approaches, systems, or techniques may be performed using nucleic acid amplification, polymerase chain reaction (PCR) (*e.g.*, digital PCR and droplet digital PCR (ddPCR), quantitative PCR, real time PCR, multiplex PCR, PCR-based singleplex methods, emulsion PCR), and/or isothermal amplification. Non-limiting examples of nucleic acid sequencing methods include Maxam-Gilbert sequencing and chain-termination methods, *de novo* sequencing methods including shotgun sequencing and bridge PCR, next-generation methods including Polony sequencing, 454 pyrosequencing, Illumina

sequencing, Ion Torrent semiconductor sequencing, SMRT® sequencing, and Oxford Nanopore Technology sequencing.

[0128] In some embodiments, the primary reaction products can be purified to remove primers, and quantified. In some embodiments, the sample can be split, and a symmetrical set of primers would be used to evaluate both sides of the target site for rearrangements in separate reactions (for this and subsequent steps). The primary reaction products can also be used to detect rearrangements at other genomic loci, such as candidate or bona fide off-target sites predicted by *in silico* algorithms, or detected by other experimental methods.

Additional Amplification

[0129] As described above, the first amplification round can be followed by a second amplification round. Thus, in some embodiments, the second amplification includes nested PCR and the introduction of second and third sequence tags. In some embodiments, the second amplification includes hemi-nested PCR.

[0130] In accordance with such embodiments where a second amplification round is used, the method disclosed herein can further include prior to step (c) performing a second amplification reaction using third and fourth nested primers. The third and fourth primers include nucleotide sequences complementary to sequences proximal to the inside ends of the amplified products of (b) and fourth and fifth sequence tags at the 5' ends of the nested primers. The second amplification reaction produces secondary amplification products that include a re-arranged target sequence and additional sequence tags. This nested amplification step enriches for the desired products, since the truncated extension products, and spurious fragments derived from mis-priming in the first amplification step, cannot serve as amplification templates.

[0131] In some embodiments where a second amplification round is used, the method disclosed herein can further include prior to (c) performing a second, hemi-nested amplification reaction using the second primer and a third nested primer wherein the third primer comprises a nucleotide sequence complementary to a sequence proximal to target-specific primer end of the amplified products of (b) and optionally a fourth and fifth sequence tags at the 5' ends of the second and/or third primers to produce secondary amplification products comprising a re-arranged target sequence and one or two additional sequence tags.

[0132] For the second amplification step, protocols for amplification can include, *e.g.*, one round of amplification or multiple rounds of amplification. For example, a first amplification round

can be followed by a second amplification round with or without one or more processing (*e.g.*, cleanup, concentrating, etc.) steps in between the two rounds of amplification. Additional rounds of amplification may be used in some embodiments, with or without one or more processing steps in between.

[0133] The number of amplification cycles can be varied depending on the embodiment. For example, each round of amplification can comprise at least 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 38, or 40 cycles. In embodiments involving more than one round of amplification, the number of amplification cycles used in each round can be the same or it can be different. In cases in which two rounds of amplification are used and the number of amplification cycles in the two rounds differ, the first round can comprise more cycles or it may comprise fewer cycles than the second round. As an example, a first round of amplification can comprise 12 cycles and a second round of amplification can comprise 15 cycles. As another example, a first round of amplification can comprise 10 cycles and a second round of amplification can comprise 12 cycles.

[0134] The temperature of the amplification reaction will vary depending upon the step of the reaction. In some embodiments, an initial denaturing can be performed at about 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99°C for about 1, 2, 3, 4, or 5 minutes, and a final extension can be performed at about 67, 68, 69, 70, 71, 72, 73, 74, or 75°C for about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 minutes.

[0135] In one embodiment, the method further includes performing a third amplification reaction using a fifth and sixth primer, or a fourth and fifth primer after hemi-nested PCR, comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence. In some embodiments, the third amplification reaction is a hemi-nested PCR.

[0136] For the third amplification step, protocols for amplification can include, *e.g.*, one round of amplification or multiple rounds of amplification. For example, a first amplification round can be followed by a second amplification round with or without one or more processing (*e.g.*, cleanup, concentrating, etc.) steps in between the two rounds of amplification. Additional rounds of amplification may be used in some embodiments, with or without one or more processing steps in between.

[0137] The number of amplification cycles can be varied depending on the embodiment. For example, each round of amplification can comprise at least 4, 8, 10, 12, 14, 16, 18, 20, 22, 24,

26, 28, 30, 32, 34, 38, or 40 cycles. In embodiments involving more than one round of amplification, the number of amplification cycles used in each round can be the same or it can be different. In cases in which two rounds of amplification are used and the number of amplification cycles in the two rounds differ, the first round can comprise more cycles or it may comprise fewer cycles than the second round. As an example, a first round of amplification can comprise 12 cycles and a second round of amplification can comprise 15 cycles. As another example, a first round of amplification can comprise 10 cycles and a second round of amplification can comprise 12 cycles.

[0138] The temperature of the amplification reaction will vary depending upon the step of the reaction. In some embodiments, an initial denaturing can be performed at about 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99°C for about 1, 2, 3, 4, or 5 minutes, and a final extension can be performed at about 67, 68, 69, 70, 71, 72, 73, 74, or 75°C for about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 minutes.

[0139] In one embodiment, the third and/or the fourth primers comprise barcode sequences.

Barcode Sequences

[0140] A “barcode” is a molecular label, or identifier, that conveys or is capable of conveying information about a sample, sequencing read, or group of samples or sequencing reads (*e.g.*, information about an analyte in a sample, a template, a bead, a microwell, a primer, a read, and/or a capture probe). A barcode can be part of an analyte, or independent of an analyte. A barcode can be attached to an analyte. A particular barcode will be unique relative to other barcodes, and may include error-correcting features (*e.g.*, Hamming codes) that enable correct identification of each barcode in a mixture even in the presence of sequencing errors.

[0141] As used in here Barcodes comprise custom polynucleotides of defined sequence that are introduced by primer extension (during PCR, for example) or by adapter ligation. Barcodes can allow for identification and/or quantification of groups of sequencing reads that share the same barcode. Barcodes can be 8 to 50 or more bases in length, but are typically between 10 and 25 bases in length.

[0142] Barcodes can be used to spatially-resolve molecular components found in biological samples, for example, at single-cell resolution (*e.g.*, a barcode can be or can include a “spatial barcode”). In some embodiments, a barcode includes two or more sub-barcodes that can together function as a single barcode. For example, a polynucleotide barcode can include two

or more polynucleotide sequences (*e.g.*, sub- barcodes) that are separated by one or more non-barcode sequences.

[0143] An index sequence is a class of barcode that is used to differentiate and group sets of reads derived from different samples after sequencing, by means of standard sequencing analysis software.

Unique Molecular Identifiers

[0144] A unique molecular identifier, or “UMI”, comprises a short stretch of random nucleotides (*e.g.*, NNNNNNNN (SEQ ID NO: 90)) or semi-random nucleotides (*e.g.*, NWYYRRV (SEQ ID NO: 91)) included in a custom primer or adapter. UMIs are typically between 6 and 25 bases, and are generally introduced by tagmentation or adapter ligation, although they may also be introduced by a single round of primer extension. UMIs serve to enable the identification and grouping of extension or amplification products (*e.g.*, PCR products) that derive from a single starting genomic DNA molecule. Such grouping enables counting of unique molecules in a starting sample that are represented by multiple reads after amplification and sequencing (*e.g.*, counting of unique recombination events). UMIs can also be used to distinguishing between SNVs and sequencing errors in amplified molecules.

[0145] Individual sequencing reads may contain multiple barcodes, indexes and UMIs in addition to the sequences being characterized.

[0146] In one embodiment, the fifth and sixth primers comprise a sequencing tag and/or an index sequence.

[0147] In one embodiment, (b) further includes performing a separate amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using a second target specific primer on the opposite side of the target site.

[0148] Embodiments of the method described herein is shown in FIG. 1 and FIG. 6 (Method IA).

In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a UMI, and transposase recognition site (FIG. 1). In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a UMI, an index sequence, and transposase recognition site (FIG. 6). As shown in FIG. 1 and FIG. 6, after tagmentation, the sample can be subjected to a three-step PCR amplification procedure to enrich for fragments that have undergone recombination (Steps 2-4). Each of these steps is carried out in the presence of a blocking oligonucleotide, which prevents the products from native (non-rearranged) sequences from undergoing exponential amplification. The amplification procedure may

include a 3-step or touchdown PCR cycling protocol to ensure that the blocking oligonucleotide hybridizes before the PCR primers are extended.

[0149] In the first PCR step (Step 2), one DNA primer (labeled as “1” in FIG. 1 and FIG. 6) matches the 5’ transposon end tag (e.g., sequencing tag) and the other (labeled as “2” in FIG. 1 and FIG. 6) is complementary to a region flanking the gene editor target site that is being investigated as a possible site of rearrangements. In the presence of the blocker (labeled as “0”), DNA polymerase extension products derived from the target specific primer (2) at a native gene editor target site (right hand side of FIG. 1 and FIG. 6, below) are blocked from extension. This prevents exponential PCR amplification from occurring (note that the 3’ ends of the tagged fragments cannot be extended if we include the optional 3’ end blocking step). If a rearrangement has occurred such that the sequence complementary to the blocking oligonucleotide has been replaced by another sequence derived from elsewhere in the genome, the blocker cannot hybridize, and PCR products derived from the rearranged fragment are productively generated (left hand side of FIG. 1 and FIG. 6). Following amplification, the PCR products are purified to remove primers, and are quantified. Typically, the sample would be split, and a symmetrical set of primers would be used to evaluate both sides of the target site for rearrangements in separate reactions (for this and subsequent steps). The PCR products can be further aliquoted and used to detect rearrangements at other genomic loci, such as the off-target sites predicted by *in silico* algorithms, or detected by other experimental methods.

[0150] In the second PCR step (Step 3), a pair of nested primers (labeled as “3” and “4”, respectively in FIG. 1) are used that are complementary to sequences proximal to the inside ends of the PCR product generated in the Step 2. This nested PCR step enriches for the desired products, since the truncated extension products, and spurious fragments derived from mis-priming at other loci in the first PCR step, cannot serve as amplification templates. The nested primers contain sequence tags that can be used as priming sites for the third PCR step (Step 4), and may also contain barcodes to enable multiplexing of products derived from different target sites in separate reactions or different samples. Alternatively, in cases where the first sequencing tag includes an index primer (FIG. 6), the second PCR step involves a hemi-nested PCR step using a pair of primers (labeled as “3” and “4” in FIG. 6) in which primer “4” is complementary to a sequence proximal to the inside end of the PCR product generated in the Step 2, and primer “3” matches the 5’ transposon end tag (e.g., sequencing

tag). Once again, in the presence of the blocking oligonucleotide (blocker 0), only the PCR products derived from rearranged target molecules can be amplified further with the nested or hemi-nested primers. Fragments derived from native, non-rearranged target sites cannot be amplified (only a single, truncated extension product can be generated in each PCR cycle). Following amplification, the PCR products are purified to remove primers and quantified.

[0151] In the third PCR step (Step 4), primers (labeled “5” and “6”) complementary to the tags that were introduced by PCR in the second PCR step (Step 3) are employed to enrich further for rearranged molecules. These primers also provide the sequences required for NGS sequencing, such as the adapter sequences for cluster generation on an Illumina flow cell, UMIs and index sequences for sample multiplexing. Following amplification, the PCR products are purified to remove primers and quantified. The resulting library is then subjected to NGS sequencing on the Illumina platform, for example.

Variations of Method I

[0152] In one embodiment, the method disclosed herein further includes prior to (b) contacting the plurality of tagmented nucleic acid fragments with ddNTP or other 3' modified dNTP and a DNA polymerase (*e.g.*, Klenow fragment or terminal deoxynucleotidyl transferase) to block all extendable 3' ends immediately after tagmentation.

[0153] Tagmentation and amplification reactions can be performed as described above for Method I.

[0154] An exemplary embodiment of the method described herein is shown in FIG. 2 (Method IB). As shown in FIG. 2 and FIG. 7, after tagmentation, 3' end of the short ME fragment and the 3' end of genomic DNA at the 9-bp gap are extended and blocked by ddNTP with DNA polymerase or terminal deoxynucleotidyl transferase (TdT). This prevents any possible extension of the ME or genomic fragment ends to generate priming sites for primers 1 and 3 during PCR, as well as preventing extension of the ends due to hybridization with repetitive sequences or short sequences with homology to the 3' fragment ends. The other steps are the same as described above for FIG. 1.

[0155] In one embodiment, the sequence tag in the methods described above comprises uracil.

[0156] In accordance with this embodiment, an exemplary embodiment of the method described herein is shown in FIG. 3 and FIG. 8 (Method IC). As shown in FIG. 3 and FIG. 8, the sequence tag used to assemble transposome carries Uracil. Unlike other current protocols, there is no strand displacement step after tagmentation. Instead, DNA is purified, then a PCR

reaction is set up with primer 2 and a uracil-tolerating DNA polymerase (like Phusion U Hot Start DNA Polymerase, Cat # F555S, Thermo Fisher Scientific). The extension from primer 2 using the non-rearranged DNA as template is blocked by the blocker 0, while blocker 0 cannot block the extension using rearranged DNA as templates. This primer 2 extension step can be performed multiple times to generate multiple copies of the primer 2 extension strands. Alternatively, isothermal amplification method can be used to produce more copies, like Recombinase polymerase amplification (RPA) and strand-invasion based amplification (SIBA). Next, purified DNA is treated with USER (Uracil-Specific Excision Reagent) Enzyme (a mixture of Uracil DNA glycosylase (UDG) and DNA glycosylase-lysase Endonuclease VIII) at uracil residues synthetically incorporated into the adapter. In some embodiments, USER enzyme cleavage can be performed by mixing 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 μ L of USER enzyme with product from the previous step, and incubating at 16, 20, 25, 30, 35, 40, or 45°C for 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 minutes. The following steps of FIG. 3 and FIG. 8 are the same as FIG. 1 from Step 2 onwards.

Method II

[0157] The present disclosure also includes another method for detecting genome-wide re-arrangements in a nucleic acid genome. The method includes (a) contacting a genomic DNA sample obtained from a cell or tissue contacted with a site-specific nuclease with a plurality of transposons comprising a first sequence tag at the 5' end of the transposon, under conditions whereby the plurality of transposons is inserted into the genomic DNA sample and the genomic DNA sample is tagmented into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment. In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a UMI, and transposase recognition site (FIG. 5). In some embodiments, the sequence tag comprises a sequence orthogonal to the genome, a UMI, an index sequence, and transposase recognition site (FIG. 9). The method then includes (b) contacting the plurality of tagmented nucleic acid fragments with a sequence specific cleavage reagent; (c) performing a first amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using (i) a first target-specific primer comprising a nucleotide sequence complementary to a region adjacent to a target site, (ii) a second primer comprising a nucleotide sequence identical to a portion of the sequence tag located at the 5' ends of the fragment and; (d) sequencing the primary amplification products.

[0158] Methods and embodiments of step (a) are described above for Method I. However, unlike the methods described *supra*, which utilize blocking oligonucleotides, the method in the accordance with this aspect of the disclosure involves contacting the plurality of tagmented nucleic acid fragments with a sequence specific cleavage reagent.

Sequence Specific Cleavage Agent

[0159] In one embodiment, the sequence specific cleavage reagent is an enzymatic reagent, or a chemical cleavage agent.

[0160] In methods presently disclosed herein, contacting step(s) with the nuclease are generally carried out under conditions favorable for the cleavage by the nuclease. That is, even though a given candidate target site or variant target site might not actually be cleaved by the nuclease, the incubation conditions are such that the nuclease would have cleaved at least a significant portion (*e.g.*, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99%) of templates containing its known target site. For known and generally well-characterized nucleases, such conditions are generally known in the art and/or can easily be discovered or optimized. For newly discovered nucleases, such conditions can generally be approximated using information about related nucleases that are better characterized (*e.g.*, homologs and orthologs).

[0161] In some embodiments, the nuclease is an endonuclease. In some embodiments, the nuclease is a site-specific endonuclease (*e.g.*, a restriction endonuclease, a meganuclease, a transcription activator-like effector nucleases (TALEN), a zinc finger nuclease, etc.).

[0162] In some embodiments, the site specificity of a site-specific nuclease is conferred by an accessory molecule. For example, the CRISPR-associated (Cas) nucleases are guided to specific sites by “guide RNAs” or gRNAs as described herein. In some embodiments, the nuclease is an RNA-guided nuclease. In some embodiments, the nuclease is a CRISPR-associated nuclease.

[0163] In some embodiments, the nuclease is a homolog or an ortholog of a previously known nuclease, for example, a newly discovered homolog or ortholog.

[0164] RNA-guided nucleases according to the present disclosure include, but are not limited to, naturally-occurring Class 2 CRISPR nucleases such as Cas9, and Cas12a, as well as other nucleases derived or obtained therefrom. In functional terms, RNA-guided nucleases are defined as those nucleases that: (a) interact with (*e.g.*, complex with) a gRNA; and (b)

together with the gRNA, associate with, and optionally cleave or modify, a target region of a DNA that includes (i) a sequence complementary to the targeting domain of the gRNA and, optionally, (ii) an additional sequence referred to as a “protospacer adjacent motif,” or “PAM,” which is described in greater detail below.

[0165] In one embodiment, the enzymatic cleavage agent comprises a CRISPR Cas/gRNA complex.

[0166] In one embodiment, a guide RNA homologous to a blocking oligonucleotide location can be used in conjunction with a Cas nuclease (*e.g.*, SpCas9) to cleave the targeted genomic DNA. This would selectively prevent amplification of native, non-rearranged fragments, while allowing rearranged loci (un-cleaved) to be amplified in exactly the same manner as in the blocking oligonucleotide procedure.

[0167] In one embodiment, the sequence specific cleavage reagent is a chemical cleavage agent. Chemical cleavage can encompass any method which utilizes a non-nucleic acid and non-enzymatic chemical reagent in order to promote/achieve cleavage of one or both strands of a double-stranded nucleic acid molecule. If required, one or both strands of the double-stranded nucleic acid molecule may include one or more non-nucleotide chemical moieties and/or non-natural nucleotides and/or non-natural backbone linkages in order to permit chemical cleavage reaction.

[0168] In one embodiment, the chemical cleavage agent can be linked to an oligonucleotide. Exemplary chemical cleavage agents linked to oligonucleotides that can be useful in the method herein include, without limitation, chelated metal ions (*e.g.*, Ag, Cu, Fe).

[0169] In one embodiment, the sequence tag comprises a sequence orthogonal to the genome, a unique molecular identifier (UMI), a transposase recognition site, an index sequence, or a combination thereof as described above for Method I.

[0170] In one embodiment, the first and second primers comprise second and third sequence tags as described above for Method I.

[0171] As described above for Method I, the first amplification round can be followed by a second amplification round. Thus, in some embodiments, the second amplification includes nested PCR and the introduction of second and third sequence tags. In some embodiments, the second amplification includes hemi-nested PCR.

[0172] In one embodiment, where a second amplification round is used, the method disclosed herein can further include prior to (d) performing a second amplification reaction using third

and fourth nested primers comprising nucleotide sequences complementary to sequences proximal to the inside ends of the amplified products of (c) and additional sequence tags at the 5' ends of the nested primers, to produce secondary amplification products comprising a re-arranged target sequence and additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fifth and sixth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

[0173] In one embodiment, where a second amplification round is used, the method disclosed herein can further include prior to (d) performing a second, hemi-nested amplification reaction using the second primer and a third nested primer wherein the third primer comprises a nucleotide sequence complementary to a sequence proximal to target-specific primer end of the amplified products of (c) and optionally a fourth and fifth sequence tags at the 5' ends of the second and/or third primers to produce secondary amplification products comprising a re-arranged target sequence and one or two additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fourth and fifth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

[0174] In one embodiment, the products of the first amplification reaction are subjected to a repeated cleavage reaction before proceeding with the second and/or third amplification reactions, respectively.

[0175] In one embodiment, the third and/or the fourth primers include barcode sequences as described above for Method I.

[0176] In one embodiment, the fifth and sixth primers include a sequencing tag and/or an index sequence as described above for Method I.

[0177] In one embodiment, the method further includes performing a separate amplification reaction to amplify the re-arranged tagged nucleic acid fragments using a second target specific primer on the opposite side of the target site.

[0178] Exemplary embodiments of the above method described herein is shown in FIG. 4 and FIG. 9. As shown in FIG. 4 and FIG. 9, instead of using blocking oligos as in Method I, Method II employs sequence specific cleavage using CRISPR/Cas, TALEN, ZFN, or similar

sequence directed enzymatic cleavage reagents. In some embodiments, cleavage reactions can be performed by mixing a guide RNA with a Cas9 nuclease at 15, 20, 25, 30, 35, 40, or 45°C for 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120 minutes.

[0179] Another alternative might be the use of a chemical cleavage reagent linked to an oligonucleotide. This would selectively prevent amplification of native, non-rearranged fragments, while allowing rearranged loci (un-cleaved) to be amplified in exactly the same manner as in the blocking oligo procedure. After the cleavage reaction is completed, all 3' ends are blocked by treatment with ddNTP and terminal deoxynucleotidyl transferase (TdT), Klenow exo-, or another suitable enzyme (TdT has the advantage that it can incorporate 3' blocked nucleotides at staggered or blunt ends). End-blocking will reduce the generation of interfering side products in the subsequent PCR reactions.

Method III

[0180] The present disclosure is also directed to a method for detecting genome-wide re-arrangements in a nucleic acid genome. The method includes (a) contacting a genomic DNA sample obtained from a cell or tissue contacted with a site specific nuclease with a plurality of transposons comprising a first sequence tag at the 5' end of the transposon, wherein the sequence tag comprises an RNA promoter sequence, under conditions whereby the plurality of transposons is inserted into the genomic DNA sample and the genomic DNA sample is tagged into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment; (b) transcribing the plurality of tagged nucleic acid fragments into RNA; (c) contacting the RNA with target sequence specific DNA oligonucleotide probes and RNase H; (d) performing a reverse transcription amplification reaction using a (i) a first primer comprising a nucleotide sequence complementary to a portion of the sequence tag located at the 3' or 5' end of the fragment and (ii) a second primer comprising a nucleotide sequence complementary to a region adjacent to a target site to produce primary amplification products; (e) sequencing the primary amplification products.

RNA Promoter Sequence

[0181] RNA promoter sequences useful in the method described herein are sequences capable of binding an RNA polymerase and contain a transcriptional start site. Accordingly, the promoter sequence usually includes between about 15 and about 250 nucleotides, preferably between about 25 and about 60 nucleotides, from a naturally occurring RNA polymerase

promoter, a consensus promoter sequence (Alberts et al., in *Molecular Biology of the Cell*, 2d Ed., Garland, N.Y. (1989), or a modified version thereof. Exemplary RNA promoters are the T3, T7, and SP6 phage promoter/polymerase systems.

[0182] Unlike previous methods described herein, this aspect of the disclosure includes the steps of (b) transcribing the plurality of tagged nucleic acid fragments into RNA; (c) contacting the RNA with target sequence specific DNA oligonucleotide probes and RNase H; (d) performing a reverse transcription amplification reaction. The earlier and later steps of the method are similar to those described for the other methods above.

[0183] In one embodiment, the sequence tag comprises a sequence orthogonal to the genome, a unique molecular identifier (UMI), a transposase recognition site, an index sequence, or a combination thereof as described above.

[0184] In one embodiment, the first and second primers comprise sequence tags as described above.

[0185] An exemplary embodiment of the above method described herein is shown in FIG. 5. As shown in FIG. 5, the special endoribonuclease activity of RNase H (Ribonuclease H), which specifically hydrolyzes the phosphodiester bonds of RNA when hybridized to DNA. In this method, T7 promoter (or other *in vitro* transcription promoter) is included in the transposon adapter. After tagmentation, DNA is *in vitro* transcribed into RNA. In some embodiments, the *in vitro* transcription reaction can be performed using MAXIscript™ T7 Transcription Kit and incubating at 25, 30, 35, 40, or 45°C for 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, or 600 minutes.

[0186] Then sequence specific DNA probes are hybridized to the RNA and treated with RNase H, which will specifically cleave the RNA region hybridized with the DNA probes. Then, one-step Reverse Transcription PCR is performed using sequence specific primer 2. In some embodiments, Primer 2 is designed specific to the right-side sequence of potential rearrangement site, and design the DNA probe specific to the left-side sequence of the same site. Thus, the non-rearranged templates will be cleaved by RNase H and cannot be amplified by primers 1 and 2, while the rearranged templates are not cleaved by the DNA probes and can be exponentially amplified. In some embodiments, for DNA probe hybridization and RNase H treatment, reactions are performed in two separate reactions for each target, one reaction for detecting genome rearrangement on the left-side of the target and the other reaction for detecting genome rearrangement on the right-side of the target. In some

embodiments, the reaction is denatured at about 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99°C for about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 minutes. In some embodiments, RNase H is added to the reaction and incubated at about 16, 20, 25, 30, 35, 40, or 45°C for about 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200 or 300 minutes. In some embodiments, RNase H is inactivated at 60, 65, 70, 75, 80, 85, or 90°C for about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, or 240 minutes. After first strand synthesis, the following steps are the same as in FIG. 4 from Step 3 onwards.

[0187] In one embodiment, the method further can include prior to (e) performing a second amplification reaction using third and fourth nested primers comprising nucleotide sequences complementary to sequences proximal to the inside ends of the amplified products of (d) and additional sequence tags, to produce secondary amplification products comprising a re-arranged target sequence and additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fifth and sixth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

[0188] In one embodiment, the method further can include prior to (e) performing a second, hemi-nested amplification reaction using the second primer and a third nested primer wherein the third primer comprises a nucleotide sequence complementary to a sequence proximal to target-specific primer end of the amplified products of (d) and optionally a fourth and fifth sequence tags at the 5' ends of the second and/or third primers to produce secondary amplification products comprising a re-arranged target sequence and one or two additional sequence tags. In one embodiment, the method further includes performing a third amplification reaction using a fourth and fifth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

[0189] In one embodiment, the third and/or fourth primers include barcode sequences as described above.

[0190] In one embodiment, the fifth and sixth primers comprise an adapter sequence and/or an index sequence as described above.

[0191] In one embodiment, the method further includes wherein (d) further comprises performing a separate amplification reaction to amplify the re-arranged tagged nucleic acid fragments using a second target specific primer on the opposite side of the target site.

DATA ANALYSIS

[0192] For the methods described herein, data analysis can be performed using paired-end sequencing reads from Illumina sequencing that are merged using a paired-end reads merging software, *e.g.*, PEAR (Zhang et al, Bioinformatics. 2014;30: 614–620), FLASH (Magoč and Salzberg, Bioinformatics. 2011;27: 2957–2963), BBMerge (Bushnell et al, PLoS One. 2017;12: e0185056), or NGmerge (Gaspar, Bioinformatics. 2018;19: 536). The merged reads can be trimmed and filtered using software such as BBmap (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>), samtools (Danecek et al, Gigascience. 2021;10. doi:10.1093/gigascience/giab008), or custom scripts to remove Illumina adaptor sequences, low quality reads, reads containing the target flanking sequence on the side of the target site where the blocking oligonucleotide or cleavage reagent binds, and reads that do not contain the target flanking sequence on the side of the target site where the target-specific primer hybridizes. Selected reads can be aligned to a human reference genome, *e.g.*, GRCh37 (hg19), GRCh38 (hg38), or Telomere-to-Telomere assembly (*e.g.*, T2T-CHM13v2.0), using an alignment software, *e.g.*, Bowtie 2 (Langmead and Salzberg, Nat Methods. 2012;9: 357–359), BWA (Li and Durbin, Bioinformatics. 2009;25: 1754–1760), or Minimap2 (Li, Bioinformatics. 2018;34: 3094–3100). The aligned BAM file can be converted into bed file using BEDTools (Quinlan, Bioinformatics. 2014;47: 11.12.1–34). For methods using UMIs, UMIs can be collapsed to remove redundant sequencing reads using software such as Gencore (Chen et al, Bioinformatics. 2019;20: 606), UMI-tools (Smith et al, Genome Res. 2017;27: 491–499), UMI-Reducer ([/github.com/smangul1/UMI-Reducer](https://github.com/smangul1/UMI-Reducer)), or custom scripts.

[0193] Reads with candidate translocation break points within a suitable window flanking the target site (*e.g.*, within 1, 3, 5, 10, 20, 40, 60, 80, 100, 200, or more bases) can be identified and counted to quantify the number of rearrangements between the target site and other genomic loci. Statistical tests can be applied to establish a confidence score to each group of rearrangements at each distal rearranged locus (*e.g.*, within a defined window of up 50, 100, 200, 500, 1000, 2000, 3000, or more bases at the distal rearranged locus). Visualization tools

such as IGV (Thorvaldsson et al, Brief Bioinform. 2013;14: 178–192) or Circos (Kryzyski et al, Genome Res. 2009;19: 1639–1645) can be used to facilitate the identification and analysis of rearrangements at genomic scale.

SYSTEMS AND KITS

[0194] Also provided herein are systems and kits including the reagents needed for performing the methods described herein as well as written instructions for making and using the same.

[0195] Any of the above-described systems and kits can further include one or more additional reagents.

[0196] In some embodiments, a system or kit can further include instructions for using the components of the kit to practice the methods. The instructions for practicing the methods are generally recorded on a suitable recording medium. For example, the instructions can be printed on a substrate, such as paper or plastic, and the like. The instructions can be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or sub-packaging), and the like. The instructions can be present as an electronic storage data file present on a suitable computer readable storage medium, *e.g.*, CD-ROM, diskette, flash drive, and the like. In some instances, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source (*e.g.*, via the internet), can be provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions can be recorded on a suitable substrate.

EXAMPLES

[0197] While particular alternatives of the present disclosure have been disclosed, it is to be understood that various modifications and combinations are possible and are contemplated within the true spirit and scope of the appended claims. There is no intention, therefore, of limitations to the exact abstract and disclosure herein presented.

EXAMPLE 1: EXEMPLARY METHOD FOR METHOD 1A

[0198] Exemplary oligonucleotides for use in the methods are listed in Table 1.

Table 1 – Exemplary oligonucleotide sequences.

Line_ID	Name	Sequence and modification	Note	Coordinate in hg38 (chr strand start end)	Method 1a Example 1 FIG. 1	Method 1b Example 3 FIG. 2	Method 1c Example 5 FIG. 3	Method 2 Example 7 FIG. 4	Method 3 Example 9 FIG. 5
Oligo-01	Tn5 adapter top	CGTACCGATACGCTCGGACAGATGTGTATAAG AGACAG (SEQ ID NO: 1)	Common		Yes	Yes	No	Yes	No
Oligo-02	Tn5 adapter bottom	CTGTCTCTTATACACATCT (SEQ ID NO: 2)	Common		Yes	Yes	Yes	No	Yes
Oligo-03	Blocker 0 - EMX1_L	+T+C+CTCCAG+CTTCT+GCCGT+T+T+G/3am MO/ (SEQ ID NO: 3)	target specific, for EMX1 Left side rearrangement	chr2 - 72933822 72933842	Yes	Yes	Yes	No	No
Oligo-04	Primer 1	GCGTACCGATACGCTCGGAC (SEQ ID NO: 4)	Common		Yes	Yes	Yes	Yes	Yes
Oligo-05	Primer 2 - EMX1_L	GGTTTGTGGTTGCCACCCT (SEQ ID NO: 5)	target specific, for EMX1 Left side rearrangement	chr2 - 72933953 72933972	Yes	Yes	Yes	Yes	Yes
Oligo-06	Primer 3	ACACTCTTTCCCTACACGAGCTCTTCCGATCT CGTACCGATACGCTCGGAC (SEQ ID NO: 6)	Common		Yes	Yes	Yes	Yes	Yes
Oligo-07	Primer 4 - EMX1_L	GTGACTGGAGTTACAGAGTGTGCTCTTCCGAT CTCTCTCCCATTTGGCTGCTTC (SEQ ID NO: 7)	target specific, for EMX1 Left side rearrangement	chr2 - 72933909 72933928	Yes	Yes	Yes	Yes	Yes
Oligo-08	Primer 5	AATGATACGGCGACACCGAGATCTACACTAT AGCCT ACACTCTTTCCCTACAGAGCGCTCTCC GATC*T (SEQ ID NO: 8)	common, bold is index, this is an example of NEBNext i501 Primer		Yes	Yes	Yes	Yes	Yes
Oligo-09	Primer 6	CAAGCAGAAGCGCATACGAGATCGAGTAA TGTGACTGGAGTTACAGAGTGTGCTCTCCGA TC*T (SEQ ID NO: 9)	common, bold is index, this is an example of NEBNext i701 Primer		Yes	Yes	Yes	Yes	Yes
Oligo-10	Tn5 adapter top U	GCGTACCGATACGCTCGGACUAGATGTGTATAA GAGACAG (SEQ ID NO: 10)	Common		No	No	Yes	No	No
Oligo-11	Tn5 adapter bottom	CTGTCTCTTATACACATCT/3ddC/ (SEQ ID NO: 11)	Common		No	No	No	Yes	No

[illegible]

Oligo- 23	Primer 2 - CCR5#1 R	ACAGGGTGAACAAGATGGAT (SEQ ID NO: 23)	target specific for CCR5#1 Right side rearrangement	chr3 + 46372888 46372908	Yes	Yes	Yes	Yes	Yes
Oligo- 24	Primer 4 - CCR5#1 R	GTGACTGGAGTTTCAGACGTGTGCTCTTCCGAT CTTACATCGGAGCCCTGCCAA (SEQ ID NO: 24)	target specific for CCR5#1 Right side rearrangement	chr3 + 46372947 46372966	Yes	Yes	Yes	Yes	Yes
Oligo- 25	Probe 0 - CCR5#1 R	GGCAACATGCTGGTCACTCCTCA (SEQ ID NO: 25)	target specific for CCR5#1 Right side rearrangement	chr3 + 46373041 46373062	No	No	No	No	Yes
Oligo- 26	Tn5 adapter top UMI	CGTACCGATACGCTCGGACNNNNNNNNNN NAGATGTGTATAAGAGACAG (SEQ ID NO: 26)	UMI version of Oligo- 01		Yes	No	Yes	No	No
Oligo- 27	Tn5 adapter top U UMI	CGTACCGAUACGCTCGGACUNNNNNNNNN NNUAGATGTGTATAAGAGACAG (SEQ ID NO: 27)	UMI version of Oligo- 10		No	Yes	No	No	No
Oligo- 28	Tn5 adapter top T7p UMI	TAATACGACTCACTATAGCGTACCGATACGCTC GGACNNNNNNNNNNNAGATGTGTATAAGA GACAG (SEQ ID NO: 28)	UMI version of Oligo- 12		No	No	No	No	Yes
For site specific oligos, the "R" or "L" in the oligo name is defined by which side the blocker/probe 0 is annealed to relative to the cleavage site, when the genomic DNA is kept its +/- strands orientation, despite the target sequence (PAM) is on (+) or (-) strand.									

[0199] Exemplary oligonucleotide sets are shown in Table 2 and Table 3 below. For example, for EMX1 or CCR5#1, a first reaction is performed to detect rearrangements on the left-side of the on-target cleavage site, and a second reaction is performed to detect rearrangements on the right-side of the on-target cleavage site. For each reaction, a set of target specific primers and blocker is to be used. For EMX1 left-side rearrangements, Oligos 03/05/07/13 are used, and for EMX1 right-side rearrangements, Oligos 14/15/16/17 are used. Similarly, for CCR5#1 left-side, Oligos 18/19/20/21 are used and for CCR5#1 right side, Oligos 22/23/24/25 are used.

Table 2. Target-specific primers and blocker

Common Name	for EMX1 Left side rearrangement EMX1_L		for EMX1 Right side rearrangement EMX1_R		for CCR5#1 Left side rearrangement CCR5_L		for CCR5#1 Right side rearrangement CCR5_R	
	Line_ID	Name	Line_ID	Name	Line_ID	Name	Line_ID	Name
Blocker 0	Oligo-03	Blocker 0 - EMX1_L	Oligo-14	Blocker 0 - EMX1_R	Oligo-18	Blocker 0 - CCR5#1_L	Oligo-22	Blocker 0 - CCR5#1_R
Primer 2	Oligo-05	Primer 2 - EMX1_L	Oligo-15	Primer 2 - EMX1_R	Oligo-19	Primer 2 - CCR5#1_L	Oligo-23	Primer 2 - CCR5#1_R
Primer 4	Oligo-07	Primer 4 - EMX1_L	Oligo-16	Primer 4 - EMX1_R	Oligo-20	Primer 4 - CCR5#1_L	Oligo-24	Primer 4 - CCR5#1_R
Probe 0	Oligo-13	Probe 0 - EMX1_L	Oligo-17	Probe 0 - EMX1_R	Oligo-21	Probe 0 - CCR5#1_L	Oligo-25	Probe 0 - CCR5#1_R

Table 3. Transposome oligonucleotides

Non-UMI version		UMI version	
Line_ID	Name	Line_ID	Name
Oligo-01	Tn5 transposome top	Oligo-26	Tn5 transposome top UMI
Oligo-10	Tn5 transposome top U	Oligo-27	Tn5 transposome top U UMI
Oligo-12	Tn5 transposome top T7p	Oligo-28	Tn5 transposome top T7p UMI

[0200] Exemplary sgRNA target sequences are shown in Table 4 below.

Table 4

Name	Sequence (PAM is <u>underlined</u>)	Coordinate in hg38 (chr strand start end)
EMX1 target sequence	GAGTCCGAGCAGAAGAAGAGG (SEQ ID NO: 29)	chr2 + 72933853 72933875
CCR5#1 target sequence	GTGAGTAGAGCGGAGGCAGGAGG (SEQ ID NO: 30)	chr3 - 46372995 46373017

[0201] Additional reagents include Tn5 transposase, *e.g.*, Robust Tn5 Transposase (Cat# EMQZ1422, Creative Biogene), DNA polymerase PCR master mix, *e.g.*, Platinum™ SuperFi II PCR Master Mix (ThermoFisher, Cat# 12368010), DNA polymerase with PCR buffer, *e.g.*, Platinum™ SuperFi II DNA Polymerase (ThermoFisher, Cat# 12361010), dNTP, *e.g.*, dNTP Mix (10 mM each) (ThermoFisher, Cat# R0194), DNA purification kit: Column kit, *e.g.*, DNA Clean & Concentrator-5 (Zymo Research, Cat# D4004) Magnetic beads, *e.g.*, AMPure XP Reagent (Beckman Coulter, Cat# A63880), Illumina sequencing kit, *e.g.*, NextSeq 1000/2000 P1 Reagents (300 Cycles) (Cat# 20050264), MiSeq Reagent Kit v3 (150-cycle) (Cat# MS-102-3001), MiSeq Reagent Kit v2 (300-cycles) (Cat# MS-102-2002), PhiX Control v3 (Illumina, Cat# FC-110-3001), STE buffer (10mM Tris-HCl (pH 8.0), 1mM EDTA, 0.1M NaCl).

[0202] In the first step, Tn5 transposomes are assembled. Transposon-end containing oligos (*e.g.*, Oligo-01 and Oligo-02 for non-UMI version, or Oligo-26 and Oligo-02 for UMI version) are annealed in STE buffer with the following program: 1–10 minutes at 95–99°C, 40–500 cycles of 1 minute at 1–99°C (decrease temperature 0.2–2°C every cycle until reaching 1–12°C). Transposome complexes are assembled in TPS buffer following the procedure below, as recommended by the manufacturer (Creative Biogene, Cat# EMQZ1422). The reagents are mixed thoroughly and incubated at 15–40°C for 5–120 minutes.

Component	Volume
Tn5 transposase	1–10 µL
Adaptor	0.5–8 µL
10x TPS buffer	2 µL
Sterile water	add to 20 µL

[0203] The workflow for this example is shown in FIG. 1. The tagmentation reaction is performed in which genomic DNA extracted from gene edited samples or controls (using standard methods) is tagmented with the assembled transposome complex following procedure. An exemplary reaction mixture is shown below:

Component	Volume
Genomic DNA	varies (5–500 ng)
5x LM buffer	6 µL
Transposome	0.1–8 µL
Sterile water	add to 30 µL

[0204] The reaction components are mixed thoroughly and incubated at 30–65°C for 5–120 minutes. The tagged DNA is purified using column-based purification.

[0205] Tagmented DNA from the previous step is used for two separate PCR reactions for each target, one reaction to detect genome rearrangement on the left-side of the target and the other reaction to detect genome rearrangement on the right-side of the target. Each PCR is then performed with a common primer annealed to the adapter region, a target specific blocker and a target specific primer. PCR is performed using 3-step PCR cycling protocol with tagmented DNA, Primer 1 (Oligo-04, common to all targets and rearrangement of both sides), Primer 2 (*e.g.*, Oligo-05 for EMX1 left-side rearrangement, Oligo-15 for EMX1 right-side rearrangement, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 20–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0206] Nested PCR is then performed with barcode tag. PCR is performed with PCR product from the previous step, Primer 3 (Oligo-06, common to all targets and rearrangement of both sides), Primer 4 (*e.g.*, Oligo-07 for EMX1 left-side rearrangement, Oligo-16 for EMX1 right-side rearrangement, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-24 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0207] Tag PCR is then performed with sequencing primers. PCR is performed with PCR product from the previous step, Primer 5 (*e.g.*, Oligo-08, an example of Illumina indexing

primer from New England Biolabs, NEB#E7603A), Primer 6 (*e.g.*, Oligo-09, an example of Illumina indexing primer from New England Biolabs, NEB#E7611A), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0208] Next generation sequencing is then performed using the PCR product from the previous step, and sequencing is performed using an Illumina platform (*e.g.*, Illumina MiSeq or NextSeq 2000) with optional ≥ 0 –95% PhiX (Illumina, Cat# FC-110-3001).

[0209] Paired-end sequencing reads from Illumina sequencing are merged using a paired-end reads merging software, *e.g.*, PEAR (Zhang et al, *Bioinformatics*. 2014;30: 614–620), FLASH (Magoč and Salzberg, *Bioinformatics*. 2011;27: 2957–2963), BBMerge (Bushnell et al, *PLoS One*. 2017;12: e0185056), or NGmerge (Gaspar, *Bioinformatics*. 2018;19: 536). The merged reads are then trimmed and filtered using software such as BBmap (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>), samtools (Danecek et al, *Gigascience*. 2021;10. doi:10.1093/gigascience/giab008), or custom scripts to remove Illumina adaptor sequences, low quality reads, reads containing the target flanking sequence on the side of the target site where the blocking oligonucleotide or cleavage reagent binds, and reads that do not contain the target flanking sequence on the side of the target site where the target-specific primer hybridizes. Selected reads are aligned to a human reference genome, *e.g.*, GRCh37 (hg19), GRCh38 (hg38), or Telomere-to-Telomere assembly (*e.g.*, T2T-CHM13v2.0), using an alignment software, *e.g.*, Bowtie 2 (Langmead and Salzberg, *Nat Methods*. 2012; 9: 357–359), BWA (Li and Durbin, *Bioinformatics*. 2009;25: 1754–1760), or Minimap2 (Li, *Bioinformatics*. 2018;34: 3094–3100). The aligned BAM file is converted into bed file using BEDTools (Quinlan, *Bioinformatics*. 2014; 47: 11.12.1–34). For methods using UMIs, UMIs are collapsed to remove redundant sequencing reads using software such as Gencore (Chen et al, *Bioinformatics*. 2019;20: 606), UMI-tools (Smith et al, *Genome Res*. 2017;27: 491–499), UMI-Reducer (github.com/smangul1/UMI-Reducer), or custom scripts.

Reads with candidate translocation break points within a suitable window flanking the target site (*e.g.*, within 1, 3, 5, 10, 20, 40, 60, 80, 100, 200, or more bases) are identified and counted to quantify the number of rearrangements between the target site and other genomic loci. Statistical tests are then applied to establish a confidence score to each group of rearrangements at each distal rearranged locus (*e.g.*, within a defined window of up 50, 100, 200, 500, 1000, 2000, 3000, or more bases at the distal rearranged locus). Visualization tools such as IGV (Thorvaldsdóttir et al, Brief Bioinform. 2013; 14: 178–192) or Circos (Krzywinski et al, Genome Res. 2009;19: 1639–1645) are used to facilitate the identification and analysis of rearrangements at genomic scale.

EXAMPLE 2: EXEMPLARY METHOD FOR METHOD 1A

[0210] Exemplary oligonucleotides used in the method are listed in Table 5.

Table 5 – Exemplary oligonucleotide sequences.

Line_ID	Name	Sequence and modification	Note	Coordinate in hg38 (chr strand start end)	Method 1A Example 2 FIG.6	Method 1B Example 4 FIG.7	Method 1C Example 6 FIG.8	Method 1I Example 8 FIG.9
Oligo-01A-501	i5_N501_UMI_Tn5 -A	AATGATACGGCGACACCGAGATCT ACACTAGATCGNNNNNNNNNTC GTCGGCAGCGTCAGATGTGTAAAG AGACAG (SEQ ID NO: 35)	Common		Yes	Yes	No	Yes
Oligo-01A-502	i5_N502_UMI_Tn5 -A	AATGATACGGCGACACCGAGATCT ACACCTCTCTATNNNNNNNNNTCG TCGGCAGCGTCAGATGTGTAAAG GACAG (SEQ ID NO: 36)	Common		Yes	Yes	No	Yes
Oligo-01A-503	i5_N503_UMI_Tn5 -A	AATGATACGGCGACACCGAGATCT ACACTATCTCTNNNNNNNNNTCG TCGGCAGCGTCAGATGTGTAAAG GACAG (SEQ ID NO: 37)	Common		Yes	Yes	No	Yes
Oligo-01A-504	i5_N504_UMI_Tn5 -A	AATGATACGGCGACACCGAGATCT ACACAGAGTAGANNNNNNNNTC GTCGGCAGCGTCAGATGTGTAAAG AGACAG (SEQ ID NO: 38)	Common		Yes	Yes	No	Yes
Oligo-01A-505	i5_N505_UMI_Tn5 -A	AATGATACGGCGACACCGAGATCT ACACGTAAGGAGNNNNNNNNNTC GTCGGCAGCGTCAGATGTGTAAAG AGACAG (SEQ ID NO: 39)	Common		Yes	Yes	No	Yes
Oligo-01A-506	i5_N506_UMI_Tn5 -A	AATGATACGGCGACACCGAGATCT ACACACTGCATNNNNNNNNNTC GTCGGCAGCGTCAGATGTGTAAAG AGACAG (SEQ ID NO: 40)	Common		Yes	Yes	No	Yes
Oligo-01A-507	i5_N507_UMI_Tn5 -A	AATGATACGGCGACACCGAGATCT ACACAAAGGAGTANNNNNNNNTC GTCGGCAGCGTCAGATGTGTAAAG AGACAG (SEQ ID NO: 41)	Common		Yes	Yes	No	Yes
Oligo-01A-508	i5_N508_UMI_Tn5 -A	AATGATACGGCGACACCGAGATCT ACACCTAAGCCTNNNNNNNNNTC GTCGGCAGCGTCAGATGTGTAAAG AGACAG (SEQ ID NO: 42)	Common		Yes	Yes	No	Yes

Oligo-02A	i5_N501_UMI_Tn5-A Uracil	AATGATACGGCGACCGAGATC/i deoxyU/ACACTAGATCGNNNNNNN/ideoxyU/CGTCGGCAGCGTCAGA TGTGTATAAGAGACAG (SEQ ID NO: 43)		No	NO	Yes	NO
Oligo-03A	Tn5-b_3' blocked	CTGTCTCTTATACACATCT/3ddC/ (SEQ ID NO: 44)	Common	Yes	Yes	Yes	Yes
Oligo-04A	Primer 1/3/5	AATGATACGGCGACCGAGATCTACAC (SEQ ID NO: 45)	Common	Yes	Yes	Yes	Yes
Oligo-05A	Blocker 0 - CCR5#1_L	+A+GG+GC+TC+CG+AT+GT+AT+AA +TA+AT+T+G/3AmMO/ (SEQ ID NO: 46)	target specific, for CCR5#1 Left side rearrangement	Yes	Yes	Yes	No
Oligo-09A	Blocker 0 - CCR5#1_R	+T+GG+GC+AA+CA+TG+CT+GG+TC +AT+CC+T+C/3AmMO/ (SEQ ID NO: 47)	target specific, for CCR5#1 Right side rearrangement	Yes	Yes	Yes	No
Oligo-13A-701	Primer 6_N701	CAAGCAGAAGACGGCATACGAGAT AGCGGAATGTGACTGGAGTTCAGA CGTGT (SEQ ID NO: 48)	common	Yes	Yes	Yes	Yes
Oligo-13A-702	Primer 6_N702	CAAGCAGAAGACGGCATACGAGAT GATCATGCGTGAATGGAGTTCAGAC GTGT (SEQ ID NO: 49)	common	Yes	Yes	Yes	Yes
Oligo-13A-703	Primer 6_N703	CAAGCAGAAGACGGCATACGAGAT AAGACGGAGTGAATGGAGTTCAGAC CGTGT (SEQ ID NO: 50)	common	Yes	Yes	Yes	Yes
Oligo-13A-704	Primer 6_N704	CAAGCAGAAGACGGCATACGAGAT CGAGTCCTGTGACTGGAGTTCAGAC GTGT (SEQ ID NO: 51)	common	Yes	Yes	Yes	Yes
Oligo-13A-705	Primer 6_N705	CAAGCAGAAGACGGCATACGAGAT TCCTCAGGTTGACTGGAGTTCAGAC GTGT (SEQ ID NO: 52)	common	Yes	Yes	Yes	Yes

Oligo-13A-706	Primer 6_N706	CAAGCAGAAGACGGGCATACGAGATGTACGGATGTGACTGGAGTTCAGACGTGT (SEQ ID NO: 53)	common		Yes	Yes	Yes	Yes	Yes
Oligo-13A-707	Primer 6_N707	CAAGCAGAAGACGGGCATACGAGATCATCTCGTGACTGGAGTTCAGACGTGT (SEQ ID NO: 54)	common		Yes	Yes	Yes	Yes	Yes
Oligo-13A-710	Primer 6_N710	CAAGCAGAAGACGGGCATACGAGATGTCGGAGCGTGACTGGAGTTCAGACGTGT (SEQ ID NO: 55)	common		Yes	Yes	Yes	Yes	Yes
Oligo-13A-711	Primer 6_N711	CAAGCAGAAGACGGGCATACGAGATACGGAGAAGTGACTGGAGTTCAGACGTGT (SEQ ID NO: 56)	common		Yes	Yes	Yes	Yes	Yes
Oligo-13A-712	Primer 6_N712	CAAGCAGAAGACGGGCATACGAGATAGGAGATGGTGACTGGAGTTCAGACGTGT (SEQ ID NO: 57)	common		Yes	Yes	Yes	Yes	Yes
Oligo-13A-714	Primer 6_N714	CAAGCAGAAGACGGGCATACGAGATAGTACTCGTGACTGGAGTTCAGACGTGT (SEQ ID NO: 58)	common		Yes	Yes	Yes	Yes	Yes
Oligo-13A-715	Primer 6_N715	CAAGCAGAAGACGGGCATACGAGATGGACTCTAGTGACTGGAGTTCAGACGTGT (SEQ ID NO: 59)	common		Yes	Yes	Yes	Yes	Yes
Oligo-60	CCRS#13	CTGATGTGTGGCAGTTTGGGAC (SEQ ID NO: 60)	Validation	Chr13 - 24886436 24886458	Yes	Yes	Yes	Yes	Yes
For site specific oligos, the "R" or "L" in the oligo name is defined by which side the blocker/probe 0 is annealed to relative to the cleavage site, when the genomic DNA is kept its +/- strands orientation, despite the target sequence (PAM) is on (+) or (-) strand.									

[0211] Exemplary oligonucleotide sets used are shown in Table 6 and Table 7 below. For example, for CCR5#1, a first reaction was performed to detect rearrangements on the left-side of the on-target cleavage site, and a second reaction was performed to detect rearrangements on the right-side of the on-target cleavage site. For each reaction, a set of target specific primers and blocker were used. For CCR5#1 left side, Oligos 04A/05A/19/20 were used and for CCR5#1 right side, Oligos 04A/09A/10A/23/24 were used.

Table 6. Target-specific primers and blocker

Common Name	for CCR5#1 Left side rearrangement CCR5_L		for CCR5#1 Right side rearrangement CCR5_R	
	Line_ID	Name	Line_ID	Name
Blocker 0	Oligo-05A	Blocker 0 - CCR5#1 L	Oligo-09A	Blocker 0 - CCR5#1 R
Primer 2	Oligo-19	Primer 2 - CCR5#1 L	Oligo-23	Primer 2 - CCR5#1 R
Primer 4	Oligo-20	Primer 4 - CCR5#1 L	Oligo-24	Primer 4 - CCR5#1 R

Table 7. Transposome oligonucleotides

UMI version	
Line_ID	Name
Oligo-01A-501	i5_N501_UMI_Tn5-A
Oligo-01A-502	i5_N502_UMI_Tn5-A
Oligo-03A	Tn5 transposome bottom

[0212] Exemplary sgRNA target sequences used are shown in Table 8 below.

Table 8

Name	Sequence (PAM is <u>underlined</u>)	Coordinate in hg38 (chr strand start end)
CCR5#1 target sequence	GTGAGTAGAGCGGAGGCAGG <u>AGG</u> (SEQ ID NO: 30)	chr3 - 46372995 46373017

[0213] Additional reagents included unloadedTn5 Transposase (Diagenode, C01070010, Diagenode), 2X Tagmentation buffer (Diagenode, C01019043), Platinum™ SuperFi II PCR Master Mix (ThermoFisher, Cat# 12368010), dNTP Mix (10 mM each) (ThermoFisher, Cat# R0194), DNA purification kit: DNA Clean & Concentrator-5 (Zymo Research, Cat# D4004), AMPure XP Reagent (Beckman Coulter, Cat# A63880), Glycerol (MilliporeSigma, 356350500ML), 10% SDS (Fisher, BP2436200), 5M NaCl (Invitrogen, AM9760G), Tris HCl

1M, pH8.0 (Invitrogen, 15568-025), Genomic DNA (extracted from edited and unedited 293T cells), NEBNext library Quant Kit for Illumina (NEB, E7630), Illumina NextSeq 1000/2000 P1 Reagents (300 Cycles) (Illumina, Cat# 20050264), PhiX Control v3 (Illumina, Cat# FC-110-3001).

[0214] Before library preparation, we tested the ability of the LNA blocker to efficiently block the amplification of unintended targets, including wild type and edits with small indels. We included LNA blocker (Oligo05-A and Oligo09-A) in the PCR to amplify the amplicon using wild type genomic DNA. The PCR mixture was prepared as follows: 50ng gDNA, 25 μ L 2x PCR Master Mix, 2.5 μ L Primer F (Oligo-19, 10 μ M), 2.5 μ L Primer R (Oligo-23, 10 μ M), different concentration of Blocker, and add ddH₂O up to 50 μ L. PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 35 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 80°C, 10 second at 60°C, seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were directly profiled in 2% EX agarose gel. As shown in FIG. 10A, when the concentration of Blockers reaches 50% of the PCR primers, both of which can efficiently block the amplification of the amplicon. To further optimize the blocking effect of the Blocker oligo, different temperatures were tested as shown in FIG. 10B. The temperatures tested, from 77°C to 82.4°C, all achieved complete elimination of PCR products. 78°C was used for 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, seconds at 72°C) for library preparation.

[0215] In the first step, transposon-end containing oligos (*e.g.*, Oligo-01A-501 and Oligo-03A, or Oligo-01A-502 and Oligo-03A) were resuspend in Annealing Buffer (40mM Tris-HCl (pH8.0), 50mM NaCl) to stock concentration of 100 μ M. In a PCR tube, 10 μ L oligo-01, 10 μ L oligo-03 were mixed, vortexed and placed in PCR tubes in a thermocycler with the following program: 5 minutes at 95°C, cool to 65°C (-0.1°C/second), 5 minutes at 65°C, cool to 25°C (-0.1°C/second), 5 minutes at 25°C, and hold at 4°C.

[0216] Transposome complexes were assembled following the procedure below, as recommended by the manufacturer. The reagents were mixed in a PCR tube: 10 μ L Tn5 transposase (2 ug/ μ L), 10 μ L annealed adaptor. The reagents were mixed thoroughly and incubated at 23°C for 30 minutes. 10 μ L glycerol was added and mixed. The assembled transposome complex was stored at -20°C.

[0217] The workflow for this example is shown in FIG. 6. The tagmentation reaction was performed in which genomic DNA extracted from gene edited samples or controls (using standard methods) was tagmented with the assembled transposome complex using the following procedure. The reaction mixture contained: 100 ng genomic DNA from unedited control or edited sample, 20 µL 2x Tagmentation buffer, 200 ng loaded Tn5 Transposase, and add H₂O up to 40 µL. The reaction mixture was mixed thoroughly and incubated at 55°C for 15 minutes. 10 µL 0.2% SDS was added. And then Tn5 was inactivated for 10 min at 70°C. The tagmented DNA was purified using a Zymo column following the manufacturer's instructions and eluted in 21 µL.

[0218] Tagmented DNA from the previous step was used for two separate PCR reactions for each target, one reaction to detect genome rearrangements on the left-side of the target and the other reaction to detect genome rearrangements on the right-side of the target. Each PCR was then performed with a common primer annealed to the Tn5 adapter region, a target specific blocker and a target specific primer. Primer 1 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 2 (*e.g.*, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixture was prepared as follows: 25 µL 2x PCR Master Mix, 10 µL Tagmented DNA, and 1.25 µL Primer 1 (10 µM), 2.5 µL Primer 2 (10 µM), 10 µL Blocker 0 (10 µM), and add H₂O up to 50 µL. PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, 60 seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads (1x) and eluted in 12 µL.

[0219] Nested PCR was then performed with the PCR product from the previous step, Primer 3 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 4 (*e.g.*, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-24 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixture was prepared as follows: 25 µL 2x PCR Master Mix, 10 µL PCR product, 1.25 µL Primer 3 (10 µM), 2.5 µL Primer 4 (10 µM), 10 µL Blocker 0 (10 µM), and add H₂O up to 50 µL. PCR was performed using the following program: initial denaturing at

98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads, and eluted in 12 µL.

[0220] Tag PCR was then performed with sequencing primers. PCR was performed with PCR products from the previous step, Primer 5 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 6 (*e.g.*, any Oligo-13A, or Illumina indexing primer from New England Biolabs, NEB#E7611A), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixture was prepared as follows: 25 µL 2x PCR Master Mix, 10 µL PCR product, 5 µL Primer 5 (10 µM), 5 µL Primer 6 (10 µM), 5 µL Blocker 0 (20 µM). PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads, and eluted in 52 µL.

[0221] 50 µL of the PCR product was transferred to a new tube for library quantification using NEBNext library Quant Kit for Illumina (NEB, E7630) following the manufacture's instructions. Libraries were normalized and pooled for sequencing.

[0222] Next generation sequencing was then performed using the PCR product from the previous step, and sequencing was performed using an Illumina NextSeq 2000 with 20%PhiX (Illumina, Cat# FC-110-3001).

[0223] Paired-end sequencing reads from Illumina sequencing were processed using a customized NGS analysis pipeline. The sequence reads were first checked for overall quality using FastQC followed by consolidating FastQC results using MultiQC. The reads were then subjected to Adapter and primer sequence trimming using Cutadapt (DOI:10.14806/ej.17.1.200) followed by trimming low quality bases using Trimmomatic (Bolger et al, Bioinformatics. 2014; 30(15):2114-20). The trimming in the reads was verified by running FastQC and MultiQC steps on trimmed reads fastq files to ascertain the removal of adaptors. The resulting reads were aligned with human genome reference sequence (hg38) using BWA-MEM (Li and Durbin, Bioinformatics, 2009; 25:1754-1760) tool with default parameters. The generated Sequence Alignment Map (SAM) file was sorted by coordinate

and converted to Binary Alignment Map (BAM) file using PicardTools (<http://broadinstitute.github.io/picard>). The BAM file was then filtered to remove low-quality reads and to keep the reads with quality score ≥ 30 using Samtools. The resulting high quality BAM file was then checked using the CollectInsertSizeMatrix and CollectAlignmentSummaryMetrics modules from PicardTools followed by FastQC and MultiQC report generation. The reads in the sorted high quality read BAM were grouped based on UMIs using UMI-Tools with paired option and saved as grouped BAM followed by BAM indexing using Samtools. The grouped BAM file was then de-duplicated using UMI-Tools with paired option to remove PCR duplicate reads. The FGSV tool was then used to discover the structural variation pileup by searching for split read mapping and read pairs that map across breakpoints in the BAM file using the FGSV SVPileup module. The AggregateSVPileup module of FGSV then aggregated information across nearby pileups to call structural variants. Only the pileups containing at least one breakpoint on the target were considered as real hits, and as high confidence hits with at least 10 split reads. IGV (Thorvaldsdóttir et al, Brief Bioinform. 2013; 14: 178–192) was used to facilitate the identification and analysis of rearrangements at genomic scale.

[0224] To validate the DNA rearrangements identified by experiment above, PCR was performed with genomic DNA after gene editing. Primer for inter-chromosomal rearrangement event CCR5#1C13 (Oligo-60, common to rearrangement of both sides), Primer for CCR5-side (*e.g.*, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). The PCR mixture was prepared as follows: 100ng gDNA, 25 μ L 2x PCR Master Mix, 5 μ L Primer CCR5#1C13 (10 μ M), 5 μ L Primer CCR5-side (10 μ M), 5 μ L Blocker 0 (20 μ M). PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 35 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, 60 seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were directly profiled in 2% EX agarose gel. Since the genomic DNA for all the experiments performed in this application is from the same batch, therefore, the PCR method is the same for all methods and the results are applicable to all method as well.

[0225] Summary of DNA rearrangement events detected for Example 2 in control and edited cells in Table 9 below. To remove the noise introduced by PCR in library preparation for next generation sequencing, we set a threshold at 10 split reads for high confident hits and one breakpoint must be located at the on-target site. In control cells, 0 hits passed the filters for high confidence for the left-side events, 18 hits were enriched for the right-side events in Table 9. However, the number of the split reads that supported these hits was very low compared to events in edited cells in FIGs. 11A-B. In edited samples, in total, 63 and 68 hits were captured for the left side and right-side events, respectively. These data support that the hits captured in edited cells are produced or increased by CRISPR/Cas9 gene editing.

Table 9. Summary of DNA rearrangement events detected for Example 2

Samples	Side	Intra Chromosomal	Inter Chromosomal
Control	Left	0	0
CCR5 edited		48	15
Control	Right	14	4
CCR5 edited		50	18
Note: DNA rearrangement events with split read number <10 were filtered out. One breakpoint must be located at the on-target cutting site +/-25 bp.			

[0226] DNA rearrangements identified by Example 2 in CCR5 edited cells are displayed in Table 10 below. The genomic coordinates of the DNA breakpoints between DNA rearrangements, number of unique split-reads based on UMI-tools, extracted reads to show the split reads, and note to further explain the potential mechanism for this DNA rearrangements were included in the table. It is worth noting that the DNA rearrangements between CCR5 and its homologous gene CCR2, inter-chromosomal translocation between CCR5 on chromosome 3 and RNF17/CENPJ on chromosome 13 were also identified by CAST-seq (Turchiano et al, 2021). The presence of fused DNA of chr3/chr13 and acentric chromosome were also validated using PCR (FIGs. 12A-B). Therefore, SAFER detection Method IA can efficiently capture intra- and inter-chromosomal rearrangements.

Table 10. Examples of split reads for DNA rearrangement events.

Catgory	Side	On-target breakpoint	Non-target breakpoint	Unique Split_reads	Exemplar read	Note
---------	------	----------------------	-----------------------	--------------------	---------------	------

Intra-chromosomal rearrangement	Left	chr3:46373018-46373029	chr3:46357590-46357600	559	Read_name: VH01033:30:AAF5TJWM5:1:1202:15618:35494:rCTACCGATCA Sequence: GCTCTTCAGCC TTTTGCAGTTTA TCAGGATGAGG ATGACCAGCAT GTTGCCACAA AACCAAAGATG AACACCTTTTTT <u>GATTATGATTA</u> <u>CGGTGCTCCCT</u> <u>GTCATAAATT</u> <u>GACGTGAAGCA</u> <u>AATTGGGGCCC</u> <u>AATCCTGCCT</u> <u>CCGCTC</u> (SEQ ID NO: 64) Alignments: chr3:46,373,018-46,373,091 (+) <u>chr3:46,357,593-46,357,671 (-)</u> <u>underlined</u>	This event represents an intergenic inversion that is caused by off-target editing or homology mediated recombination. Predicted off-target site: CACCAGCGAGTAG AGCGGAGGCAGGA G (SEQ ID NO: 61) chr3:46357657-46357683 and homology sequence in CCR2 and CCR5: CTGGTGTTTCATCTT TGGTTTTGTGGCA ACATGCTGGTC (SEQ ID NO: 62) Gene: CCR2, CCR5
	Left	chr3:46373000-46373015	chr3:46372658-46372689	2643	Read_name: VH01033:30:AAF5TJWM5:1:2404:56614:2950:rTCTACCGGAA Sequence: <u>GTGCTATAGAG</u> <u>CACAAGATTTT</u> <u>ATTGGTGAGA</u> <u>TGGTGCTCTAC</u> TCACTGGTGTT CATCTTTGGTTT TGTGGCAACA TGCTGGTCATC CTCATCCTGAT AAAC TGCAAAA GGCTGAAGAGC (SEQ ID NO: 65) Alignments: <u>chr3:46,372,619-46,372,658 (+)</u> <u>underlined</u> chr3:46,373,007-46,373,092 (+)	The event represents a large deletion caused by CRISPR/Cas9 editing. No off-target site was predicted at the non-target breakpoint site. Gene: CCR5

	Left	chr3:46373000-46373015	chr3:46374369-46374387	1284	Read_name: VH01033:30:AAF5TJWM5:1:1102:51274:48462:rTTA CCTATCA Sequence: <u>CCCCATAGCAA</u> <u>GACAAAGACCT</u> <u>GTTTTAAACAC</u> <u>ATTTACAACCT</u> <u>ATATGTCCTCG</u> CTCTACTCACT GGTGTTTCATCTT TGGTTTTGTGG GCAACATGCTG GTCATCCTCAT CCTGATAAACT GCAAAAGGCTG AAGAGC (SEQ ID NO: 66) Alignments: chr3:46,373,002-46,373,091 (+) <u>chr3:46,374,369-46,374,419 (-)</u> <u>underlined</u>	The event represents a large deletion caused by CRISPR/Cas9 editing. This intragenic Inversion may be occurred between on-target CRISPR editing and naturally happened breakpoint. No off-target site was predicted at the non-target breakpoint site. Gene: CCR5
	Right	chr3:46373000-46373000	chr3:46357607-46357607	28	Read_name: VH01033:30:AAF5TJWM5:1:1404:8328:41230:rCAGC ACCATG Sequence: TACATCGGAGC CCTGCCAAAAA ATCAATGTGAA GCAAATCGCAG CCCGCCTCCTC <u>ATAATCAAAAA</u> <u>AGGTGGTGACT</u> <u>TCTTCACGCTC</u> <u>TCGTGGTATTI</u> <u>CTGGTGTTCGTT</u> (SEQ ID NO: 67) Alignments: chr3:46,372,947-46,373,000 (+) <u>chr3:46,357,556-46,357,606 (-)</u> <u>underlined</u>	This event represents an intergenic inversion that is caused by off-target editing or homology mediated recombination. Predicted off-target site: CACCAGCGAGTAG AGCGGAGGCAGGA G (SEQ ID NO: 61) chr3:46357657-46357683 and homology sequence in CCR2 and CCR5: CTGGTGTTCATCTT TGGTTTTGTGGGCA ACATGCTGGTC (SEQ ID NO: 62) Gene: CCR2, CCR5
	Right	chr3:46372992-46373001	chr3:46357616-46357619	391	Read_name: VH01033:30:AAF5TJWM5:1:1301:13725:35115:rGTC GACCCAA Sequence: TACATCGGAGC CCTGCCAAAAA ATCAATGTGAA	This event represents an intergenic inversion that is caused by off-target editing or homology mediated recombination. Predicted off-target site: CACCAGCGAGTAG

					GCAAATCGCAG <u>CCCGCTCCCT</u> <u>GTCATAAATT</u> <u>GACGTGAAGCA</u> <u>AATTGGGGCCC</u> <u>AACTCCTGCC</u> (SEQ ID NO: 68) Alignments: chr3:46,372,947- 46,372,999 (+) <u>chr3:46,357,616-</u> <u>46,357,664 (+)</u> <u>underlined</u>	AGCGGAGGCAGGA G (SEQ ID NO: 61) chr3:46357657- 46357683 and homology sequence in CCR2 and CCR5: <u>CTGGTGTTTCATCTT</u> <u>TGGTTTTGTGGGCA</u> <u>ACATGCTGGTC</u> (SEQ ID NO: 62) Gene: CCR2, CCR5
	Right	chr3:46372986-46373002	chr3:46373246-46373288	4074	Read_name: VH01033:30:AAF 5TJWM5:1:1303:3 1164:20443:rAGA TCCGGTT Sequence: TACATCGGAGC CCTGCCAAAAT CAATGTGAAGC AAATCGCAGCC CGCCTCCTGAC <u>AATCGATAGGT</u> <u>ACCTGGCTGTC</u> <u>GTCCATGCTGT</u> <u>GTTTGCTTAA</u> <u>AAGCCAGGACG</u> <u>GTCACCTTTGG</u> <u>GGTGGTGACAA</u> <u>GTGTGATCACT</u> <u>TGGGTGG</u> (SEQ ID NO: 69) Alignments: chr3:46,372,946- 46,372,995 (+) <u>chr3:46,373,262-</u> <u>46,373,365 (+)</u> <u>underlined</u>	The event represents a large deletion caused by CRISPR/Cas9 editing. No off-target site was predicted at the non- target breakpoint site. Gene: CCR5
	Right	chr3:46372984-46373001	chr3:46368151-46368163	1686	Read_name: VH01033:30:AAF 5TJWM5:1:2104:5 1122:6774:rCCGG GTGTCT Sequence: <u>CTCCCCATCCT</u> <u>GTTGACGATGC</u> <u>TCTGAAAATAT</u> <u>GGTCCAGAAAT</u> <u>GGTGTGAGCAA</u> <u>GGAGACAGCAA</u> <u>AGCAATGCTTG</u> <u>GAACATAGGTG</u> <u>CAGTGACTAGA</u> <u>CATGAGGCGGG</u> <u>CTGCGATATGA</u> <u>TTCACATTGATT</u>	The event represents a large deletion caused by CRISPR/Cas9 editing. This intragenic Inversion may be occurred between on- target CRISPR editing and naturally happened breakpoint. No off-target site was predicted at the non- target breakpoint site. Gene: CCR5

					TTTTGGCAGGG CTCCGAT (SEQ ID NO: 70) Alignments: <u>chr3:46,368,056-</u> <u>46,368,162 (+)</u> chr3:46,372,946- 46,372,997 (-)	
Inter chromosomal rearrangemen t	Left	chr3:46373002 -46373008	chr13:2488633 7-24886341	279	Read_name: VH01033:21:AAC CM73M5:1:2403:5 0308:18720:rACT AGCCAAG Sequence: GATGAGGATGA CCAGCATGTTG CCCACAAAACC AAAGATGAACA CCAGTGAGTAG AGCGGGTGGTC AGCCAACAAGC <u>TGCTGGGCGTG</u> <u>TGAGGCGGCTG</u> <u>TGCTGTGCCTG</u> <u>TGAGACAGGAA</u> <u>TGGTTCCATTA</u> <u>CAC (SEQ ID NO:</u> <u>71)</u> Alignments: chr3:46373004- 46373063 (-) <u>chr13:248863338-</u> <u>248863413 (+)</u> <u>underlined</u>	These events represents an interchromosomal translocation that is mediated by Off-target editing. On-target site: chr3:46372995- 46373017 (-) GTGAGTAGAGCGG AGGCAGGAGG (SEQ ID NO: 30) OFF-target site: G()GAGGAGAGCGG AGGCAGGTGG Chr13:24886320- 24886342 (+) Gene: CCR5 and RNF17/CENPJ
	Right	chr3:46372994 -46373001	chr13:2488633 5-24886341	464	Read_name: VH01033:30:AAF 5TJWM5:1:2402:2 1375:40681:rATC CCGGCTC Sequence: TACATCGGAGC CCTGCCAAAAA ATCAATGTGAA GCAAATCGCAG CCCGCAGGTGG TCAGCCAACAA <u>GCTGCTGGGCG</u> <u>TGTGAGGCGGC</u> <u>TGTGCTGTGCC</u> <u>TGTGAGACAGG</u> <u>AATGGTTCCAT</u> <u>TACAC (SEQ ID</u> <u>NO: 72)</u> Alignments: chr3:46372947-	

					46372995 (+) <u>chr13:248863336-</u> <u>248863412 (+)</u> <u>underlined</u>	
--	--	--	--	--	--	--

Example 3: Exemplary Method for Method IB

[0227] Exemplary oligonucleotides are listed in Table 1.

[0228] Additional reagents include Tn5 transposase, *e.g.*, Robust Tn5 Transposase (Cat# EMQZ1422, Creative Biogene), DNA polymerase PCR master mix, *e.g.*, Platinum™ SuperFi II PCR Master Mix (ThermoFisher, Cat# 12368010), DNA polymerase with PCR buffer, *e.g.*, Platinum™ SuperFi II DNA Polymerase (ThermoFisher, Cat# 12361010), dNTP, *e.g.*, dNTP Mix (10 mM each) (ThermoFisher, Cat# R0194), ddNTP, *e.g.*, Dideoxynucleoside Triphosphate Set (MilliporeSigma, Cat# 03732738001), DNA purification kit: Column kit, *e.g.*, DNA Clean & Concentrator-5 (Zymo Research, Cat# D4004), Magnetic beads, *e.g.*, AMPure XP Reagent (Beckman, Cat# A63880), Illumina sequencing kit, *e.g.*, NextSeq 1000/2000 P1 Reagents (300 Cycles) (Cat# 20050264), MiSeq Reagent Kit v3 (150-cycle) (Cat# MS-102-3001), MiSeq Reagent Kit v2 (300-cycles) (Cat# MS-102-2002), PhiX Control v3 (Illumina, Cat# FC-110-3001), STE buffer (10mM Tris-HCl (pH 8.0), 1mM EDTA, 0.1M NaCl).

[0229] In the first step, Tn5 transposomes are assembled. Transposon-end containing oligos (*e.g.*, Oligo-01 and Oligo-02 for non-UMI version, or Oligo-26 and Oligo-02 for UMI version) are annealed in STE buffer with the following program: 1–10 minutes at 90–99°C, 40–500 cycles of 1 minutes at 90–99°C (decrease temperature 0.2–2°C every cycle until reaching 1–12°C). Transposome complexes are assembled in TPS buffer following the procedure below, as recommended by the manufacturer (Creative Biogene, Cat# EMQZ1422). The reagents are mixed thoroughly and incubated at 15–40°C for 5–120 minutes.

Component	Volume
Tn5 transposase	1–10 µL
Adaptor	0.5–8 µL
10x TPS buffer	2 µL
Sterile water	add to 20 µL

[0230] The workflow for this example is shown in FIG. 2. The tagmentation reaction is performed in which genomic DNA extracted from gene edited samples or controls (using standard methods) is tagmented with the assembled transposome complex following procedure:

Component	Volume
Genomic DNA	varies (5–500 ng)
5x LM buffer	6 μ L
Transposome	0.1–8 μ L
Sterile water	add to 30 μ L

The reaction components are mixed thoroughly and incubated at 30–65°C for 5–120 minutes. Tagmented DNA is purified using column-based purification.

[0231] The 3' ends of DNA molecules are then blocked. 3' ends of DNA molecules are blocked using ddNTP and DNA polymerase SuperFi II (ThermoFisher, Cat# 12361010). Product is purified using column-based purification.

[0232] Tagmented DNA from the previous step is used for two separate PCR reactions for each target, one reaction to detect genome rearrangement on the left-side of the target and the other reaction to detect genome rearrangement on the right-side of the target. Each PCR is then performed with a common primer annealed to the adapter region, a target specific blocker and a target specific primer. PCR is performed using 3-step PCR cycling protocol with tagmented DNA, Primer 1 (Oligo-04, common to all targets and rearrangement of both sides), Primer 2 (*e.g.*, Oligo-05 for EMX1 left-side rearrangement, Oligo-15 for EMX1 right-side rearrangement, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 20–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0233] Nested PCR is then performed with the barcode tag. PCR is performed with PCR product from the previous step, Primer 3 (Oligo-06, common to all targets and rearrangement of both sides), Primer 4 (*e.g.*, Oligo-07 for EMX1 left-side rearrangement, Oligo-16 for EMX1 right-

side rearrangement, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-24 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0234] Tag PCR is then performed with sequencing primers. PCR is performed with PCR product from the previous step, Primer 5 (*e.g.*, Oligo-08, an example of Illumina indexing primer from New England Bio Labs, NEB#E7603A), Primer 6 (*e.g.*, Oligo-09, an example of Illumina indexing primer from New England Bio Labs, NEB#E7611A), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0235] Next generation sequencing is then performed using the PCR product from the previous step, and sequencing is performed using an Illumina platform (*e.g.*, Illumina MiSeq or NextSeq 2000) with optional ≥ 0 –95% PhiX (Illumina, Cat# FC-110-3001).

[0236] Paired-end sequencing reads from Illumina sequencing are merged using a paired-end reads merging software, *e.g.*, PEAR (Zhang et al, Bioinformatics. 2014;30: 614–620), FLASH (Magoč and Salzberg, Bioinformatics. 2011;27: 2957–2963), BBMerge (Bushnell et al, PLoS One. 2017;12: e0185056), or NGMerge (Gaspar, Bioinformatics. 2018;19: 536). The merged reads are then trimmed and filtered using software such as BBmap (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>), samtools (Danecek et al, Gigascience. 2021;10. doi:10.1093/gigascience/giab008), or custom scripts to remove Illumina adaptor sequences, low quality reads, reads containing the target flanking sequence

on the side of the target site where the blocking oligonucleotide or cleavage reagent binds, and reads that do not contain the target flanking sequence on the side of the target site where the target-specific primer hybridizes. Selected reads are aligned to a human reference genome, *e.g.*, GRCh37 (hg19), GRCh38 (hg38), or Telomere-to-Telomere assembly (*e.g.*, T2T-CHM13v2.0), using an alignment software, *e.g.*, Bowtie 2 (Langmead and Salzberg, *Nat Methods*. 2012; 9: 357–359), BWA (Li and Durbin, *Bioinformatics*. 2009; 25: 1754–1760), or Minimap2 (Li, *Bioinformatics*. 2018; 34: 3094–3100). The aligned BAM file is converted into bed file using BEDTools (Quinlan, *Bioinformatics*. 2014; 47: 11.12.1–34). For methods using UMIs, UMIs are collapsed to remove redundant sequencing reads using software such as Gencore (Chen et al, *Bioinformatics*. 2019;20: 606), UMI-tools (Smith et al, *Genome Res*. 2017;27: 491–499), UMI-Reducer (github.com/smangul1/UMI-Reducer), or custom scripts.

[0237] Reads with candidate translocation break points within a suitable window flanking the target site (*e.g.*, within 1, 3, 5, 10, 20, 40, 60, 80, 100, 200, or more bases) are identified and counted to quantify the number of rearrangements between the target site and other genomic loci. Statistical tests are then applied to establish a confidence score to each group of rearrangements at each distal rearranged locus (*e.g.*, within a defined window of up 50, 100, 200, 500, 1000, 2000, 3000, or more bases at the distal rearranged locus). Visualization tools such as IGV (Thorvaldsdóttir et al, *Brief Bioinform*. 2013; 14: 178–192) or Circos (Krzywinski et al, *Genome Res*. 2009; 19: 1639–1645) are used to facilitate the identification and analysis of rearrangements at genomic scale.

EXAMPLE 4: EXEMPLARY METHOD FOR METHOD 1B

[0238] Exemplary oligonucleotides used are listed in Table 5.

[0239] Additional reagents included unloaded Tn5 Transposase (Diagenode, C01070010, Diagenode), 2X Tagmentation buffer (Diagenode, C01019043), Platinum™ SuperFi II PCR Master Mix (ThermoFisher, Cat# 12368010), dNTP Mix (10 mM each) (ThermoFisher, Cat# R0194), DNA purification kit: DNA Clean & Concentrator-5 (Zymo Research, Cat# D4004), AMPure XP Reagent (Beckman Coulter, Cat# A63880), Glycerol (MilliporeSigma, 356350500ML), 10% SDS (Fisher, BP2436200), 5M NaCl (Invitrogen, AM9760G), Tris HCl 1M, pH8.0 (Invitrogen, 15568-025), Genomic DNA (extracted from edited and unedited 293T cells), ddNTP, Dideoxynucleoside Triphosphate Set (MilliporeSigma, Cat# 03732738001), NEBNext library Quant Kit for Illumina (NEB, E7630), Terminal transferase(TdT, NEB, M0315S), USER, (NEB #M5508), Illumina NextSeq 1000/2000 P1

Reagents (300 Cycles) (Illumina, Cat# 20050264), PhiX Control v3 (Illumina, Cat# FC-110-3001).

[0240] In the first step, transposon-end containing oligos (*e.g.*, Oligo-01A-501 and Oligo-03A, or Oligo-01A-502 and Oligo-03A) were resuspend in annealing buffer (40mM Tris-HCl (pH8.0), 50mM NaCl) to stock concentration of 100 μ M. In a PCR tube, 10 μ l oligo-01A, 10 μ l oligo-03A were mixed, vortexed and placed in PCR tubes in a thermocycler with the following program: 5 minutes at 95°C, cool to 65°C (-0.1°C/second), 5 minutes at 65°C, cool to 25°C (-0.1°C/second), 5 minutes at 25°C, and hold at 4°C.

[0241] Transposome complexes were assembled following the procedure below, as recommended by the manufacturer. The reagents were mixed in a PCR tube: 10 μ L Tn5 transposase (2 ug/ μ l), 10 μ L annealed adaptor. The reagents were mixed thoroughly and incubated at 23°C for 30 minutes. 10 μ l glycerol was added and mixed. The assembled transposome complex was stored at -20°C.

[0242] The workflow for this example is shown in FIG. 7. The tagmentation reaction was performed using genomic DNA extracted from gene edited samples or controls (using standard methods) with the assembled transposome complex using the following procedure: 100 ng genomic DNA from unedited control or edited sample, 20 μ L 2x Tagmentation buffer, 200 ng loaded Tn5 Transposase, and added H₂O up to 40 μ L. The reaction mixtures were mixed thoroughly and incubated at 55°C for 15 minutes. 10 μ L 0.2% SDS was added. And then Tn5 was inactivated for 10 min at 70°C. The tagmented DNA was purified using a Zymo column following the manufacturer's instructions and eluted in 11 μ L.

[0243] The 3' ends of DNA molecules were then blocked using ddNTP and terminal transferase (TdT, NEB, M0315S) using the following procedure: 10 μ L tagmented DNA, 5 μ L 10X TdT buffer, 2 μ L (40U) TdT, 5 μ L CoCl₂ solution, 1 μ L ddNTP (10mM), and added up to 50 μ L. The reactions were incubated at 37°C for 2 hours, 70°C for 10min, and cooled down to 4°C. The DNA was then purified using a Zymo column following the manufacturer's instructions and eluted in 21 μ L.

[0244] Tagmented and 3' blocked DNA from the previous step was used for two separate PCR reactions for each target, one reaction to detect genome rearrangement on the left-side of the target and the other reaction to detect genome rearrangement on the right-side of the target. Each PCR was then performed with a common primer annealed to the adapter region, a target specific blocker and a target specific primer. PCR was performed using a 4-step PCR cycling

protocol with tagmented DNA, Primer 1 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 2 (*e.g.*, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixture was prepared as below: 25 μ L 2x PCR Master Mix, 10 μ L Tagmented DNA, 1.25 μ L Primer 1 (10 μ M), 2.5 μ L Primer 2 (10 μ M), 10 μ L Blocker 0 (10 μ M), and add H₂O up to 50 μ L. PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, 60 seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads (1x) and eluted in 12 μ L.

[0245] Nested PCR was then performed with a barcode tag. PCR was performed with PCR product from the previous step, Primer 3 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 4 (*e.g.*, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-24 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixture was prepared as below: 25 μ L 2x PCR Master Mix, 10 μ L PCR product, 1.25 μ L Primer 3 (10 μ M), 2.5 μ L Primer 4 (10 μ M), 10 μ L Blocker 0 (10 μ M), and add H₂O up to 50 μ L. PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads, and eluted in 12 μ L.

[0246] Tag PCR was then performed with sequencing primers. PCR was performed with PCR product from the previous step, Primer 1 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 6 (*e.g.*, any Oligo-13A, or Illumina indexing primer from New England Biolabs, NEB#E7611A), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixture was prepared as follows: 25 μ L 2x PCR Master Mix, 10 μ L PCR product, 5 μ L Primer 5 (10 μ M), 5 μ L Primer 6 (10 μ M), 5 μ L Blocker 0 (20 μ M). PCR is performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-

amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads, and eluted in 52 µL.

[0247] 50 µL of the supernatant was transferred to a new tube for library quantification using NEBNext library Quant Kit for Illumina (NEB, E7630) following the manufacture's instructions. Libraries were normalized and pooled for sequencing.

[0248] Next generation sequencing was then performed using the PCR product from the previous step, and sequencing was performed using an Illumina platform (NextSeq 2000) with 20% PhiX (Illumina, Cat# FC-110-3001).

[0249] Paired-end sequencing reads from Illumina sequencing were processed using a customized NGS analysis pipeline. The sequence reads were first checked for overall quality using FastQC followed by consolidating FastQC results using MultiQC. The reads were then subjected to Adapter and primer sequence trimming using Cutadapt (DOI:10.14806/ej.17.1.200) followed by trmming low quality bases using Trimmomatic (Bolger et al, Bioinformatics. 2014; 30(15):2114-20). The trimming in the reads was verified by running FastQC and MultQC steps on trimmed reads fastq files to ascertain the removal of adaptors. The resulting reads were aligned with human genome reference sequence (hg38) using BWA-MEM (Li and Durbin, Bioinformatics, 2009; 25:1754-1760) tool with default parameters. The generated Sequence Alignment Map (SAM) file was sorted by coordinate and converted to Binary Alignment Map (BAM) file using PicardTools (<http://broadinstitute.github.io/picard>). The BAM file was then filtered to remove low-quality reads and to keep the reads with quality score ≥ 30 using Samtools. The resulting high quality BAM file was then checked using the CollectInsertSizeMatrix and CollectAlignemntSummaryMetrics modules from PicardTools followed by FastQC and MultiQC report generation. The reads in the sorted high quality read BAM were grouped based on UMIs using UMI-Tools with paired option and saved as grouped BAM followed by BAM indexing using Samtools. The grouped BAM file was then de-duplicated using UMI-Tools with paired option to remove PCR duplicate reads. The FGSV tool was then used to discover the structural variation pileup by searching for split read mapping and read pairs that map across breakpoints in the BAM file using the FGSV SVPileup module. The AggregateSvPileup module of FGSV then aggregated information across nearby pileups to call structural variants. Only the pileups containing at least one breakpoint on the target were

considered as real hits, and as high confidence hits with at least 10 split reads. IGV (Thorvaldsdóttir et al, Brief Bioinform. 2013; 14: 178–192) was used to facilitate the identification and analysis of rearrangements at genomic scale.

[0250] DNA rearrangement events detected for Example 4 in control and edited cells are shown in Table 11 below. The criteria used to filter out noise is the same with Method 1A. As shown in Table 11, 36 and 37 hits of high confidence passed the filters in edited cells, for the left-side and right-side DNA rearrangement events, respectively, while only 0 and 10 hits were captured in control cells for the left-side events, and right-side events, respectively. However, besides more hits, the numbers of the split reads to support these hits were also much higher in edited cells in FIGs. 13A-B.

Table 11. Summary of DNA rearrangement by Example 4 in control and CCR5 edited cell.

Samples	Side	Intra Chromosomal	Inter Chromosomal
Control	Left	0	0
CCR5 edited		27	9
Control	Right	9	1
CCR5 edited		25	12
Note: DNA rearrangement events with split read number <10 were filtered out. One breakpoint must be located at ON target cutting site +/-25 bp.			

[0251] DNA rearrangements identified by Example 4 in CCR5 edited cells are displayed in Table 12 below. The genomic coordinates of the DNA breakpoints between DNA rearrangements, number of unique split-reads based on UMI-tools, extracted reads to show the split reads, and note to further explain the potential mechanism for this DNA rearrangements are included in the Table 12. It is worth noting that the DNA rearrangements between CCR5 and its homologous gene CCR2, inter-chromosomal translocation between CCR5 on chromosome 3 and RNF17/CENPJ on chromosome 13 were also identified by CAST-seq (Turchiano et al, 2021). The presence of fused DNA of chr3/chr13 and acentric chromosome were also validated using PCR (FIGs. 13A-B). Therefore, SAFER detection Method 1B can efficiently capture intra- and inter-chromosomal rearrangements.

Table 12. Examples of DNA rearrangements identified by Example 4 in CCR5 edited cells.

Catego ry	Side	On-target breakpoint	Non-target breakpoint	Unique Split_re ads	Exemplar read	Note
--------------	------	-------------------------	--------------------------	---------------------------	---------------	------

Intra-chromosomal rearrangement	Left	chr3:46373001-46373011	chr3:46372167-46372170	21	Read_name: VH01033:21:AACCM73M5:1:1103:21602:42423:rGCCTTGGGAG Sequence: <u>GAACAGTGATTGGCATCC</u> <u>AGTAGGTGCCCTCTAGGC</u> <u>CTCCGCTCTACTCACTGG</u> TGTTTCATCTTTGGTTTGT GGGCAACATGCTGGTCAT CCTCATCCTGATAAACTG CAAAAGGCTGAAGAGC (SEQ ID NO: 73) Alignments: <u>chr3:46,372,128-46,372,167</u> (+) <u>chr3:46,373,005-46,373,091</u> (+)	The event represents a large deletion caused by CRISPR/Cas9 editing. No off-target site was predicted at the non-target breakpoint site. Gene: CCR5
	Left	chr3:46373024-46373032	chr3:46357598-46357604	525	Read_name: VH01033:21:AACCM73M5:1:1104:14160:45016:rCTACCTTCCT Sequence: GCTCTTCAGCCTTTTGCA GTTTATCAGGATGAGGAT GACCAGCATGTTGCCAC AAAACCAAAGATGATTA <u>TGATTACGGTGCTCCCTG</u> <u>TCATAAATTTGACGTGAA</u> <u>GCAAATTTGGGGCCCAACT</u> <u>CCTGCCTCCGCTCTACTC</u> <u>CCTGGTGT</u> (SEQ ID NO: 74) Alignments: <u>chr3:46,373,026-46,373,091</u> (-) <u>chr3:46,357,598-46,357,684</u> (+) underlined	This event represents an intergenic inversion that is caused by off-target editing or homology mediated recombination. Predicted off-target site: CACCAGCGA GTAGAGCGG AGGCAGGAG (SEQ ID NO: 61) chr3:46357657-46357683 and homology sequence in CCR2 and CCR5: CTGGTGTTC ATCTTTGGTT TTGTGGGCA ACATGCTGG TC (SEQ ID NO: 62) Gene: CCR2, CCR5
	Right	chr3:46372992-46373001	chr3:46357616-46357619	15	Read_name: VH01033:30:AAF5TJWM5:1:1101:57788:4048:rCTAATAACCA Sequence: TACATCGGAGCCCTGCCA AAAAATCAATGTGAAGC AAATCGCAGCCCGCCTCC <u>CTGTCATAAATTTGACGT</u> <u>GAAGCAAATTTGGGGCCC</u>	This event represents an intergenic inversion that is caused by off-target editing or homology mediated recombination.

				<u>AACTCCTGCC (SEQ ID NO: 75)</u> <u>Alignments:</u> chr3:46,372,946-46,372,999 (+) <u>chr3:46,357,619-46,357,664 (+) underlined</u>	Predicted off-target site: CACCAGCGA GTAGAGCGG AGGCAGGAG (SEQ ID NO: 61) chr3:46357657-46357683 and homology sequence in CCR2 and CCR5: CTGGTGTTT ATCTTTGGTT TTGTGGGCA ACATGCTGGTC (SEQ ID NO: 62) Gene: CCR2, CCR5
Right	chr3:46372987-46373001	chr3:46372579-46372588	1293	Read_name: VH01033:30:AAF5TJWM5:1:1103:14974:28678:rCGCCACCCCC Sequence: TACATCGGAGCCCTGCCA AAAAATCAATGTGAAGC AAATCGCAGCCCGCCTCC TGAAACTGGGTCAAGCAT <u>GGTGCAGCTCAGCCTTTTG (SEQ ID NO: 76)</u> <u>Alignments:</u> chr3:46,372,946-46,373,000 (+) <u>chr3:46,372,547-46,372,584 (-) underlined</u>	Predicted off-target site: CACCAGCGA GTAGAGCGG AGGCAGGAG (SEQ ID NO: 61) chr3:46357657-46357683. Intergenic inversion caused by off-target editing or homology mediated recombination. Gene: CCR2, CCR5
Right	chr3:46372986-46372991	chr3:46373766-46373769	1355	Read_name: VH01033:30:AAF5TJWM5:1:2204:19027:32162:rCAGTAGCGGG Sequence: TACATCGGAGCCCTGCCA AAAAATCAATGTGAAGC AAATCGCAGCACTGCTGC <u>ATCAACCCCATCATCTATGCCTTTGTCGGGGAGAAGTTCAGAACTACCTCTTAGTC (SEQ ID NO: 77)</u> <u>Alignments:</u> chr3:46,372,946-46,372,990 (+) <u>chr3:46,373,766-46,373,832 (+) underlined</u>	The event represents a large deletion caused by CRISPR/Cas9 editing. No off-target site was predicted at the non-target breakpoint site. Gene: CCR5

Inter chromosomal rearrangement	Left	chr3:46373002-46373003	chr13:24886326-24886326	15	Read_name: VH01033:21:AACCM73M5:1:2103:56102:51699:rACCGGATCGC Sequence: GATGAGGATGACCAGCATGTTGCCCACAAAACCATAGATGAACACCAAGTGAGTAGAGCGGAGAGAGGTGTGCAGCCAAACAAGCTGCTGGGCGTGTGAGGCGGCTGTGCTGTGCCTGTGAGACAGGAATGGTTCCATTACA C (SEQ ID NO: 78) Alignments chr3:46,373,003-46,373,064 (-) chr13:24,886,326-24,886,413 (+) <u>underlined</u>	These events represent interchromosomal translocation that is mediated by Off-target editing. On-target site: chr3:46372995-46373017 (-) GTGAGTAGA GCGGAGGCA GGAGG (SEQ ID NO: 30) OFF-target site: G()GAGGAGA GCGGAGGCA GGTGG Chr13:24886320-24886342 (+) Gene: CCR5 and RNF17/CENPJ
	Right	chr3:46372995-46373002	chr13:24886335-24886340	418	Read_name: VH01033:30:AAF5TJWM5:1:1302:22757:39640:rTGAACCGGGG Sequence: TACATCGGAGCCCTGCCA AAAAATCAATGTGAAGCAAAATCGCAGCCCGCCTCC TAGGTGGTCAGCCAACAAGCTGTGGGCGTGTGAGGCGGCTGTGCTGTGCCTGTGAGACAGGAATGGTTCCATTACAC (SEQ ID NO: 79) Alignments: chr3:46,372,946-46,373,000 (+) chr13:24,886,336-24,886,413 (+) <u>underlined</u>	

EXAMPLE 5: EXEMPLARY METHOD FOR METHOD IC

[0252] Exemplary oligonucleotides are listed in Table 1.

[0253] Additional reagents include Tn5 transposase, *e.g.*, Robust Tn5 Transposase (Cat#

EMQZ1422, Creative Biogene), DNA polymerase PCR master mix, *e.g.*, Platinum™ SuperFi II PCR Master Mix (ThermoFisher, Cat# 12368010), DNA polymerase with PCR buffer, *e.g.*, Platinum™ SuperFi II DNA Polymerase (ThermoFisher, Cat# 12361010), dNTP, *e.g.*, dNTP Mix (10 mM each) (ThermoFisher, Cat# R0194), DNA purification kit: Column kit, *e.g.*, DNA Clean & Concentrator-5 (Zymo Research, Cat# D4004), Magnetic beads, *e.g.*, AMPure XP Reagent (Beckman, Cat# A63880), Illumina sequencing kit, *e.g.*, NextSeq 1000/2000 P1 Reagents (300 Cycles) (Cat# 20050264), MiSeq Reagent Kit v3 (150-cycle) (Cat# MS-102-

3001), MiSeq Reagent Kit v2 (300-cycles) (Cat# MS-102-2002), PhiX Control v3 (Illumina, Cat# FC-110-3001), USER® Enzyme (New England Biolabs, Cat# M5505S), STE buffer (10mM Tris-HCl (pH 8.0), 1mM EDTA, 0.1M NaCl).

[0254] In the first step, Tn5 transposomes are assembled. Transposon-end containing oligos (*e.g.*, Oligo-10 and Oligo-02 for non-UMI version, or Oligo-27 and Oligo-02 for UMI version) are annealed in STE buffer with the following program: 1–10 minutes at 90–99°C, 40–500 cycles of 1 minutes at 90–99°C (decrease temperature 0.2–2°C every cycle until reaching 1–12°C). Transposome complexes are assembled in TPS buffer following the procedure below, as recommended by the manufacturer (Creative Biogene, Cat# EMQZ1422). The reagents are mixed thoroughly and incubated at 15–40°C for 5–120 minutes.

Component	Volume
Tn5 transposase	1–10 µL
Adaptor	0.5–8 µL
10x TPS buffer	2 µL
Sterile water	add to 20 µL

[0255] The workflow for this example is shown in FIG. 3. The tagmentation reaction is performed in which genomic DNA extracted from gene edited samples or controls (using standard methods) is tagged with the assembled transposome complex following procedure. An exemplary reaction mixture is shown below:

Component	Volume
Genomic DNA	varies (5–500 ng)
5x LM buffer	6 µL
Transposome	0.1–8 µL
Sterile water	add to 30 µL

[0256] The reaction components are mixed thoroughly and incubated at 30–65°C for 5–120 minutes. The tagged DNA is purified using column-based purification.

[0257] First new strand synthesis is then performed. The PCR reaction is set up with fragmented DNA from the previous step, Primer 2 (*e.g.*, Oligo-05 for EMX1 left-side rearrangement, Oligo-15 for EMX1 right-side rearrangement, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: denaturing at 90–99°C for 1–10 minutes, blocker annealing at 75–92°C for 5–120 seconds,

primer annealing at 55–75°C for 1 minute, primer extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. Product is purified using column-based purification.

[0258] USER enzyme cleavage is then performed. Briefly, USER enzyme cleavage is performed with the following procedure: 0.5–10 µL of USER enzyme is added to the product from the previous step, mixed and incubated at 16–45°C for 5–60 minutes. Product is purified using a DNA purification kit.

[0259] USER treated DNA from the previous step is used for two separate PCR reactions for each target, one reaction to detect genome rearrangement on the left-side of the target and the other reaction to detect genome rearrangement on the right-side of the target. Each PCR is then performed with a common primer annealed to the adapter region, a target specific blocker and a target specific primer. PCR is performed using 3-step PCR cycling protocol with tagged DNA, Primer 1 (Oligo-04, common to all targets and rearrangement of both sides), Primer 2 (*e.g.*, Oligo-05 for EMX1 left-side rearrangement, Oligo-15 for EMX1 right-side rearrangement, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 20–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0260] Nested PCR is then performed with barcode tag. PCR is performed with the PCR product from the previous step, Primer 3 (Oligo-06, common to all targets and rearrangement of both sides), Primer 4 (*e.g.*, Oligo-07 for EMX1 left-side rearrangement, Oligo-16 for EMX1 right-side rearrangement, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-24 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–

75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0261] Tag PCR is then performed with sequencing primers. PCR is performed with PCR product from the previous step, Primer 5 (*e.g.*, Oligo-08, an example of Illumina indexing primer from New England Bio Labs, NEB#E7603A), Primer 6 (*e.g.*, Oligo-09, an example of Illumina indexing primer from New England Bio Labs, NEB#E7611A), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0262] Next generation sequencing is then performed using the PCR product from the previous step, and sequencing is performed using an Illumina platform (*e.g.*, Illumina MiSeq or NextSeq 2000) with optional ≥ 0 –95% PhiX (Illumina, Cat# FC-110-3001).

[0263] Paired-end sequencing reads from Illumina sequencing are merged using a paired-end reads merging software, *e.g.*, PEAR (Zhang et al, Bioinformatics. 2014;30: 614–620), FLASH (Magoč and Salzberg, Bioinformatics. 2011;27: 2957–2963), BBMerge (Bushnell et al, PLoS One. 2017;12: e0185056), or NGMerge (Gaspar, Bioinformatics. 2018;19: 536). The merged reads are then trimmed and filtered using software such as BBmap (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>), samtools (Danecek et al, Gigascience. 2021;10. doi:10.1093/gigascience/giab008), or custom scripts to remove Illumina adaptor sequences, low quality reads, reads containing the target flanking sequence on the side of the target site where the blocking oligonucleotide or cleavage reagent binds, and reads that do not contain the target flanking sequence on the side of the target site where the target-specific primer hybridizes. Selected reads are aligned to a human reference genome, *e.g.*, GRCh37 (hg19), GRCh38 (hg38), or Telomere-to-Telomere assembly (*e.g.*, T2T-CHM13v2.0), using an alignment software, *e.g.*, Bowtie 2 (Langmead and Salzberg, Nat Methods. 2012;9: 357–359), BWA (Li and Durbin, Bioinformatics. 2009;25: 1754–1760), or Minimap2 (Li, Bioinformatics. 2018;34: 3094–3100). The aligned BAM file is converted into

bed file using BEDTools (Quinlan, Bioinformatics. 2014; 47: 11.12.1–34). For methods using UMIs, UMIs are collapsed to remove redundant sequencing reads using software such as Gencore (Chen et al, Bioinformatics. 2019;20: 606), UMI-tools (Smith et al, Genome Res. 2017;27: 491–499), UMI-Reducer (github.com/smangul1/UMI-Reducer), or custom scripts. Reads with candidate translocation break points within a suitable window flanking the target site (*e.g.*, within 1, 3, 5, 10, 20, 40, 60, 80, 100, 200, or more bases) are identified and counted to quantify the number of rearrangements between the target site and other genomic loci. Statistical tests are then applied to establish a confidence score to each group of rearrangements at each distal rearranged locus (*e.g.*, within a defined window of up 50, 100, 200, 500, 1000, 2000, 3000, or more bases at the distal rearranged locus). Visualization tools such as IGV (Thorvaldsdóttir et al, Brief Bioinform. 2013; 14: 178–192) or Circos (Krzywinski et al, Genome Res. 2009; 19: 1639–1645) are used to facilitate the identification and analysis of rearrangements at genomic scale.

EXAMPLE 6: EXEMPLARY METHOD FOR METHOD IC

[0264] Exemplary oligonucleotides used are listed in Table 5.

[0265] Additional reagents included unloaded Tn5 Transposase (Diagenode, C01070010, Diagenode), 2X Tagmentation buffer (Diagenode, C01019043), Platinum™ SuperFi II PCR Master Mix (ThermoFisher, Cat# 12368010), dNTP Mix (10 mM each) (ThermoFisher, Cat# R0194), DNA purification kit: DNA Clean & Concentrator-5 (Zymo Research, Cat# D4004), AMPure XP Reagent (Beckman Coulter, Cat# A63880), Glycerol (MilliporeSigma, 356350500ML), 10% SDS (Fisher, BP2436200), 5M NaCl (Invitrogen, AM9760G), Tris HCl 1M, pH8.0 (Invitrogen, 15568-025), Genomic DNA (extracted from edited and unedited 293T cells), ddNTP, Dideoxynucleoside Triphosphate Set (MilliporeSigma, Cat# 03732738001), NEBNext library Quant Kit for Illumina (NEB, E7630), Terminal transferase (TdT, NEB, M0315S), USER, (NEB #M5508), Phusion U Hot Start DNA Polymerase (ThermoFisher, F555S), Illumina NextSeq 1000/2000 P1 Reagents (300 Cycles) (Illumina, Cat# 20050264), PhiX Control v3 (Illumina, Cat# FC-110-3001).

[0266] In the first step, transposon-end containing oligos (*e.g.*, Oligo-02A and Oligo-03A) were resuspend in annealing buffer (40mM Tris-HCl (pH8.0), 50mM NaCl) to stock concentration of 100 μM. In a PCR tube, 10 μl oligo-01A, 10 μl oligo-03A were mixed, vortexed and placed in a thermocycler with the following program: 5 minutes at 95°C, cool to 65°C (-0.1°C/second), 5 minutes at 65°C, cool to 25°C (-0.1°C/second), 5 minutes at 25°C, and hold at 4°C.

[0267] Transposome complexes were assembled using the following procedure, as recommended by the manufacturer. The reagents were mixed in a PCR tube: 10 µL Tn5 transposase (2 ug/ul), 10 µL annealed adaptor. The reagents were mixed thoroughly and incubated at 23°C for 30 minutes. 10 µl glycerol was added and mixed. The assembled transposome complex was stored at -20°C.

[0268] The workflow for this example is shown in FIG 8. The tagmentation reaction was performed using genomic DNA extracted from gene edited samples or controls (using standard methods) with the following procedure:
100 ng genomic DNA from unedited control or edited sample, 20 µL 2x Tagmentation buffer, 200 ng loaded Tn5 Transposase, and add H2O up to 40 µL. The reaction mixtures were mixed thoroughly and incubated at 55°C for 15 minutes. 10 µL 0.2% SDS is added. And then Tn5 was inactivated for 10 min at 70°C. The tagged DNA was purified using a Zymo column following the manufacturer's instructions and eluted in 11 µL.

[0269] The 3' ends of DNA molecules were then blocked using ddNTP and terminal transferase (TdT, NEB, M0315S) using the following procedure: 10 µL tagged DNA, 5 µL 10X TdT buffer, 2 µL (40U) TdT, 5 µL CoCl₂ solution, 1 µL ddNTP (10mM), and added H2O up to 50 µL. The reactions were incubated at 37°C for 2 hours, 70°C for 10min, and cooled down to 4°C. The DNA was then purified using a Zymo column following the manufacturer's instructions and eluted in 21 µL.

[0270] First strand synthesis was then performed using DNA from the previous step in two separate PCR reactions for each target, one reaction to detect genome rearrangement on the left-side of the target and the other reaction to detect genome rearrangement on the right-side of the target. Each PCR is then performed with a common primer annealed to the adapter region, a target specific blocker and a target specific primer. The PCR reaction was set up with fragmented and 3' blocked DNA from the previous step, Primer 2 (*e.g.*, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using Phusion U Hot Start DNA Polymerase (Thermofisher, F555S). The reaction mixture was prepared as follows: 10 µL Blocked DNA, 10 µL 5X Phusion buffer, 0.5 µL Phusion U polymerase, 1 µL dNTP (10 mM), 2.5 µL Primer 2 (10 µM), 10 µL Blocker-0 (10 µM), and add H2O up to 50 µL. PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 1 or 20 cycles of the 4-step

denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, 60 seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads (1x) and eluted in 12 µL.

[0271] USER enzyme cleavage was then performed. Briefly, USER enzyme cleavage was performed using the following procedure: 10 µL Purified DNA, 5 µL(10X) rCutsmart buffer, 2 µL USER, and add H₂O up to 50 µL. The mixture was incubated at 37°C for 30 minutes. The DNA was purified using Zymo columns and eluted in 11 ul for PCR

[0272] USER treated DNA from the previous step was used for the first PCR amplification. Each PCR was performed with a common primer annealed to the adapter region, a target specific blocker and a target specific primer. PCR was performed using 4-step PCR cycling protocol with tagmented DNA, Primer 1 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 2 (*e.g.*, Oligo-06A for CCR5#1 left-side rearrangement, or Oligo-10A for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixture was prepared as below: 25 µL 2x PCR Master Mix, 10 µL Tagmented DNA, 1.25 µL Primer 1 (10 µM), 2.5 µL Primer 2 (10 µM), 10 µL Blocker 0 (10 µM), and add H₂O up to 50 µL. PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, 60 seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads (1x) and eluted in 12 µL.

[0273] Nested PCR was then performed with a barcode tag. PCR was performed using the PCR products from the previous step: Primer 3 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 4 (*e.g.*, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-24 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). The PCR mixture was processed as follows: 25 µL 2x PCR Master Mix, 10 µL PCR product, 1.25 µL Primer 3 (oligo-04, 10 µM), 2.5 µL Primer 4 (10 µM), 10 µL Blocker 0 (10 µM), and add H₂O up to 50 µL. PCR is performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at

78°C, 10 second at 60°C, seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads, and eluted in 12 µL.

[0274] Tag PCR was then performed with sequencing primers. PCR was performed with the product from the previous step, Primer 5 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 6 (*e.g.*, any Oligo-13A, or Illumina indexing primer from New England Biolabs, NEB#E7611A), and Blocker 0 (*e.g.*, Oligo-05A for CCR5#1 left-side rearrangement, or Oligo-09A for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). The PCR mixture was prepared as follows: 25 µL 2x PCR Master Mix, 10 µL PCR product, 5 µL Primer 5 (10 µM), 5 µL Primer 6 (10 µM), 5 µL Blocker 0 (20 µM). PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. The PCR products were purified using AMPure XP beads, and eluted in 52 µL.

[0275] 50 µL of the supernatant was transferred to a new tube for library quantification using the NEBNext library Quant Kit for Illumina (NEB, E7630) following the manufacture's instructions. The libraries were normalized and pooled for sequencing.

[0276] Next generation sequencing was then performed using the PCR product from the previous step, and sequenced using an Illumina platform (NextSeq 2000) with 20% PhiX (Illumina, Cat# FC-110-3001).

[0277] Paired-end sequencing reads from Illumina sequencing were processed using a customized NGS analysis pipeline. The sequence reads were first checked for overall quality using FastQC followed by consolidating FastQC results using MultiQC. The reads were then subjected to Adapter and primer sequence trimming using Cutadapt (DOI:10.14806/ej.17.1.200) followed by trmning low quality bases using Trimmomatic (Bolger et al, Bioinformatics. 2014; 30(15):2114-20). The trimming in the reads was verified by running FastQC and MultQC steps on trimmed reads fastq files to ascertain the removal of adaptors. The resulting reads were aligned with human genome reference sequence (hg38) using BWA-MEM (Li and Durbin, Bioinformatics, 2009; 25:1754- 1760) tool with default parameters. The generated Sequence Alignment Map (SAM) file was sorted by coordinate and converted to Binary Alignment Map (BAM) file using PicardTools (<http://broadinstitute.github.io/picard>). The BAM file was then filtered to remove low-quality

reads and to keep the reads with quality score ≥ 30 using Samtools. The resulting high quality BAM file was then checked using the CollectInsertSizeMatrix and CollectAlignmentSummaryMetrics modules from PicardTools followed by FastQC and MultiQC report generation. The reads in the sorted high quality read BAM were grouped based on UMIs using UMI-Tools with paired option and saved as grouped BAM followed by BAM indexing using Samtools. The grouped BAM file was then de-duplicated using UMI-Tools with paired option to remove PCR duplicate reads. The FGSV tool was then used to discover the structural variation pileup by searching for split read mapping and read pairs that map across breakpoints in the BAM file using the FGSV SVPileup module. The AggregateSVPileup module of FGSV then aggregated information across nearby pileups to call structural variants. Only the pileups containing at least one breakpoint on the target were considered as real hits, and as high confidence hits with at least 10 split reads. IGV (Thorvaldsdóttir et al, Brief Bioinform. 2013; 14: 178–192) was used to facilitate the identification and analysis of rearrangements at genomic scale.

[0278] DNA rearrangement by Example 6 in control and CCR5 edited cells are shown in Table 13 below. The criteria used to filter out noise is the same with Method 1A. For this example, only libraries for left-side events were performed. As shown in Table 13, 38 and 57 hits of high confidence passed the filters in control and in edited cells, respectively. However, the number of unique split support these hits were still substantially higher in edited cells (FIG.14).

Table 13. Summary of DNA rearrangement by Example 6 in control and CCR5 edited cell.

Samples	Side	Intra Chromosomal	Inter Chromosomal
Control	Left	22	16
CCR5 edited		40	17
Note: DNA rearrangement event with split read number <10 were filtered out. One breakpoint must be located at ON target cutting site +/-25 bp. Only the DNA rearrangements of the left side was examined.			

[0279] DNA rearrangements identified by Example 6 in CCR5 edited cells are displayed in Table 14 below. The genomic coordinates of the DNA breakpoints between DNA rearrangements, number of unique split-reads based on UMI-tools, extracted reads to show the split reads, and note to further explain the potential mechanism for this DNA rearrangements were included in the table. It is worth noting that the inter-chromosomal

translocation between CCR5 on chromosome 3 and RNF17/CENPJ on chromosome 13 were also identified by CAST-seq (Turchiano et al, 2021). The presence of fused DNA of chr3/chr13 and acentric chromosome were also validated using PCR (FIG.13A-B). Therefore, SAFER detection Method IC can efficiently capture intra- and inter-chromosomal rearrangements.

Table 14. Examples of DNA rearrangements identified by Example 6 in CCR5 edited cells.

Category	Side	On-target breakpoint	Non-target breakpoint	Unique Split-reads	Exemplar read	Note
Intra chromosomal rearrangement	Left	chr3:46373001-46373004	chr3:46372462-46372467	39	Read_name: VH01033:21:AACCM73M5:1:1204:40178:38655:rGGATTACCT Sequence: <u>CTGTAGTCCCCAGCCACTTGG</u> <u>AGGGGTGAGGTGAGAGGATT</u> <u>GCTTGAGCCCGGGATGGTCCA</u> <u>GGCTCCGCTCTACTACTGGT</u> GTTCATCTTTGGTTATGTGGGC AA (SEQ ID NO: 80) Alignments: <u>chr3:46,372,401-46,372,465 (+)</u> <u>chr3:46,373,003-46,373,046 (+)</u>	The event represents a large deletion caused by CRISPR/Cas9 editing. No off-target site was predicted at the non-target breakpoint site. Gene: CCR5
	Left	chr3:46373002-46373011	chr3:46376507-46376515	706	Read_name: VH01033:21:AACCM73M5:1:2204:49532:17527:rCAACTAAAAG Sequence: GCTCTTCAGCCTTTTGCAGTTT ATCAGGATGAGGATGACCAGC ATGTTGCCACAAAACCAAAG ATGAACACTCAAAAGAATTGC <u>CATGTTTTAAAAAAGAGAATT</u> <u>GGTGCCATAAGTTAGTTGTCT</u> <u>ATGTTTGAAAATGAAGAAGAT</u> AT (SEQ ID NO: 81) Alignments: chr3:46,373,002-46,373,047 (-) chr3:46,376,506-46,376,586 (+) underlined	The event represents a large deletion caused by CRISPR/Cas9 editing. This intragenic Inversion may be occurred between on-target CRISPR editing and naturally happened breakpoint. No off-target site was predicted at the non-target breakpoint site. Gene: CCR5
Inter chromosomal rearrangement	Left	chr3:46373005-46373006	chr13:24886338-24886339	21	Read_name: VH01033:21:AACCM73M5:1:2403:34061:51586:rAACAGCTTGT Sequence: GATGAGGATGACCAGCATGTT GCCACAAAACCAAAGATGA ACACCAGTGAGTAGAGCGTGG	This event represents an interchromosomal translocation that is mediated by Off-target editing.

					TCAGCCAACAAGCTGCTGGGC <u>GTGTGAGGCGGCTGTGCTGTG</u> <u>CCTGTGAGACAGGAATGGTTC</u> <u>CATTACACCATGGGAAATCAC</u> (SEQ ID NO: 82) Alignments: chr3:46,373,005- 46,373,064 (-) chr13:24,886,340-24,886,426 (+)	On-target site: chr3:46372995- 46373017 (-) GTGAGTAGAG CGGAGGCAGG AGG (SEQ ID NO: 30) OFF-target site: G()GAGGAGAG CGGAGGCAGG TGG Chr13:24886320- 24886342 (+) Gene: CCR5 and RNF17/CENPJ
--	--	--	--	--	--	--

EXAMPLE 7: EXEMPLARY METHOD FOR METHOD II

[0280] Exemplary oligonucleotides are listed in Table 1.

[0281] Additional reagents include Tn5 transposase, *e.g.*, Robust Tn5 Transposase (Cat# EMQZ1422, Creative Biogene), DNA polymerase PCR master mix, *e.g.*, Platinum™ SuperFi II PCR Master Mix (ThermoFisher, Cat# 12368010), DNA polymerase with PCR buffer, *e.g.*, Platinum™ SuperFi II DNA Polymerase (ThermoFisher, Cat# 12361010), dNTP, *e.g.*, dNTP Mix (10 mM each) (ThermoFisher, Cat# R0194), DNA purification kit: Column kit, *e.g.*, DNA Clean & Concentrator-5 (Zymo Research, Cat# D4004), Magnetic beads, *e.g.*, AMPure XP Reagent (Beckman, Cat# A63880), Illumina sequencing kit, *e.g.*, NextSeq 1000/2000 P1 Reagents (300 Cycles) (Cat# 20050264), MiSeq Reagent Kit v3 (150-cycle) (Cat# MS-102-3001), MiSeq Reagent Kit v2 (300-cycles) (Cat# MS-102-2002), PhiX Control v3 (Illumina, Cat# FC-110-3001), Sequence-specific nuclease and ribonucleoprotein assembly buffer, *e.g.*, Cas9 Nuclease, *S. pyogenes* (New England Biolabs, Cat# M0386S), EnGen® Lba Cas12a (Cpf1) (New England Biolabs, Cat# M0653S), SpRY, Guide RNA specific to target region, STE buffer (10mM Tris-HCl (pH 8.0), 1mM EDTA, 0.1M NaCl).

Exemplary guide RNA sequences are shown in Table 15 below.

Table 15.

Design group	Sequence (PAM is <u>underlined</u> , not included in sgRNA)	Coordinate in hg38 (chr strand start end)
EMX1_L	CAAACGGCAGAAAGCTGGAGGAGG (SEQ ID NO: 31)	chr2 + 72933822 72933844
EMX1_R	CTCCCCATTGGCCTGCTTCG <u>TGG</u> (SEQ ID NO: 32)	chr2 - 72933905 72933927
CCR5_L	TGACATCAATTATTATACAT <u>CGG</u> (SEQ ID NO: 33)	chr3 + 46372932 46372954

CCR5_R	TTTTCAGTTTATCAGGATGAGG (SEQ ID NO: 34)	chr3 - 46373058 46373080
--------	---	--------------------------

[0282] In the the first step, Tn5 transposomes are assembled. Transposon-end containing oligos (e.g., Oligo-01 and Oligo-11 for non-UMI version, or Oligo-26 and Oligo-11 for UMI version) are annealed in STE buffer with the following program: 1–10 minutes at 90–99°C, 40–500 cycles of 1 minutes at 90–99°C (decrease temperature 0.2–2°C every cycle until reaching 1–12°C). Transposome complexes are assembled in TPS buffer following the procedure below, as recommended by the manufacturer (Creative Biogene, Cat# EMQZ1422). The reagents are mixed thoroughly and incubated at 15–40°C for 5–120 minutes.

Component	Volume
Tn5 transposase	1–10 µL
Adaptor	0.5–8 µL
10x TPS buffer	2 µL
Sterile water	add to 20 µL

[0283] The workflow for this example is shown in FIG. 4. The tagmentation reaction is performed in which genomic DNA extracted from gene edited samples or controls (using standard methods) is tagmented with the assembled transposome complex following procedure. An exemplary reaction mixture is shown below:

Component	Volume
Genomic DNA	varies (5–500 ng)
5x LM buffer	6 µL
Transposome	0.1–8 µL
Sterile water	add to 30 µL

[0284] The reaction components are mixed thoroughly and incubated at 30–65°C for 5–120 minutes. The tagmented DNA is purified using column-based purification.

[0285] Two separate ribonucleoprotein (RNP) assembly reactions and corresponding in vitro cleavage (IVC) reactions are then performed for each target, one reaction to detect genome rearrangement on the left-side of the target and the other reaction to detect genome rearrangement on the right-side of the target. CRISPR/Cas9 Nuclease, *S. pyogenes* (New England Biolabs, Cat# M0386S) and sgRNA (Table 5) are assembled into RNP with the following procedure:

NEBuffer r3.1	3 µL
300 nM sgRNA	0.6–15 µL (6–150 nM final)
1 µM Cas9 Nuclease	0.2–5 µL (6–150 nM final)

Sterile water add to 30 μ L

The reaction mix is mixed thoroughly and incubated at 15–45°C for 1–120 minutes. RNP is added to the tagged DNA at an RNP:DNA ratio of 0.01:1 to 50:1. Cleavage reaction is incubated at 15–45°C for 5–120 minutes. Product is purified using column-based purification.

[0286] Optionally, the 3' ends of DNA molecules are then blocked. 3' ends of DNA molecules are blocked using ddNTP and DNA polymerase SuperFi II (ThermoFisher, Cat# 12361010). Product is purified using column-based purification.

[0287] PCR using IVC DNA from the previous step is then performed with a common primer annealed to the adapter region and a target specific primer. PCR is performed using 3-step PCR cycling protocol with IVC DNA, Primer 1 (Oligo-04, common to all targets and rearrangement of both sides) and Primer 2 (*e.g.*, Oligo-05 for EMX1 left-side rearrangement, Oligo-15 for EMX1 right-side rearrangement, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 20–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0288] Nested PCR is then performed with barcode tag. PCR is performed with PCR product from the previous step, Primer 3 (Oligo-06, common to all targets and rearrangement of both sides), Primer 4 (*e.g.*, Oligo-07 for EMX1 left-side rearrangement, Oligo-16 for EMX1 right-side rearrangement, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-24 for CCR5#1 right-side rearrangement), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

- [0289] Tag PCR is then performed with sequencing primers. PCR is performed with PCR product from the previous step, Primer 5 (*e.g.*, Oligo-08, an example of Illumina indexing primer from New England Bio Labs, NEB#E7603A), Primer 6 (*e.g.*, Oligo-09, an example of Illumina indexing primer from New England Bio Labs, NEB#E7611A), and Blocker 0 (*e.g.*, Oligo-03 for EMX1 left-side rearrangement, Oligo-14 for EMX1 right-side rearrangement, Oligo-18 for CCR5#1 left-side rearrangement, or Oligo-22 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 4-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 seconds at 75–92°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.
- [0290] Next generation sequencing is then performed using the PCR product from the previous step, and sequencing is performed using an Illumina platform (*e.g.*, Illumina MiSeq or NextSeq 2000) with optional ≥ 0 –95% PhiX (Illumina, Cat# FC-110-3001).
- [0291] Paired-end sequencing reads from Illumina sequencing are merged using a paired-end reads merging software, *e.g.*, PEAR (Zhang et al, *Bioinformatics*. 2014;30: 614–620), FLASH (Magoč and Salzberg, *Bioinformatics*. 2011;27: 2957–2963), BBMerge (Bushnell et al, *PLoS One*. 2017;12: e0185056), or NGmerge (Gaspar, *Bioinformatics*. 2018;19: 536). The merged reads are then trimmed and filtered using software such as BBmap (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>), samtools (Danecek et al, *Gigascience*. 2021;10. doi:10.1093/gigascience/giab008), or custom scripts to remove Illumina adaptor sequences, low quality reads, reads containing the target flanking sequence on the side of the target site where the blocking oligonucleotide or cleavage reagent binds, and reads that do not contain the target flanking sequence on the side of the target site where the target-specific primer hybridizes. Selected reads are aligned to a human reference genome, *e.g.*, GRCh37 (hg19), GRCh38 (hg38), or Telomere-to-Telomere assembly (*e.g.*, T2T-CHM13v2.0), using an alignment software, *e.g.*, Bowtie 2 (Langmead and Salzberg, *Nat Methods*. 2012; 9: 357–359), BWA (Li and Durbin, *Bioinformatics*. 2009; 25: 1754–1760), or Minimap2 (Li, *Bioinformatics*. 2018;34: 3094–3100). The aligned BAM file is converted into bed file using BEDTools (Quinlan, *Bioinformatics*. 2014; 47: 11.12.1–34). For methods using UMIs, UMIs are collapsed to remove redundant sequencing reads using software such

as Gencore (Chen et al, Bioinformatics. 2019;20: 606), UMI-tools (Smith et al, Genome Res. 2017;27: 491–499), UMI-Reducer (github.com/smangul1/UMI-Reducer), or custom scripts. Reads with candidate translocation break points within a suitable window flanking the target site (*e.g.*, within 1, 3, 5, 10, 20, 40, 60, 80, 100, 200, or more bases) are identified and counted to quantify the number of rearrangements between the target site and other genomic loci. Statistical tests are then applied to establish a confidence score to each group of rearrangements at each distal rearranged locus (*e.g.*, within a defined window of up 50, 100, 200, 500, 1000, 2000, 3000, or more bases at the distal rearranged locus). Visualization tools such as IGV (Thorvaldsdóttir et al, Brief Bioinform. 2013; 14: 178–192) or Circos (Krzywinski et al, Genome Res. 2009; 19: 1639–1645) are used to facilitate the identification and analysis of rearrangements at genomic scale.

EXAMPLE 8: EXEMPLARY METHOD FOR METHOD II

[0292] The Exemplary oligonucleotides used are listed in Table 5.

[0293] Additional reagents included unloaded Tn5 Transposase (Diagenode, C01070010, Diagenode), 2X Tagmentation buffer (Diagenode, C01019043), Platinum™ SuperFi II PCR Master Mix (ThermoFisher, Cat# 12368010), dNTP Mix (10 mM each) (ThermoFisher, Cat# R0194), DNA purification kit: DNA Clean & Concentrator-5 (Zymo Research, Cat# D4004), AMPure XP Reagent (Beckman Coulter, Cat# A63880), Glycerol (MilliporeSigma, 356350500ML), 10% SDS (Fisher, BP2436200), 5M NaCl (Invitrogen, AM9760G), Tris HCl 1M, pH8.0 (Invitrogen, 15568-025), Genomic DNA (extracted from edited and unedited 293T cells), ddNTP, Dideoxynucleoside Triphosphate Set (MilliporeSigma, Cat# 03732738001), NEBNext library Quant Kit for Illumina (NEB, E7630), Terminal transferase (TdT, NEB, M0315S), Alt-R™ S.p. HiFi Cas9 Nuclease V3 (IDT, 1081060), NEB buffer r3.1 (NEB, B6003S), Illumina NextSeq 1000/2000 P1 Reagents (300 Cycles) (Illumina, Cat# 20050264), PhiX Control v3 (Illumina, Cat# FC-110-3001).

Table 16. Exemplary guide RNA sequences are shown in below.

OligoID	Design group	Sequence (PAM is <u>underlined</u> , not included in sgRNA)	Coordinate in hg38 (chr strand start end)
Oligo-14	CCR5_L	TGACATCAATTATTATACAT <u>CCG</u> (SEQ ID NO: 33)	chr3 + 46372932 46372954
Oligo-15	CCR5_R	TTTTCAGTTTATCAGGATG <u>AGG</u> (SEQ ID NO: 34)	chr3 - 46373058 46373080

[0294] In the first step, transposon-end containing oligos (*e.g.*, Oligo-01A and Oligo-03A, or Oligo-02A and Oligo-03A) were resuspended in annealing buffer (40mM Tris-HCl (pH8.0), 50mM NaCl) to stock concentration of 100 μ M. In a PCR tube, 10 μ l oligo-01, 10 μ l oligo-03 were mixed, vortexed and placed in PCR tubes in a thermocycler with the following program: 5 minutes at 95°C, cool to 65°C (-0.1°C/second), 5 minutes at 65°C, cool to 25°C (-0.1°C/second), 5 minutes at 25°C, and hold at 4°C.

Transposome complexes were assembled following the procedure below, as recommended by the manufacturer. The reagents were mixed in a PCR tube: 10 μ L Tn5 transposase (2 ug/ μ l), 10 μ L annealed adaptor. The reagents were mixed thoroughly and incubated at 23°C for 30 minutes. 10 μ l glycerol was added and mixed. The assembled transposome complex was stored at -20°C.

[0295] The workflow for this example is shown in FIG. 9. The tagmentation reaction was performed using genomic DNA extracted from gene edited samples or controls (using standard methods) using the following procedure: 100 ng genomic DNA from unedited control or edited sample was combined with 20 μ L 2x Tagmentation buffer, 200 ng loaded Tn5 Transposase, and added H₂O up to 40 μ L. The reaction mixtures were mixed thoroughly and incubated at 55°C for 15 minutes. 10 μ L 0.2% SDS was added, and then Tn5 was inactivated for 10 min at 70°C. The tagmented DNA was purified using Zymo columns following the manufacturer's instructions and eluted in 21 μ L.

[0296] Two separate ribonucleoprotein (RNP) assembly reactions and corresponding *in vitro* cleavage (IVC) reactions were then performed for each target, one reaction to detect genome rearrangements on the left-side of the target and the other reaction to detect genome rearrangements on the right-side of the target. CRISPR/Cas9 Nuclease, Alt-R™ S.p. HiFi Cas9 Nuclease V3 (IDT, 1081060), and CCR5 sgRNA (*e.g.*, CCR5_L for left-side rearrangement, or CCR5_R right-side rearrangement) (Table 5A) were assembled into RNP with the following procedure: 32.5 μ L H₂O, 5 μ L (10X) NEBuffer r3.1, 1.5 μ L CCR5 sgRNA (10 μ M), 0.8 μ L HiFi Cas9 Nuclease (diluted to 6.2 μ M). The RNP reactions were mixed thoroughly and incubated at room temperature for 10 minutes.

[0297] Tagmented DNA was cleaved *in vitro* using Cas9 RNP assembled in the previous step. 10 μ L DNA was added into the RNP solution and incubated at 37°C for 2 hours. The products were purified using Zymo columns following manufacture's instructions and eluted in 11 μ L.

[0298] The 3' ends of DNA molecules were then blocked using ddNTP and terminal transferase (TdT, NEB, M0315S) using following procedure: 10 μ L tagmented DNA, 5 μ L 10X TdT buffer, 2 μ L (40U) TdT, 5 μ L CoCl₂ solution, 1 μ L ddNTP (10mM), and add up to 50 μ L. The reactions were incubated at 37°C for 2 hours, 70°C for 10min, and cooled down to 4°C. The DNA was then purified using Zymo column following the manufacturer's instructions and eluted in 11 μ L.

[0299] PCR using blocked IVC DNA from the previous step was then performed with a common primer annealed to the adapter region, and a target specific primer. PCR was performed using 4-step PCR cycling protocol with tagmented DNA, Primer 1 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 2 (*e.g.*, Oligo-06A for CCR5#1 left-side rearrangement, or Oligo-10A for CCR5#1 right-side rearrangement), using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixture was prepared as follows: 25 μ L 2x PCR Master Mix, 10 μ L Tagmented DNA, 1.25 μ L Primer 1 (10 μ M), 2.5 μ L Primer 2 (10 μ M), 10 μ L Blocker 0 (10 μ M), and add H₂O up to 50 μ L. PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, 60 seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads (1x) and eluted in 12 μ L.

[0300] Nested PCR was then performed with a barcode tag. PCR was performed with the PCR products from the previous step, Primer 3 (Oligo-04A, common to all targets and rearrangement of both sides), Primer 4 (*e.g.*, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-24 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixtures were prepared as follows: 25 μ L 2x PCR Master Mix, 10 μ L PCR product, 1.25 μ L Primer 3 (10 μ M), 2.5 μ L Primer 4 (10 μ M), 10 μ L Blocker 0 (10 μ M), and add H₂O up to 50 μ L. PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads, and eluted in 12 μ L.

[0301] Tag PCR was then performed with sequencing primers. PCR was performed using the PCR products from the previous step, Primer 1 (Oligo-04A, common to all targets and rearrangement of both sides), (*e.g.*, any Oligo-13A, or Illumina indexing primer from New

England Biolabs, NEB#E7611A) and SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010). PCR mixtures were prepared as follows: 25 µL 2x PCR Master Mix, 10 µL PCR product, 5 µL Primer 5 (10 µM), 5 µL Primer 6 (10 µM), 5 µL Blocker 0 (20 µM). PCR was performed using the following program: initial denaturing at 98°C for 30 seconds, 15 cycles of the 4-step denaturing-annealing-amplification (10 seconds at 98°C, 10 seconds at 78°C, 10 second at 60°C, seconds at 72°C), and final extension at 72°C for 5 minutes, and hold at 4°C. PCR products were purified using AMPure XP beads, and eluted in 52 µL.

[0302] 50 µL of the supernatant was transferred to a new tube for library quantification using the NEBNext library Quant Kit for Illumina (NEB, E7630) following the manufacture's instructions. Libraries were normalized and pooled for sequencing.

[0303] Next generation sequencing was then performed using the PCR product from the previous step, and sequencing was performed using an Illumina platform (NextSeq 2000) with 20% PhiX (Illumina, Cat# FC-110-3001).

[0304] Paired-end sequencing reads from Illumina sequencing were processed using a customized NGS analysis pipeline. The sequence reads were first checked for overall quality using FastQC followed by consolidating FastQC results using MultiQC. The reads were then subjected to Adapter and primer sequence trimming using Cutadapt (DOI:10.14806/ej.17.1.200) followed by trimming low quality bases using Trimmomatic (Bolger et al, Bioinformatics. 2014; 30(15):2114-20). The trimming in the reads was verified by running FastQC and MultiQC steps on trimmed reads fastq files to ascertain the removal of adapters. The resulting reads were aligned with human genome reference sequence (hg38) using BWA-MEM (Li and Durbin, Bioinformatics, 2009; 25:1754-1760) tool with default parameters. The generated Sequence Alignment Map (SAM) file was sorted by coordinate and converted to Binary Alignment Map (BAM) file using PicardTools (<http://broadinstitute.github.io/picard>). The BAM file was then filtered to remove low-quality reads and to keep the reads with quality score ≥ 30 using Samtools. The resulting high quality BAM file was then checked using the CollectInsertSizeMatrix and CollectAlignmentSummaryMetrics modules from PicardTools followed by FastQC and MultiQC report generation. The reads in the sorted high quality read BAM were grouped based on UMIs using UMI-Tools with paired option and saved as grouped BAM followed by BAM indexing using Samtools. The grouped BAM file was then de-duplicated using UMI-

Tools with paired option to remove PCR duplicate reads. The FGSV tool was then used to discover the structural variation pileup by searching for split read mapping and read pairs that map across breakpoints in the BAM file using the FGSV SVPileup module. The AggregateSVPileup module of FGSV then aggregated information across nearby pileups to call structural variants. Only the pileups containing at least one breakpoint on the target were considered as real hits, and as high confidence hits with at least 10 split reads. IGV (Thorvaldsdóttir et al, Brief Bioinform. 2013; 14: 178–192) was used to facilitate the identification and analysis of rearrangements at genomic scale.

[0305] DNA rearrangements detected by Example 8 in control and CCR5 edited cells are shown in Table 17 below. The criteria used to filter out noise is the same with Method 1A. As shown in Table 17, 27 and 30 hits of high confidence passed the filters in edited cells, for the left-side and right-side DNA rearrangement events, respectively, while only 0 and 12 hits were captured in control cells for the left-side events, and right-side events, respectively. However, besides more hits, the numbers of the split reads to support these hits were also much higher in edited cells in FIG. 15.

Table 17. Summary of DNA rearrangement by Example 8 in control and CCR5 edited cells.

Samples	Side	Intra Chromosomal	Inter Chromosomal
Control	Left	0	0
CCR5 edited		20	7
Control	Right	11	1
CCR5 edited		23	7
Note: DNA rearrangement event with split read number <10 were filtered out. One breakpoint must be located at ON target cutting site +/-25 bp.			

Comparison of split reads between control and CCR5 edited cells shown in FIG. 15.

[0306] DNA rearrangements identified by Example 8 in CCR5 edited cells are displayed in Table 18 below. The genomic coordinates of the DNA breakpoints between DNA rearrangements, number of unique split-reads based on UMI-tools, extracted reads to show the split reads, and note to further explain the potential mechanism for this DNA rearrangements were included in the table. It is worth noting that the DNA rearrangements between CCR5 and its homologous gene CCR2, inter-chromosomal translocation between CCR5 on chromosome 3 and RNF17/CENPJ on chromosome 13 were also identified by CAST-seq (Turchiano et al, 2021). The presence of fused DNA of chr3/chr13 and acentric

chromosome were also validated using PCR (FIGs. 13A-B). Therefore, SAFER detection Method II can efficiently capture intra- and inter-chromosomal rearrangements.

Table 18. Summary of DNA rearrangement by Example 8 in CCR5 edited cells.

Category	Side	On-target breakpoint	Non-target breakpoint	Unique Split_reads	Exemplar read	Note
Intra chromosomal rearrangement	Left	chr3:46373001-46373007	chr3:46372545-46372561	1023	Read_name: VH01033:21:AACCM73M5:1:1303:41825:48898:rTGATGGCGGG Sequence: <u>CCTGGGCGACAGAGTGA</u> <u>GACCCTGTCTCACAACA</u> <u>ACAACAACAACA</u> <u>ACTCC</u> GCTCTACTCACTGGTGTTCATCTTTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAAACTGC AAAAGGCTGAAGAGC (SEQ ID NO: 83) Alignments: chr3:46,373,003-46,373,091 (+) <u>chr3:46,372,500-46,372,548 (+) underlined</u>	The event represents a large deletion caused by CRISPR/Cas9 editing. No off-target site was predicted at the non-target breakpoint site. Gene: CCR5
	Left	chr3:46373003-46373003	chr3:46486211-46486214	372	Read_name: VH01033:21:AACCM73M5:1:1202:27775:37879:rACGGCCTCCT Sequence: ATGAGGATGACCAGCATGTTGCCCAAAAACCAAAGATGAACACCAAGTGAGTAGAGCGGAGAGCAGACAAACAGACAAACAAACAAAACACCTGAAGCCCCTGACTAACAATCAATAGGCAACATCCAGGAAGATTGTAACCCCATAGTAC (SEQ ID NO: 84) Alignments: chr3:46,373,003-46,373,063 (-) <u>chr3:46,486,210-46,486,295 (+) underlined</u>	This event represents an intergenic inversion that is caused by off-target editing or homology mediated recombination. Predicted off-target site: CACCAGCGAGTAGAGCGGAGGCAGGAG (SEQ ID NO: 61) chr3:46357657-46357683 and homology sequence in CCR2 and CCR5: CTGGTGTTTCATCTTTGGTTTTGTGGGCAACATGCTGGTC (SEQ ID NO: 62) Gene: CCR2, CCR5
	Right	chr3:46372990-46373002	chr3:46373945-46373956	6566	Read_name: VH01033:30:AAF5TJWM5:1:2203: Sequence: 57636:20822:rACCTCGGA AA TACATCGGAGCCCTGCC	The event represents a large deletion caused by CRISPR/Cas9 editing. No off-target site was predicted at the non-target breakpoint site.

					AAAAAATCAATGTGAAG CAAATCGCAGCCCGCT CTGTGGGCTTGTGACAC GGACTCAAGTGGGCTGG TGACCCAGTCAGAGTTG TGCACATGGCTTAGTTTT CATACACAGCCTGGGCT GGGG (SEQ ID NO: 85) Alignments: chr3:46,372,946-46,372,995 (+) <u>chr3:46,373,946-46,374,037</u> (+) <u>underlined</u>	Gene: CCR5
	Righ ht	chr3:4637 2978- 46373001	chr3:463 72540- 4637265 4	10651	Read_name: VH01033:30:AAF5TJWM5: 1:1101:32982:28773:rCTAA TCAAAC Sequence: AGCCTGGGCGACAGAGT GAGACCCTGTCTCACAA CAACAACAACAACAACA AAAAGGCTGAGCTGCAC CATGCTTGACCCAGTTT CTTAACTGGAGGCGGGC TGCGATTGCTTCACATT GATTTTTTGGCAGGGCT CCGATGTA (SEQ ID NO: 86) Alignments: <u>chr3:46,372,498-46,372,588</u> (+) <u>underlined</u> chr3:46,372,946-46,373,000 (-)	This event represents an intergenic inversion that is caused by off- target editing or homology mediated recombination. Predicted off-target site: CACCAGCGAGTAG AGCGGAGGCAGGA G (SEQ ID NO: 61) chr3:46357657- 46357683 and homology sequence in CCR2 and CCR5: CTGGTGTTCATCTT TGGTTTTGTGGGCA ACATGCTGGTC (SEQ ID NO: 62) Gene: CCR2, CCR5
	right	chr3:4637 2986- 46373001	chr3:463 57615- 4635763 5	3805	Read_name: VH01033:30:AAF5TJWM5: 1:2104:20844:22204:rGAAC ATGCTT Sequence: TACATCGGAGCCCTGCC AAAAAATCAATGTGAAG CAAATCGCAGCCCGCT CCCTGTCATAAATTTGA CGTGAAGCAAATTGGGG CCCAACTCCTGCC (SEQ ID NO: 87) Alignments: chr3:46,372,946-46,372,999 (+) <u>chr3:46,357,619-46,357,664</u> (+) <u>underlined</u>	This event represents an intergenic inversion that is caused by off- target editing or homology mediated recombination. Predicted off-target site: CACCAGCGAGTAG AGCGGAGGCAGGA G (SEQ ID NO: 61) chr3:46357657- 46357683 and homology sequence in CCR2 and CCR5: CTGGTGTTCATCTT TGGTTTTGTGGGCA ACATGCTGGTC (SEQ ID NO: 62) Gene: CCR2, CCR5
Inter chromo somal rearrang ement	Lef t	chr3:4637 2999- 46373012	chr13:24 886337- 2488635 3	459	Read_name: VH01033:21:AACCM73M5: 1:1303:21394:11999:rGCGG CGGACG Sequence:	These events represent interchromosomal translocation that is mediated by Off-target editing.

					<u>CAGCACAGCCGCCTCAC</u> <u>ACGCCCAGCAGCTTGTT</u> <u>GGCTGACCTCCGCTCTA</u> CTCACTGGTGTTTCATCTT TGGTTTTGTGGGCAACA TGCTGGTCATCCTCATCC TGATAAACTGCAAAAGG CTGAAGAGC (SEQ ID NO: 88) Alignments: chr3:46,373,001-46,373,090 (+) <u>chr13:24,886,342-24,886382</u> <u>(-) underlined</u>	On-target site: chr3:46372995- 46373017 (-) GTGAGTAGAGCGG AGGCAGGAGG (SEQ ID NO: 30) OFF-target site: G()GAGGAGAGCGG AGGCAGGTGG Chr13:24886320- 24886342 (+) Gene: CCR5 and RNF17/CENPJ
	Rig ht	chr3:4637 2995- 46373001	chr13:24 886335- 2488633 7	127	Read_name: VH01033:30:AAF5TJWM5: 1:2303:30350:25763:rCTCC CGCCCC Sequence: TACATCGGAGCCCTGCC AAAAAATCAATGTGAAG CAAATCGCAGCCCGCAG <u>GTGGTCAGCCACAAGC</u> <u>TGCTGGGCGTGTGAGGC</u> <u>GGCTGTGCTGTGCCTGT</u> GAGACAGGAATGGTTCC ATTACTC (SEQ ID NO: 89) Alignments: chr3:46,372,946-46,372,995 (+) <u>chr13:24,886,336-24,886,413</u> <u>(+) underlined</u>	

EXAMPLE 9: **EXEMPLARY METHOD FOR METHOD III**

[0307] Exemplary oligonucleotides are listed in Table 1.

[0308] Additional reagents include Tn5 transposase, *e.g.*, Robust Tn5 Transposase (Cat# EMQZ1422, Creative Biogene), DNA polymerase PCR master mix, *e.g.*, Platinum™ SuperFi II PCR Master Mix (ThermoFisher, Cat# 12368010), DNA polymerase with PCR buffer, *e.g.*, Q5 High-Fidelity DNA Polymerase (New England Biolabs, Cat# M0491S), dNTP, *e.g.*, dNTP Mix (10 mM each) (ThermoFisher, Cat# R0194), DNA purification kit: Column kit, *e.g.*, DNA Clean & Concentrator-5 (Zymo Research, Cat# D4004), Magnetic beads, *e.g.*, AMPure XP Reagent (Beckman, Cat# A63880), RNA purification kit, *e.g.*, RNA Clean & Concentrator-5 (Zymo Research, Cat# R1013), Illumina sequencing kit, *e.g.*, NextSeq 1000/2000 P1 Reagents (300 Cycles) (Cat# 20050264), MiSeq Reagent Kit v3 (150-cycle) (Cat# MS-102-3001), MiSeq Reagent Kit v2 (300-cycles) (Cat# MS-102-2002), PhiX

Control v3 (Illumina, Cat# FC-110-3001), Strand displacement polymerase, *e.g.*, phi29 DNA Polymerase (New England Biolabs, Cat# M0269S), or Bst 3.0 DNA Polymerase (New England Biolabs, Cat# M0374S), DNase I (RNase-free) (New England Biolabs, Cat# M0303S), RNase H, *e.g.*, (New England Biolabs, Cat# M0297S), In vitro transcription kit, *e.g.*, MAXIscript™ T7 Transcription Kit (ThermoFisher, Cat# AM1312), Reverse transcription kit, *e.g.*, SuperScript IV First-Strand Synthesis System (ThermoFisher, Cat# 18091050), STE buffer (10mM Tris-HCl (pH 8.0), 1mM EDTA, 0.1M NaCl), Magnesium Chloride (MgCl₂) Solution (New England Biolabs, Cat# B9021S).

[0309] In the the first step, Tn5 transposomes are assembled. Transposon-end containing oligos (*e.g.*, Oligo-01 and Oligo-02 for non-UMI version, or Oligo-26 and Oligo-02 for UMI version) are annealed in STE buffer with the following program: 1–10 minutes at 90–99°C, 40–500 cycles of 1 minutes at 90–99°C (decrease temperature 0.2–2°C every cycle until reaching 1–12°C). Transposome complexes are assembled in TPS buffer following the procedure below, as recommended by the manufacturer (Creative Biogene, Cat# EMQZ1422). The reagents are mixed thoroughly and incubated at 15–40°C for 5–120 minutes.

Component	Volume
Tn5 transposase	1–10 µL
Adaptor	0.5–8 µL
10x TPS buffer	2 µL
Sterile water	add to 20 µL

[0310] The workflow for this example is shown in FIG. 5. The tagmentation reaction is performed in which genomic DNA extracted from gene edited samples or controls (using standard methods) is tagmented with the assembled transposome complex following procedure. An exemplary reaction mixture is shown below:

Component	Volume
Genomic DNA	varies (5–500 ng)
5x LM buffer	6 µL
Transposome	0.1–8 µL
Sterile water	add to 30 µL

[0311] The reaction components are mixed thoroughly and incubated at 30–65°C for 5–120 minutes. Transposase is then removed by 60–80°C for 5–120 minutes in the presence of 0.1–10 mM EDTA. Both ends of the transposed fragments is filled and extended by Q5 High-Fidelity DNA Polymerase (New England Biolabs, Cat# M0491S) at 60–75°C for 5–600

seconds in the presence of 0.2–4 mM MgCl₂ and 20–1000 μM dNTPs, without the supplement of specific Q5 reaction buffer. Products are purified using column-based purification.

[0312] In vitro transcription is then performed using MAXIscript™ T7 Transcription Kit (ThermoFisher, Cat# AM1312) and the tagmented DNA from the previous step, by incubating at 25–45°C for 1–600 min. The next step is performed immediately.

[0313] TURBO DNase from the MAXIscript™ T7 Transcription Kit (ThermoFisher, Cat# AM1312) is added to the previous reaction mixture and incubated at 20–45°C for 1–240 min. RNA product is purified using an RNA purification kit, *e.g.*, RNA Clean & Concentrator-5 (Zymo Research, Cat# R1013).

[0314] DNA probe hybridization and following RNase H treatment are then performed in two separate reactions for each target, one reaction for detecting genome rearrangement on the left-side of the target and the other reaction for detecting genome rearrangement on the right-side of the target. Probe 0 (*e.g.*, Oligo-13 for EMX1 left-side rearrangement, Oligo-17 for EMX1 right-side rearrangement, Oligo-21 for CCR5#1 left-side rearrangement, or Oligo-25 for CCR5#1 right-side rearrangement)) is added to RNA products purified from the previous step with RNase H Reaction Buffer. Mixture is denatured at 85–99°C for 1–10 minutes and cool down to 4–28°C. RNase H is added to the reaction and incubated at 16–45°C for 1–300 minutes. RNase H is then inactivated at 60–90°C for 5–240 minutes. Product is purified using an RNA purification kit, *e.g.*, RNA Clean & Concentrator-5 (Zymo Research, Cat# R1013).

[0315] Reverse transcription is then performed. Briefly, RNase H treated RNA is reverse transcribed to DNA using target specific Primer 2 (*e.g.*, Oligo-05 for EMX1 left-side rearrangement, Oligo-15 for EMX1 right-side rearrangement, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement) and SuperScript IV First-Strand Synthesis System (ThermoFisher, Cat# 18091050). First, RNA, dNTPs and primer are mixed and heated at 65°C for 1–20 minutes, then incubated at 1–12°C for 10–600 seconds for primer annealing. Next, 5x SSIV buffer, DTT, Ribonuclease inhibitor and SuperScript IV Reverse Transcriptase are added to the mixture, and the combined reaction mixture is incubated at 40–60°C for 2–120 minutes. The reaction is inactivated by incubating it at 60–90°C for 5–240 minutes. To remove RNA, 0.1–10 μL of RNase H is added to the reaction mixture, mixed, and incubated at 20–55°C for 5–120 minutes. Product is purified using column-based purification.

[0316] The first-strand cDNA synthesized from the previous step is used in the first PCR with common Primer 1 (Oligo-04, common to all targets and rearrangement of both sides), target specific Primer 2 (*e.g.*, Oligo-05 for EMX1 left-side rearrangement, Oligo-15 for EMX1 right-side rearrangement, Oligo-19 for CCR5#1 left-side rearrangement, or Oligo-23 for CCR5#1 right-side rearrangement), and SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 20–40 cycles of the 3-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0317] Nested PCR is then performed with barcode tags. PCR is performed with PCR product from the previous step, Primer 3 (Oligo-06, common to all targets and rearrangement of both sides) and Primer 4 (*e.g.*, Oligo-07 for EMX1 left-side rearrangement, Oligo-16 for EMX1 right-side rearrangement, Oligo-20 for CCR5#1 left-side rearrangement, or Oligo-24 for CCR5#1 right-side rearrangement) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 3-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0318] Tag PCR is then performed with sequencing primers. PCR is performed with PCR product from the previous step, Primer 5 (*e.g.*, Oligo-08, an example of Illumina indexing primer from New England Bio Labs, NEB#E7603A), and Primer 6 (*e.g.*, Oligo-09, an example of Illumina indexing primer from New England Bio Labs, NEB#E7611A) using SuperFi II Polymerase PCR Master Mix (ThermoFisher, Cat# 12368010), with the following program: initial denaturing at 90–99°C for 1–5 minute, 4–40 cycles of the 3-step denaturing-annealing-amplification (5–60 seconds at 90–99°C, 5–60 second at 55–75°C, 5–60 seconds at 67–75°C), and final extension at 67–75°C for 1–10 minutes, and hold at 4–12°C. PCR products are purified using column-based purification.

[0319] Next generation sequencing is then performed using the PCR product from the previous step, and sequencing is performed using an Illumina platform (*e.g.*, Illumina MiSeq or NextSeq 2000) with optional $\geq 0-95\%$ PhiX (Illumina, Cat# FC-110-3001).

- [0320] Paired-end sequencing reads from Illumina sequencing are merged using a paired-end reads merging software, *e.g.*, PEAR (Zhang et al, *Bioinformatics*. 2014;30: 614–620), FLASH (Magoč and Salzberg, *Bioinformatics*. 2011;27: 2957–2963), BBMerge (Bushnell et al, *PLoS One*. 2017;12: e0185056), or NGmerge (Gaspar, *Bioinformatics*. 2018;19: 536). The merged reads are then trimmed and filtered using software such as BBmap (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>), samtools (Danecek et al, *Gigascience*. 2021;10. doi:10.1093/gigascience/giab008), or custom scripts to remove Illumina adaptor sequences, low quality reads, reads containing the target flanking sequence on the side of the target site where the blocking oligonucleotide or cleavage reagent binds, and reads that do not contain the target flanking sequence on the side of the target site where the target-specific primer hybridizes. Selected reads are aligned to a human reference genome, *e.g.*, GRCh37 (hg19), GRCh38 (hg38), or Telomere-to-Telomere assembly (*e.g.*, T2T-CHM13v2.0), using an alignment software, *e.g.*, Bowtie 2 (Langmead and Salzberg, *Nat Methods*. 2012; 9: 357–359), BWA (Li and Durbin, *Bioinformatics*. 2009; 25: 1754–1760), or Minimap2 (Li, *Bioinformatics*. 2018;34: 3094–3100). The aligned BAM file is converted into bed file using BEDTools (Quinlan, *Bioinformatics*. 2014; 47: 11.12.1–34). For methods using UMIs, UMIs are collapsed to remove redundant sequencing reads using software such as Gencore (Chen et al, *Bioinformatics*. 2019;20: 606), UMI-tools (Smith et al, *Genome Res*. 2017;27: 491–499), UMI-Reducer (github.com/smangul1/UMI-Reducer), or custom scripts.
- [0321] Reads with candidate translocation break points within a suitable window flanking the target site (*e.g.*, within 1, 3, 5, 10, 20, 40, 60, 80, 100, 200, or more bases) are identified and counted to quantify the number of rearrangements between the target site and other genomic loci. Statistical tests are then applied to establish a confidence score to each group of rearrangements at each distal rearranged locus (*e.g.*, within a defined window of up 50, 100, 200, 500, 1000, 2000, 3000, or more bases at the distal rearranged locus). Visualization tools such as IGV (Thorvaldsdóttir et al, *Brief Bioinform*. 2013; 14: 178–192) or Circos (Kryzyski et al, *Genome Res*. 2009; 19: 1639–1645) are used to facilitate the identification and analysis of rearrangements at genomic scale.
- [0322] No admission is made that any reference cited herein constitutes prior art. The discussion of the references states what their authors assert, and the Applicant reserves the right to challenge the accuracy and pertinence of the cited documents. It will be clearly understood that, although a number of information sources, including scientific journal articles, patent

documents, and textbooks, are referred to herein; this reference does not constitute an admission that any of these documents forms part of the common general knowledge in the art.

[0323] The discussion of the general methods given herein is intended for illustrative purposes only. Other alternative methods and alternatives will be apparent to those of skill in the art upon review of this disclosure, and are to be included within the spirit and purview of this application.

[0324] Throughout this specification, various patents, patent applications and other types of publications (*e.g.*, journal articles, electronic database entries, *etc.*) are referenced. The disclosure of all patents, patent applications, and other publications cited herein are hereby incorporated by reference in their entirety for all purpose.

INFORMAL SEQUENCE LISTING

Name	Sequence	SEQ ID NO
Tn5 adapter top	CGTACCGATACGCTCGGACAGATGTGTATAAGAGACAG	1
Tn5 adapter bottom	CTGTCTCTTATACACATCT	2
Blocker 0 - EMX1_L	+T+C+CTCCAG+CTTCT+GCCGT+T+T+G/3AmMO	3
Primer 1	GCGTACCGATACGCTCGGAC	4
Primer 2	GGTTTGTGGTTGCCACCCCT	5
Primer 3	ACACTCTTCCCTACACGACGCTCTTCCGATCTCGTACCGATACGCTCGGAC	6
Primer 4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCCATTTGGCCTGCTTC	7
Primer 5	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTCCCTACACGACGCTCTTCCGATC*T	8
Primer 6	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	9
Tn5 adapter top U	CGTACCGAUACGCTCGGACUAGATGTGTATAAGAGACAG	10
Tn5 adapter bottom 3' blocked	CTGTCTCTTATACACATCT/3ddC/	11
Tn5 adapter top T7p	TAATACGACTCACTATAGCGTACCGATACGCTCGGACAGATGTGTATAAGAGACAG	12
Probe 0 - EMX1_L	TCCTCCAGCTTCTGCCGTTTG	13
Blocker 0 - EMX1_R	+C+G+AAGCA+GGCCA+ATGGG+G+A+G/3AmMO	14
Primer 2 - EMX1_R	TCTCATCTGTGCCCCCTCCCT	15
Primer 4 - EMX1_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGTGGTTCCAGAACCGGAGG	16
Probe 0 - EMX1_R	CGAAGCAGGCCAATGGGGAG	17
Blocker 0 - CCR5#1_L	+A+G+GGCTCC+GATGTA+TAATAA+T+T+G/3AmMO/	18
Primer 2 - CCR5#1_L	AGGTAGATGTCAGTCATGCT	19
Primer 4 - CCR5#1_L	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCTTCAGCCTTTTGCAGTTTATCAG	20
Probe 0 - CCR5#1_L	AGGGCTCCGATGTATAATAATTG	21
Blocker 0 - CCR5#1_R	+G+G+CAACAT+GCTGGT+CATCC+T+C+A/3AmMO/	22
Primer 2 - CCR5#1_R	ACAGGGTGGAACAAGATGGAT	23
Primer 4 - CCR5#1_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTACATCGGAGCCCTGCCAAA	24

Probe 0 - CCR5#1_R	GGCAACATGCTGGTCATCCTCA	25
Tn5 adapter top UMI	CGTACCGATACGCTCGGACNNNNNNNNNNNNNAGATG TGTATAAGAGACAG	26
Tn5 adapter top U UMI	CGTACCGAUACGCTCGGACUNNNNNNNNNNNNUAG ATGTGTATAAGAGACAG	27
Tn5 adapter top T7p UMI	TAATACGACTCACTATAGCGTACCGATACGCTCGGAC NNNNNNNNNNNNAGATGTGTATAAGAGACAG	28
EMX1 target sequence	GAGTCCGAGCAGAAGAAGAAGGG	29
CCR5#1 target sequence	GTGAGTAGAGCGGAGGCAGGAGG	30
EMX1_L	CAAACGGCAGAAAGCTGGAGGAGG	31
EMX1_R	CTCCCCATTGGCCTGCTTCGTGG	32
CCR5_L	TGACATCAATTATTATACATCGG	33
CCR5_R	TTTTGCAGTTTATCAGGATGAGG	34
i5_N501_UMI_Tn5 -A	AATGATACGGCGACCACCGAGATCTACACTAGATCG CNNNNNNNNNNNTCGTCGGCAGCGTCAGATGTGTATA AGAGACAG	35
i5_N502_UMI_Tn5 -A	AATGATACGGCGACCACCGAGATCTACACCTCTCTAT NNNNNNNNNNNNTCGTCGGCAGCGTCAGATGTGTATAA GAGACAG	36
i5_N503_UMI_Tn5 -A	AATGATACGGCGACCACCGAGATCTACACTATCCTCT NNNNNNNNNNNNTCGTCGGCAGCGTCAGATGTGTATAA GAGACAG	37
i5_N504_UMI_Tn5 -A	AATGATACGGCGACCACCGAGATCTACACAGAGTAG ANNNNNNNNNNNNTCGTCGGCAGCGTCAGATGTGTATA AGAGACAG	38
i5_N505_UMI_Tn5 -A	AATGATACGGCGACCACCGAGATCTACACGTAAGGA GNNNNNNNNNNNTCGTCGGCAGCGTCAGATGTGTATA AGAGACAG	39
i5_N506_UMI_Tn5 -A	AATGATACGGCGACCACCGAGATCTACACACTGCAT ANNNNNNNNNNNNTCGTCGGCAGCGTCAGATGTGTATA AGAGACAG	40
i5_N507_UMI_Tn5 -A	AATGATACGGCGACCACCGAGATCTACACAAGGAGT ANNNNNNNNNNNNTCGTCGGCAGCGTCAGATGTGTATA AGAGACAG	41
i5_N508_UMI_Tn5 -A	AATGATACGGCGACCACCGAGATCTACACCTAAGCC TNNNNNNNNNNNTCGTCGGCAGCGTCAGATGTGTATA AGAGACAG	42
i5_N501_UMI_Tn5 -A_Uracil	AATGATACGGCGACCACCGAGATC/ideoxyU/ACACTA GATCGCNNNNNNNNNN/ideoxyU/CGTCGGCAGCGTCA GATGTGTATAAGAGACAG	43
Tn5-B_3'blocked	CTGTCTCTTATACACATCT/3ddC/	44
Primer 1/3/5	AATGATACGGCGACCACCGAGATCTACAC	45
Blocker 0 - CCR5#1_L	+A+GG+GC+TC+CG+AT+GT+AT+AA+TA+AT+T+G/3Am MO/	46

Blocker 0 - CCR5#1_R	+T+GG+GC+AA+CA+TG+CT+GG+TC+AT+CC+T+C/3Am MO/	47
Primer 6_N701	CAAGCAGAAGACGGCATAACGAGATAGCGGAATGTGA CTGGAGTTCAGACGTGT	48
Primer 6_N702	CAAGCAGAAGACGGCATAACGAGATGATCATGCGTGA CTGGAGTTCAGACGTGT	49
Primer 6_N703	CAAGCAGAAGACGGCATAACGAGATAAGACGGAGTGA CTGGAGTTCAGACGTGT	50
Primer 6_N704	CAAGCAGAAGACGGCATAACGAGATCGAGTCCTGTGA CTGGAGTTCAGACGTGT	51
Primer 6_N705	CAAGCAGAAGACGGCATAACGAGATTCCTCAGGGTGA CTGGAGTTCAGACGTGT	52
Primer 6_N706	CAAGCAGAAGACGGCATAACGAGATGTACGGATGTGA CTGGAGTTCAGACGTGT	53
Primer 6_N707	CAAGCAGAAGACGGCATAACGAGATCATCTCTCGTGA CTGGAGTTCAGACGTGT	54
Primer 6_N710	CAAGCAGAAGACGGCATAACGAGATGTCGGAGCGTGA CTGGAGTTCAGACGTGT	55
Primer 6_N711	CAAGCAGAAGACGGCATAACGAGATACGGAGAAGTGA CTGGAGTTCAGACGTGT	56
Primer 6_N712	CAAGCAGAAGACGGCATAACGAGATAGGAGATGGTGA CTGGAGTTCAGACGTGT	57
Primer 6_N714	CAAGCAGAAGACGGCATAACGAGATAGTACTCGGTGA CTGGAGTTCAGACGTGT	58
Primer 6_N715	CAAGCAGAAGACGGCATAACGAGATGGACTCTAGTGA CTGGAGTTCAGACGTGT	59
CCR5#1C13	CTGATGTGTGGCAGTTTGGGAC	60
Predicted_OT_CC R2	CACCAGCGAGTAGAGCGGAGGCAGGAG	61
Homology_CCR2_CCR5	CTGGTGTTTCATCTTTGGTTTTGTGGGCAACATGCTGGT C	62
Predicted_OT_Ch r13	GGAGGAGAGCGGAGGCAGGTGG	63
VH01033:30:AAF5 TJWM5:1:1202:15 618:35494:rCTAC CGATCA	GCTCTTCAGCCTTTTGCAGTTTATCAGGATGAGGATG ACCAGCATGTTGCCACAAAACCAAAGATGAACACC TTTTTTGATTATGATTACGGTGCTCCCTGTCATAAATT TGACGTGAAGCAAATTGGGGCCCAACTCCTGCCTCCG CTC	64
VH01033:30:AAF5 TJWM5:1:2404:56 614:2950:rTCTAC CGGAA	GTGCTATAGAGCACAAAGATTTTATTTGGTGAGATGGT GCTCTACTCACTGGTGTTTCATCTTTGGTTTTGTGGGCA ACATGCTGGTCATCCTCATCCTGATAAACTGCAAAAG GCTGAAGAGC	65
VH01033:30:AAF5 TJWM5:1:1102:51 274:48462:rTTAC CTATCA	<u>CCCCATAGCAAGACAAAGACCTGTTTTAAACACATTT</u> <u>ACAACCTATATGTCTCCGCTCTACTCACTGGTGTTTCAT</u> <u>CTTTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCC</u> <u>TGATAAACTGCAAAAGGCTGAAGAGC</u>	66
VH01033:30:AAF5 TJWM5:1:1404:83	TACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCA AATCGCAGCCCGCTCCTCATAATCAAAAAAGGTGGT	67

28:41230:rCAGCA CCATG	<u>GACTTCTTCACCGCTCTCGTTGGTATTTCTGGTGTTCG TT</u>	
VH01033:30:AAF5 TJWM5:1:1301:13 725:35115:rGTCG ACCCAA	<u>TACATCGGAGCCCTGCCAAAAATCAATGTGAAGCA AATCGCAGCCCGCTCCCTGTCATAAATTTGACGTGA AGCAAATTGGGGCCCAACTCCTGCC</u>	68
VH01033:30:AAF5 TJWM5:1:1303:31 164:20443:rAGAT CCGGTT	<u>TACATCGGAGCCCTGCCAAAAATCAATGTGAAGCAAA TCGCAGCCCGCTCCTGACAATCGATAGGTACCTGGC TGTCGTCCATGCTGTGTTTGCTTTAAAAGCCAGGACG GTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGG TGG</u>	69
VH01033:30:AAF5 TJWM5:1:2104:51 122:6774:rCCGGG TGTCT	<u>CTCCCCATCCTGTTGACGATGCTCTGAAAATATGGTC CAGAAATGGTGTGAGCAAGGAGACAGCAAAGCAATG CTTGGAACATAGGTGCAGTGACTAGACATGAGGCGG GCTGCGATATGATTCACATTGATTTTTTGGCAGGGCT CCGAT</u>	70
VH01033:21:AAC CM73M5:1:2403:5 0308:18720:rACT AGCCAAG	<u>GATGAGGATGACCAGCATGTTGCCACAAAACCAAA GATGAACACCAGTGAGTAGAGCGGGTGGTCAGCCAA CAAGCTGCTGGGCGTGTGAGGCGGCTGTGCTGTGCCT GTGAGACAGGAATGGTTCCATTACAC</u>	71
VH01033:30:AAF5 TJWM5:1:2402:21 375:40681:rATCC CGGCTC	<u>TACATCGGAGCCCTGCCAAAAATCAATGTGAAGCA AATCGCAGCCCGCAGGTGGTCAGCCAACAAGCTGCT GGGCGTGTGAGGCGGCTGTGCTGTGCCTGTGAGACA GGAATGGTTCCATTACAC</u>	72
VH01033:21:AAC CM73M5:1:1103:2 1602:42423:rGCCT TGGGAG	<u>GAACAGTGATTGGCATCCAGTAGGTGCCCTCTAGGCC TCCGCTCTACTACTGGTGTTCATCTTTGGTTTTGTGG GCAACATGCTGGTCATCCTCATCCTGATAAACTGCAA AAGGCTGAAGAGC</u>	73
VH01033:21:AAC CM73M5:1:1104:1 4160:45016:rCTAC CTTCCT	<u>GCTCTTCAGCCTTTTGCAGTTTATCAGGATGAGGATG ACCAGCATGTTGCCACAAAACCAAAAGATGATTATG ATTACGGTGCTCCCTGTCATAAATTTGACGTGAAGCA AATTGGGGCCCAACTCCTGCCTCCGCTCTACTCCCTG GTGT</u>	74
VH01033:30:AAF5 TJWM5:1:1101:57 788:4048:rCTAAT AACCA	<u>TACATCGGAGCCCTGCCAAAAATCAATGTGAAGCA AATCGCAGCCCGCTCCCTGTCATAAATTTGACGTGA AGCAAATTGGGGCCCAACTCCTGCC</u>	75
VH01033:30:AAF5 TJWM5:1:1103:14 974:28678:rCGCC ACCCCC	<u>TACATCGGAGCCCTGCCAAAAATCAATGTGAAGCA AATCGCAGCCCGCTCCTGAAACTGGGTCAAGCATG GTGCAGCTCAGCCTTTTTG</u>	76
VH01033:30:AAF5 TJWM5:1:2204:19 027:32162:rCAGT AGCGGG	<u>TACATCGGAGCCCTGCCAAAAATCAATGTGAAGCA AATCGCAGCACTGCTGCATCAACCCCATCATCTATGC CTTTGTGCGGGGAGAAGTTCAGAAACTACCTCTTAGTC</u>	77
VH01033:21:AAC CM73M5:1:2103:5	<u>GATGAGGATGACCAGCATGTTGCCACAAAACCATA GATGAACACCAGTGAGTAGAGCGGAGAGAGGTGGTC</u>	78

6102:51699:rACC GGATCGC	<u>AGCCAACAAGCTGCTGGGCGTGTGAGGCGGCTGTGC</u> <u>TGTGCCTGTGAGACAGGAATGGTTCCATTACAC</u>	
VH01033:30:AAF5 TJWM5:1:1302:22 757:39640:rTGAA CCGGGG	TACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCA AATCGCAGCCCGCCTCCTAGGTGGTCAGCCAACAAG <u>CTGCTGGGCGTGTGAGGCGGCTGTGCTGTGCCTGTGA</u> <u>GACAGGAATGGTTCCATTACAC</u>	79
<u>VH01033:21:AAC</u> <u>CM73M5:1:1204:4</u> <u>0178:38655:rGGA</u> <u>TTCACCT</u>	<u>CTGTAGTCCCCAGCCACTTGGAGGGGTGAGGTGAGA</u> <u>GGATTGCTTGAGCCCGGGATGGTCCAGGCTCCGCTCT</u> ACTCACTGGTGTTCATCTTTGGTTATGTGGGCAA	80
VH01033:21:AAC CM73M5:1:2204:4 9532:17527:rCAA CTAAAAG	GCTCTTCAGCCTTTTGCAGTTTATCAGGATGAGGATG ACCAGCATGTTGCCACAAAACCAAAGATGAACACT <u>CAAAAGAATTGCCATGTTTTAAAAAAGAGAATTGGT</u> <u>GCCATAAGTTAGTTGTCTATGTTTGAAATGAAGAAG</u> <u>ATAT</u>	81
VH01033:21:AAC CM73M5:1:2403:3 4061:51586:rAAC AGCTTGT	GATGAGGATGACCAGCATGTTGCCACAAAACCAAA GATGAACACCAGTGAGTAGAGCGTGGTCAGCCAACA <u>AGCTGCTGGGCGTGTGAGGCGGCTGTGCTGTGCCTGT</u> <u>GAGACAGGAATGGTTCCATTACACCATGGGAAATCA</u> <u>C</u>	82
VH01033:21:AAC CM73M5:1:1303:4 1825:48898:rTGAT GGCGGG	<u>CCTGGGCGACAGAGTGAGACCCTGTCTCACAACAAC</u> <u>AACAACAACAAC</u> <u>TCCGCTCTACTCACTGGTGTTCATC</u> TTTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCT GATAAACTGCAAAAGGCTGAAGAGC	83
VH01033:21:AAC CM73M5:1:1202:2 7775:37879:rACG GCATCCT	ATGAGGATGACCAGCATGTTGCCACAAAACCAAAG ATGAACACCAGTGAGTAGAGCGGAGAGCAGACAAAC <u>AGACAAACAACAACAACACCTGAAGCCCCTGACTAA</u> <u>CAATCAATAGGCAACATCCAGGAAGATTGTAACCCC</u> <u>ATAGTAC</u>	83
VH01033:30:AAF5 TJWM5:1:2203:57 636:20822:rACCT CGGAAA	TACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCA AATCGCAGCCCGCCTCTGTGGGCTTGTGACACGGACT <u>CAAGTGGGCTGGTGACCCAGTCAGAGTTGTGCACAT</u> <u>GGCTTAGTTTTTCATACACAGCCTGGGCTGGGG</u>	85
<u>VH01033:30:AAF5</u> <u>TJWM5:1:1101:32</u> <u>982:28773:rCTAA</u> <u>TCAAAC</u>	<u>AGCCTGGGCGACAGAGTGAGACCCTGTCTCACAACA</u> <u>ACAACAACAACAACAAGGCTGAGCTGCACCATG</u> <u>CTTGACCCAGTTTCTTAACTGGAGGCGGGCTGCGATT</u> <u>TGCTTCACATTGATTTTTTGGCAGGGCTCCGATGTA</u>	86
VH01033:30:AAF5 TJWM5:1:2104:20 844:22204:rGAAC ATGCTT	TACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCA AATCGCAGCCCGCCTCCCTGTCATAAATTTGACGTGA <u>AGCAAATTGGGGCCCAACTCCTGCC</u>	87
VH01033:21:AAC CM73M5:1:1303:2 1394:11999:rGCG GCGGACG	<u>CAGCACAGCCGCCTCACACGCCCAGCAGCTTGTGGC</u> <u>TGACCTCCGCTCTACTCACTGGTGTTCATCTTTGGTTT</u> TGTGGGCAACATGCTGGTCATCCTCATCCTGATAAAC TGCAAAAGGCTGAAGAGC	88
VH01033:30:AAF5 TJWM5:1:2303:30	TACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCA AATCGCAGCCCGCAGGTGGTCAGCCAACAAGCTGCT	89

350:25763:rCTCC CGCCCC	<u>GGGCGTGTGAGGCGGCTGTGCTGTGCCTGTGAGACA</u> <u>GGAATGGTTCCATTACTC</u>	
UMI	NNNNNNNN	90
UMI	NWYYRRV	91

WHAT IS CLAIMED IS:

1. A method for detecting genome-wide re-arrangements in a nucleic acid genome, said method comprising:

(a) contacting a genomic DNA sample obtained from a cell or tissue contacted with a site-specific nuclease with a plurality of transposons comprising a first sequence tag at the 5' end of the transposon, under conditions whereby the plurality of transposons is inserted into the genomic DNA sample and the genomic DNA sample is tagmented into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment;

(b) performing a first amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using (i) a first target-specific primer comprising a nucleotide sequence complementary to a region adjacent to a target site, (ii) a second primer comprising a nucleotide sequence identical to a portion of the sequence tag located at the 5' ends of the fragment and (iii) a blocking oligonucleotide comprising a nucleotide sequence complementary to a region on the distal side of the target site relative to the first target-specific primer to produce primary amplification products comprising a nucleotide sequence comprising a re-arranged target sequence;

(c) sequencing the primary amplification products.

2. The method of claim 1, wherein the sequence tag comprises a sequence orthogonal to the genome, a unique molecular identifier (UMI), a transposase recognition site, an index sequence or a combination thereof.

3. The method of claim 1, wherein the blocking oligonucleotide comprises an absent or blocked 3' OH, spacers, inverted nucleotides, or other modifications to block extension of the 3' end.

4. The method of claim 1, wherein the blocking nucleotide comprises one or more phosphothorothioate bonds, spacers, or other modifications at the 3' and 5' ends, to block exonuclease digestion at the 3' and 5' ends, LNA, BNA, PNA, RNA, DNA, modified nucleic acids, or a combination thereof.

5. The method of claim 1, wherein the first and second primers comprise second and third sequence tags.

6. The method of claim 1 further comprising:
prior to (c) performing a second amplification reaction using third and fourth nested primers comprising nucleotide sequences complementary to sequences proximal to the inside ends of the amplified products of (b) and fourth and fifth sequence tags at the 5' ends of the nested primers to produce secondary amplification products comprising a re-arranged target sequence and additional sequence tags.

7. The method of claim 6 further comprising:
performing a third amplification reaction using a fifth and sixth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

8. The method of claim 6, wherein the third and/or the fourth primers comprise barcode sequences.

9. The method of claim 7, wherein the fifth and sixth primers comprise a sequencing tag and/or an index sequence.

10. The method of claim 1 further comprising:
prior to (c) performing a second, hemi-nested amplification reaction using the second primer and a third nested primer wherein the third primer comprises a nucleotide sequence complementary to a sequence proximal to target-specific primer end of the amplified products of (b) and optionally a fourth and fifth sequence tags at the 5' ends of the second and/or third primers to produce secondary amplification products comprising a re-arranged target sequence and one or two additional sequence tags.

11. The method of claim 10 further comprising:

performing a third amplification reaction using a fourth and fifth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

12. The method of claim 11, wherein the fourth and fifth primers comprise a sequencing tag and/or an index sequence.

13. The method of claim 1 wherein (b) further comprises performing a separate amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using a second target specific primer on the opposite side of the target site.

14. The method of claim 1 further comprising:

prior to (b) contacting the plurality of tagmented nucleic acid fragments with ddNTP or other 3' modified dNTP and a DNA polymerase or terminal deoxynucleotidyl transferase to block all extendable 3' ends.

15. The method of claim 2, wherein the sequence tag comprises uracil.

16. A method for detecting genome-wide re-arrangements in a nucleic acid genome, said method comprising:

(a) contacting a genomic DNA sample obtained from a cell or tissue contacted with a site-specific nuclease with a plurality of transposons comprising a first sequence tag at the 5' end of the transposon, under conditions whereby the plurality of transposons is inserted into the genomic DNA sample and the genomic DNA sample is tagmented into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment;

(b) contacting the plurality of tagmented nucleic acid fragments with a sequence specific cleavage reagent;

(c) performing a first amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using (i) a first target-specific primer comprising a nucleotide sequence

complementary to a region adjacent to a target site, (ii) a second primer comprising a nucleotide sequence identical to a portion of the sequence tag located at the 5' ends of the fragment and ;

(d) sequencing the primary amplification products.

17. The method of claim 16, wherein the sequence specific cleavage reagent is an enzymatic reagent, or a chemical cleavage agent.

18. The method of claim 17, wherein the enzymatic cleavage agent comprises a CRISPR Cas/gRNA complex.

19. The method of claim 16 further comprising:
contacting the plurality of tagged nucleic acid fragments with ddNTP or other 3' modified dNTP and a DNA polymerase or deoxynucleotidyl transferase after (b) to block all extendable 3' ends.

20. The method of claim 16, wherein the sequence tag comprises a sequence orthogonal to the genome, a unique molecular identifier (UMI), a transposase recognition site, an index sequence, or a combination thereof.

21. The method of claim 16, wherein the first and second primers comprise second and third sequence tags.

22. The method of claim 16 further comprising:
prior to (d) performing a second amplification reaction using third and fourth nested primers comprising nucleotide sequences complementary to sequences proximal to the inside ends of the amplified products of (c) and additional sequence tags at the 5' ends of the nested primers, to produce secondary amplification products comprising a re-arranged target sequence and additional sequence tags.

23. The method of claim 22 further comprising:
performing a third amplification reaction using a fifth and sixth primer comprising nucleotide sequences complementary to the amplified products of the second amplification

reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

24. The method of claim 22, wherein the third and/or the fourth primers comprise barcode sequences.

25. The method of claim 22, wherein the fifth and sixth primers comprise a sequencing tag and/or an index sequence.

26. The method of claim 16 further comprising:
prior to (d) performing a second, hemi-nested amplification reaction using the second primer and a third nested primer wherein the third primer comprises a nucleotide sequence complementary to a sequence proximal to target-specific primer end of the amplified products of (c) and optionally a fourth and fifth sequence tags at the 5' ends of the second and/or third primers to produce secondary amplification products comprising a re-arranged target sequence and one or two additional sequence tags.

27. The method of claim 26 further comprising:
performing a third amplification reaction using a fourth and fifth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

28. The method of claim 26, wherein the second and/or the third primers comprise barcode sequences.

29. The method of claim 27, wherein the fourth and fifth primers comprise a sequencing tag and/or an index sequence.

30. The method of claim 16 wherein (b) further comprises performing a separate amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using a second target specific primer on the opposite side of the target site.

31. A method for detecting genome-wide re-arrangements in a nucleic acid genome, said method comprising:

(a) contacting a genomic DNA sample obtained from a cell or tissue contacted with a site-specific nuclease with a plurality of transposons comprising a first sequence tag at the 5' end of the transposon, wherein the sequence tag comprises an RNA promoter sequence, under conditions whereby the plurality of transposons is inserted into the genomic DNA sample and the genomic DNA sample is tagmented into a plurality of nucleic acid fragments each comprising the sequence tag at the 5' ends of a partially double-stranded nucleic acid fragment;

(b) transcribing the plurality of tagmented nucleic acid fragments into RNA;

(c) contacting the RNA with target sequence specific DNA oligonucleotide probes and RNase H;

(d) performing a reverse transcription amplification reaction using (i) a first primer comprising a nucleotide sequence complementary to a portion of the sequence tag located at the 3' or 5' end of the fragment and (ii) a second primer comprising a nucleotide sequence complementary to a region adjacent to a target site to produce primary amplification products;

(e) sequencing the primary amplification products.

32. The method of claim 31, wherein the sequence tag comprises a sequence orthogonal to the genome, a unique molecular identifier (UMI), a transposase recognition site, an index sequence, or a combination thereof.

33. The method of claim 31, wherein the first and second primers comprise sequence tags.

34. The method of claim 31 further comprising:

prior to (e) performing a second amplification reaction using third and fourth nested primers comprising nucleotide sequences complementary to sequences proximal to the inside ends of the amplified products of (d) and additional sequence tags, to produce secondary amplification products comprising a re-arranged target sequence and additional sequence tags.

35. The method of claim 34 further comprising:

performing a third amplification reaction using a fifth and sixth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

36. The method of claim 34, wherein the third and/or the fourth primers comprise barcode sequences.

37. The method of claim 35, wherein the fifth and sixth primers comprise an adapter sequence and/or an index sequence.

38. The method of claim 31 further comprising:

prior to (e) performing a second, hemi-nested amplification reaction using the second primer and a third nested primer wherein the third primer comprises a nucleotide sequence complementary to a sequence proximal to target-specific primer end of the amplified products of (d) and optionally a fourth and fifth sequence tags at the 5' ends of the second and/or third primers to produce secondary amplification products comprising a re-arranged target sequence and one or two additional sequence tags.

39. The method of claim 38 further comprising:

performing a third amplification reaction using a fourth and fifth primer comprising nucleotide sequences complementary to the amplified products of the second amplification reaction to produce further enriched tertiary amplification products comprising a re-arranged target sequence.

40. The method of claim 38, wherein the second and/or the third primers comprise barcode sequences.

41. The method of claim 39, wherein the fourth and fifth primers comprise an adapter sequence and/or an index sequence.

42. The method of claim 31, wherein (d) further comprises performing a separate amplification reaction to amplify the re-arranged tagmented nucleic acid fragments using a second target specific primer on the opposite side of the target site.