

Methods for Detection of Single Strand Breaks in Nucleic Acid

Field

The present invention relates to methods for detecting and mapping sites of single strand breaks (SSBs) in double stranded nucleic acid, such as double-stranded deoxyribonucleic acid (dsDNA).

5

Background

Shortly after the elucidation of DNA structure, it became apparent that DNA is subject to endogenous gaps (Lindahl and Nyberg, 1972) resulting from a variety of processes that include oxidation, hydrolysis, alkylation, mismatch of DNA bases and in mediating changes in cell responses. Estimates suggest each cell undergoes
10 as many as 70,000 gaps per day (Lindahl and Barnes, 2000). Though such DNA gaps have historically been implicated in mutations and various diseases including cancer (Vilenchik et al., 2003), recently, there is a growing appreciation for the role of endogenous DNA gaps in normal cell function (Wu et al., 2021).

Accordingly, several technologies have been developed that aim to measure the genomic distribution and
15 frequency of DNA gaps. The majority of these techniques can measure, with reasonable accuracy, the genomic landscape of double strand DNA gaps (Yan et al., 2017, Amente et al., 2021). Despite the prevalence of single strand DNA gaps (Lindahl and Barnes, 2000), there are fewer than a handful of methods that could afford genome-wide detection of single strand DNA gaps. These methods include SSiNGLe (single strand break mapping at nucleotide genome level; Cao et al (2019), SSB-Seq (Baranello et
20 al (2014); Baranello et al (2018)) XR-Seq (Hu et al (2015) and GLOE-Seq (Sriramachandran et al (2020)). Invariably, these methods rely on capturing 3' OH ends of DNA gaps to map their location, and since 3' OH ends are also present at sites of double stranded DNA gaps, the resulting DNA gap measurement may not represent single strand DNA gaps but rather may profile nicks in DNA.

25 There remains a need for simple techniques for the rapid high-resolution mapping of SSBs.

Summary

The present inventors have developed a method (referred to herein as "DNA GAPS Sequencing" or "GAPS-seq") that allows the rapid high-resolution detection of single strand breaks (SSBs) in double stranded (ds)
30 DNA and may, for example, be useful in mapping or assessing SSBs in genomic DNA in a strand-specific manner.

A first aspect of the invention provides a method of detecting sites of single strand breaks (SSBs) in double stranded (ds) DNA comprising;

- 35 (i) contacting a sample of dsDNA with an oligonucleotide probe comprising a first sequencing adaptor and a tag,
(ii) contacting the sample of dsDNA with DNA ligase, such that the DNA ligase covalently links the oligonucleotide probe to a strand of the dsDNA at the site of a single strand break in the dsDNA,
(iii) contacting the sample of dsDNA with a transposase loaded with a second sequencing adaptor,
40 such that the transposase cleaves the dsDNA to generate a population of DNA fragments comprising the second sequencing adaptor,

(iv) isolating from said population DNA strands that comprise the first and second sequencing adaptors
and;

(v) determining the sequence of the isolated DNA strands.

5

The sequences of the isolated strands may be indicative of the sites of SSBs within the dsDNA.

Methods of the first aspect may comprise determining the sites of SSBs and/or the frequencies of SSBs at sites within a first and a second sample of dsDNA. Differences in SSB sites or differences in the frequencies of SSBs at sites between the first and second samples may be determined.

10

One of the first and second samples may have been subjected to a treatment or may be obtained from a disease cell, or a cell of a defined age or stage of development. The other sample may be a reference or control. The effect of the treatment, disease, developmental stage or age on the location or frequency of SSBs may be determined.

15

A second aspect of the invention provides a kit for detecting single strand break (SSB) sites in double stranded (ds) DNA; the kit comprising;

an oligonucleotide probe comprising a first sequencing adaptor and a tag,
a DNA ligase,
a transposase and
a second sequencing adaptor loaded or loadable onto the transposase.

20

The kit may further comprise a binding member that specifically binds to the tag; sequencing primers, solid support, such as beads, PCR reagents and/or other reagents.

25

Other aspects and embodiments of the invention are described in more detail below.

Brief Description of the Figures

Figure 1 shows examples of methods for strand-specific measurement of DNA gaps genome-wide (GAPS-Seq). Nuclei from cells or tissues are fixed in situ and then ligated with a biotinylated DNA adaptor attached to a hexamer (of a random sequence). Nuclei are then lysed and genomic DNA is fragmented and attached to a second (R2) adaptor, before capturing of biotinylated fragments with streptavidin and strand separation. These are then PCR amplified, sequenced and mapped, to reveal genomic loci bearing DNA gaps.

30

Figure 2 shows measurement of ssDNA gaps using GAPS-seq. GAPS-seq reads in HEK293 cells, are shown normalised to library size then averaged over 40bp tiles. Plotted are the mean signal over all protein coding genes in 40bp tiles relative to the transcription start site. The line represents the mean of 4 replicates, shading is the standard deviation.

35

40

Figure 3 shows an example of ss DNA gaps profile in HEK293 cells treated and untreated with etoposide. Etoposide treatment is known to increase endogenous DNA gaps by inhibiting topoisomerase. As predicted, GAPS-seq detected more ssDNA gaps in etoposide treated vs DMSO only treated controls. Interestingly, compared to surrounding regions, GAPS-seq also shows more ssDNA gaps in enhancer sequences, suggesting regulatory regions undergo dynamic cycles of ssDNA gap and repair.

Figure 4 shows tapestation traces of DNA GAP-seq libraries obtained from different input samples. Input ranging from 1 million to 100 cells.

Figure 5 shows the validation of GAPS-seq sensitivity and specificity with restriction enzymes. Shown are GAPS-seq reads in HEK293 cells treated with restriction enzymes prior to GAPS-seq. Reads are normalised to library size then averaged over 40bp tiles. Plotted are the mean signal over all restriction enzyme sites in the human genome in 40bp tiles relative to the middle position of the site. Each panel shows a different restriction enzyme site and each coloured line shows a different sample - each sample was treated with a different enzyme.

Figure 6 shows the validation of GAPS-seq specificity to single-stranded gaps and not nicks in the DNA. Shown are GAPS-seq reads in BJF cells treated with a nickase enzyme and exonuclease III prior to GAPS-seq. Reads are normalized to library size then averaged over 40bp tiles. Plotted are the mean signal over all nickase motif instances in the human genome in 40bp tiles relative to the middle position of the site. The left panel shows control cells which were not treated with either enzyme. Middle panel shows cells treated only with the nickase. Right panel shows cells treated with both nickase and exonuclease.

Figure 7 shows measurement of DNA gaps using GAPS-seq at promoters of genes with different expression levels. GAPS-seq reads in HEK293 cells, are normalised to library size then averaged over 400bp tiles. The line represents the mean of 3 replicates. This result demonstrates that GAPS-seq can detect the known relationship between DNA breaks and transcription.

Figure 8 shows a dendrogram of hierarchical clustering of different samples profiled using GAPS-seq and quantified at gene promoters ($\pm$ 2kb from TSS). NPC are neural progenitor cells derived from in vitro differentiation of human embryonic stem cells. hESCs are human embryonic stem cells. HEK are HEK293 cells. BJF are primary human foreskin fibroblasts.

Figure 9 shows a scatter plot of the first two principal components derived from GAPS-seq or sBLISS performed on HEK293 cells and quantified over gene promoters. This plot clearly shows the separation of samples by the method used to produce the library and therefore demonstrates that GAPS-seq is measuring a distinct aspect of the genome from double-strand break profiling methods such as sBLISS.

35

Detailed Description

This invention relates to methods of detecting SSBs in ds nucleic acids, such as dsDNA. A sample of dsDNA is contacted with an oligonucleotide probe comprising a first sequencing adaptor and a tag and then contacted with a DNA ligase. The oligonucleotide probe is covalently linked by the DNA ligase to a strand of the dsDNA at the site of a single strand break or gap in the dsDNA. The dsDNA is then contacted with a transposase loaded with a second sequencing adaptor. The transposase cleaves the dsDNA and generates a population of fragments that have the second sequencing adaptor at one or both ends. DNA strands from this population of fragments that comprise the second sequencing adaptor at one end and the first sequencing adaptor at the other end are then isolated and the sequence of the isolated DNA strands determined, for example by sequencing. The sequences of the isolated strands may be indicative of the sites of single strand breaks (SSBs) within the dsDNA. Strand-specific SSB sites within the dsDNA in the sample may be identified or mapped from the sequences of the isolated strands.

Methods described herein may be useful for example in distinguishing SSBs from other genetic lesions, determining the sites of SSBs and/or the frequency of SSBs at sites in a sample of dsDNA, such as genomic DNA, and allowing the high-resolution, strand-specific mapping of SSBs.

Methods described herein may also be useful for detecting SSBs in low-input samples from cells to tissue sections. Because cells or tissues do not need to be labelled before fixation, the methods described herein directly detect nascent gaps. Furthermore, the fixation of samples allows the samples to be stored and the methods described herein performed long after initial sample collection, without affecting the numbers of DNA SSBs or gaps detected. The sensitivity of the methods described herein allow the evolution of DNA SSBs or gaps to be tracked with treatment (e.g. chemotherapy), developmental stage and/or ageing.

The methods described herein may be used on dsDNA in solution, cell-free dsDNA or any cell or tissue that can be obtained in suspension and allow the rapid generation of sequencing libraries and provision of DNA SSB or gap information, making the methods suitable for clinical applications. The methods described herein may also be used on a sample of dsDNA obtained from, e.g., a cell, organoid or tissue, or e.g., cell-free DNA.

The methods described herein are also compatible with a bisulfite conversion and allow the measurement of DNA methylation at sites adjacent to gaps. This allows the presence or absence of gaps and DNA methylation to be assessed and the spatial relationship to be assessed between gap location and DNA methylation.

SSBs are a common form of DNA damage or genetic lesion. An SSB (interchangeably termed a single strand gap) is a discontinuity in one strand of a dsDNA (the "broken strand"), whilst the other strand (the "intact strand") is undamaged. One or more phosphodiester bonds is missing in the broken strand at the SSB, causing a discontinuity or break in the chain of nucleotides that form the strand. One or more nucleotides may be missing from the broken strand at the SSB site, causing a gap in the nucleotide sequence of the broken strand relative to the intact strand.

dsDNA may include any dsDNA, including, for example genomic DNA, plasmid DNA or viral DNA. Preferably, the dsDNA is genomic DNA. For example, the dsDNA may be part of a chromosome, a mini-chromosome, a whole chromosome or more than one chromosome or from a circulating DNA. Suitable
5 dsDNA may also include mitochondrial, plasmid or viral DNA inserted into a chromosome or genome.

Suitable dsDNA may include prokaryotic dsDNA, such as bacterial dsDNA. For example, dsDNA for use as described herein may include part or all of the genome of a prokaryotic cell, such as a bacterial cell, more preferably the whole genome of a prokaryotic cell.

10

Preferably, dsDNA is eukaryotic dsDNA, for example mammalian dsDNA, such as human DNA. For example, dsDNA for use as described herein may include part or all of the nuclear and/or mitochondrial genome of a eukaryotic cell, more preferably the whole genome of a eukaryotic cell.

15 A sample of dsDNA for use as described herein may contain 1 or more, 10 or more 100 or more or 1000 or more cell genomes. A sample of dsDNA for use as described herein may contain 3pg, 6pg or more, 60pg or more, 600pg or more, or 6ng or more of DNA.

20 In some embodiments, a sample of dsDNA may be obtained from a tissue, organoid, cell, cell extract, cell fraction, for example a cell organelle, such as a nucleus or mitochondrion, or a biological fluid, such as blood, sperm, cerebrospinal fluid, amniotic fluid, pleural fluid, peritoneal fluid or saliva. The dsDNA may be cellular DNA or cell-free DNA (cfDNA).

25 In some embodiments, the dsDNA may be obtained from a frozen cell or tissue sample.

In some embodiments, the dsDNA may be within a cell, a cell nucleus or a nuclear extract, such as isolated genomic DNA (gDNA). A cell nucleus or nuclear extract may be obtained from a cell using standard techniques. In some preferred embodiments, the dsDNA may be within a cell nucleus. A sample comprising one or more cell nuclei may be contacted with an oligonucleotide probe as described herein, such that the
30 probe contacts dsDNA in the nuclei. This may allow the sites of SSBs within the genome of a cell to be mapped. For example, a method of mapping sites of SSBs within a cell genome may comprise;

- (i) contacting a cell nucleus with an oligonucleotide probe comprising a first sequencing adaptor and a tag,
- (ii) contacting the cell nucleus with DNA ligase, such that the DNA ligase covalently links the
35 oligonucleotide probe to a strand of the dsDNA at the site of a single strand break in the dsDNA,
- (iii) contacting the cell nucleus with a transposase loaded with a second sequencing adaptor, such that the transposase cleaves the dsDNA to generate a population of fragments comprising a terminal second sequencing adaptor,
- (iv) isolating from said population DNA stands that comprise the first and second sequencing
40 adaptors and;
- (v) determining the sequence of the isolated DNA strands.

The sequences of the isolated DNA strands may be indicative of the sites of SSBs within the cell genome.

In some embodiments, dsDNA for use as described herein may be obtained from a eukaryotic cell.

5 Eukaryotic cells may be isolated, for example as immortalised cell lines or primary cells obtained from an individual (e.g. a human subject) or may be in the form of tissues or organoids. For example, a cell nucleus may be obtained from eukaryotic cells within a sample obtained from an individual, such as a biopsy or xenograft sample.

10 Suitable eukaryotic cells may include mammalian cells, preferably human cells. For example, eukaryotic cells may include somatic and germ-line cells and may be at any stage of development, including fully or partially differentiated cells or non-differentiated or pluripotent cells, including stem cells, such as adult or somatic stem cells, foetal stem cells or embryonic stem cells. Suitable eukaryotic cells also include induced pluripotent stem cells (iPSCs), which may be derived from any type of somatic cell in accordance with
15 standard techniques. Eukaryotic cells may also include neural cells, including neurons and glial cells; contractile muscle cells; smooth muscle cells; liver cells; hormone synthesising cells; sebaceous cells; pancreatic islet cells; adrenal cortex cells; fibroblasts; mesenchymal cells; epithelial cells; keratinocytes; endothelial cells; urothelial cells; osteocytes; chondrocytes; immune cells; such as leukocytes; mesothelial cells and adipocytes; bone marrow cells or samples.

20 Suitable eukaryotic cells also include normal cells and disease cells, for example cells associated with disease conditions, including cancer cells, such as carcinoma, sarcoma, lymphoma, blastoma or germ-line tumour cells, and cells with the genotype of a genetic disorder, such as Huntington's disease, cystic fibrosis, sickle cell disease, phenylketonuria, Down syndrome, or Marfan syndrome.

25 Suitable eukaryotic cells also include embryonic cells, extra-embryonic cells, and cell culture models.

In some embodiments, dsDNA for use as described herein may from a disease cell, for example a cell obtained from an individual with a disease, such as a cancer cell, e.g. such a cell obtained from a human
30 subject. The frequency or distribution of SSBs within the dsDNA may be indicative of the diagnosis, severity, penetrance grade or prognosis of a disease, such as cancer, in the individual from whom the dsDNA is obtained. The disease or cancer may for example be caused or characterised by a defect in a DNA repair pathway.

35 In other embodiments, the dsDNA may be from a cell at a specific developmental stage, for example a cell obtained from an embryo; an age-associated cell, for example a cell obtained from an individual of a specific age, such as an elderly person, or a treated cell. Treated cells may include cells treated with reprogramming factors, such as Oct3/4, Sox2, Klf4 and c-Myc (the "Yamanaka factors" or "OSKM"), radiation, chemical compounds, and drugs, such as chemotherapy agents.

40

For example, methods described herein may be useful in cytotoxicity assays. A cytotoxicity assay may comprise subjecting a cell to a stimulus and determining the frequency or distribution of gaps or SSBs in dsDNA from the cell. The frequency or distribution of gaps or SSBs in dsDNA from the cell may be indicative of the cytotoxicity of the stimulus. Suitable stimuli include exposure to chemical compounds and radiation.

5

In other embodiments, the dsDNA may be from a cell treated with a nuclease such as a CRISPR nuclease (e.g. Cas9 or Cpf1). The frequency or distribution of SSBs within the dsDNA may be indicative of off-target effects of the nuclease in the treated cell.

10 In other embodiments, the dsDNA may be from an egg or sperm cell obtained from a donor e.g. a human donor. The frequency or distribution of SSBs within the dsDNA may be indicative of the quality of the donor egg or sperm cell. This may be useful for example in assisted reproductive technology (ART) methods.

In some embodiments, a cell may be fixed and permeabilised before step (i) of a method described herein.

15

Fixation may be useful for example, in reducing background by mitigating against SSBs caused in the handling of the sample prior to adding the ligase. Suitable methods for fixing cells are well known in the art and include contacting the cells with an aldehyde fixative such as formaldehyde, formalin, or glutaraldehyde; or an alcohol fixative, such as methanol, ethanol, or acetone. For example, cells may be fixed by exposure to 2% formaldehyde.

20

Permeabilization may be useful for example, in allowing the oligonucleotide probe and DNA ligase to contact dsDNA in a cell or in extracting the nucleus or dsDNA from a cell. Suitable methods for permeabilising cells are well known in the art and include contacting the cells with a detergent, such as 2-[4-(2,4,4-trimethylpentan-2-yl)phenoxy]ethanol (e.g. triton X-100™), nonyl phenoxypolyethoxyethanol (e.g. NP-40™), polyoxyethylene sorbitan monolaurate (e.g. Tween™), saponin, or digitonin. For example, cells may be permeabilised by exposure to 2-[4-(2,4,4-trimethylpentan-2-yl)phenoxy]ethanol, for example 0.2% 2-[4-(2,4,4-trimethylpentan-2-yl)phenoxy]ethanol.

25

30 A cell nucleus may be isolated or extracted from the fixed and permeabilised cells before step (i). For example, a method described may comprise providing a cell sample and extracting a cell nucleus from the cell sample. The cell nucleus or genomic DNA therein may then be used in the present methods.

In some embodiments, a sample of dsDNA may be in solution in methods described herein. For example, dsDNA or cells containing nucleic acid may be contacted with the probe, DNA ligase and/or transposase in solution. The dsDNA, nucleic acid containing dsDNA or cells containing dsDNA may be washed, for example by centrifugation and resuspension between steps.

35

In other embodiments, a sample of dsDNA may be immobilised on a solid support in methods described herein. A solid support is an insoluble, non-gelatinous body which presents a surface on which a capture molecule can be immobilised for capture of dsDNA or a cell containing dsDNA. Examples of suitable

40

supports include glass slides, microwells, membranes, or microbeads. The support may be in particulate or solid form, including for example a plate, a test tube, bead, a ball, filter, fabric, polymer or a membrane. Capture molecules may bind to proteins, glycoproteins or other molecules on the surface of a cell. Suitable capture molecules for cells are well-known in the art and include lectins that bind to extracellular
5 glycoproteins on the cell, such as concanavalin A.

The sequence of an DNA strand isolated from the population of generated fragments may be indicative of the sequence of the broken strand at an SSB site in the dsDNA. Multiple DNA strands may be isolated from the population. The sequences of the multiple isolated DNA strands may be indicative of the sequence of the
10 broken strand at multiple SSB sites in the dsDNA.

In methods described herein, the sample of nucleic acid, e.g. dsDNA is contacted with an oligonucleotide probe. The oligonucleotide probe comprises a first sequencing adaptor and a tag.

15 Sequencing adaptors are double-stranded oligonucleotides that are attached to nucleic acid molecules to allow sequencing, for example by facilitating the amplification of a nucleic acid molecule using sequencing primers. Suitable sequence adaptors for any method of sequencing are available in the art. In some embodiments, a sequencing adaptor may comprise a primer recognition sequence. A primer recognition sequence is a nucleotide sequence that is complementary to the sequence of an amplification primer. The
20 presence of primer recognition sequences allows DNA strands that are tagged with the adaptor to be amplified, for example by PCR using amplification primers that target the primer recognition sequences. The primer recognition sequences may be heterologous sequences that are not naturally present in the dsDNA. Suitable primer recognition sequences are well-known in the art and include Illumina-compatible barcoded i7/i5 primers. The first sequencing adaptor may comprise a first primer recognition sequence.

25 In some embodiments, the first sequencing adaptor may further comprise a barcode. A barcode is a nucleotide sequence that is unique for the sample from which the population of nucleic acids is obtained. A suitable barcode sequence may be 6-10 nucleotides. The barcode allows sequence reads from a specific sample to be unambiguously identified in a pooled multiplex sequencing reaction. Each sample may have a
30 unique barcode, so that all of the nucleic acids from the same sample receive the same barcode. Once prepared, populations of nucleic acids from different samples may be mixed into a single pool and sequenced. The sample from which a sequence read from the pool originates may then be identified from the barcode. For example, a suitable barcode for the multiplex sequencing of 24 samples (a 24-plex reaction) may consist of at least 6 nucleotides, preferably 6 nucleotides (Craig DW et al. 2008. Nat Methods
35 5, 887; Cronn R et al. 2008. Nucleic Acids Res, 36, e122). The use of barcodes in sequencing reactions is well-known in the art. For example, the Illumina sequencing 8-mer barcodes i5 and i7 may be used. In other embodiments, one or both sequencing primers may comprise a barcode.

The tag allows the isolation of the DNA strand to which it is attached. Suitable tags include any label,
40 molecule or group which allows the specific binding of a binding member to the generated fragment or strand to which it is attached. The tag may allow covalent or more preferably non-covalent binding of the binding

member. Suitable tags include immunogens, such as digoxigenin; short peptides, such as glutathione and FLAGTM; or small organic compounds such as biotin and trimethoprim (TMP).

Preferably, the tag is biotin, and the oligonucleotide probe is a biotinylated oligonucleotide.

5

The oligonucleotide probe may further comprise a hybridisation region. The hybridisation region may comprise 2-10 nucleotides, preferably 4-8 nucleotides, more preferably 6 nucleotides, that anneals to the intact strand of the dsDNA at a SSB site.

10 In some embodiments, the hybridisation region may be non-specific and may bind to the intact strand at any SSB site. For example, the hybridisation region may have a random sequence. The nucleotide at each position in the hybridisation region may be random ("N"; i.e. a nucleotide randomly selected from G, C, T or A), such that the hybridisation region of the oligonucleotide probe is diverse.

15 In other embodiments, the hybridisation region may be specific and may bind to the intact strand at a particular SSB site. For example, the hybridisation region may be complementary to the sequence of the intact strand at a target site in the dsDNA, for example a site at which SSBs are to be assessed or detected. The target sequence may be a site of interest; or a known site of SSBs, such as a promoter.

20 Examples of oligonucleotide probes are found in Table 4, which provides a suitable ssBreak_3prime sequence and a ssBreak_5prime sequence.

The oligonucleotide probe is ligated to the strand of the dsDNA that comprises the gap or SSB (i.e. the broken strand) by the DNA ligase. Suitable DNA ligases are well known in the art and available from
25 commercial suppliers. Examples of DNA ligases that are suitable for use in the methods described herein include: T4 DNA ligase, T7 DNA ligase, Ligase 3, and Ligase 1. In an embodiment, methods described herein are carried out using T4 DNA ligase.

The dsDNA may be contacted with the oligonucleotide probe and the DNA ligase at the same time or more
30 preferably sequentially, with the dsDNA being contacted with the oligonucleotide probe and then contacted with the DNA ligase.

The oligonucleotide probe may be contacted with dsDNA, such that the oligonucleotide probe anneals to the intact strand of the dsDNA at one or more gaps or SSB sites. The DNA ligase may then be contacted with
35 the dsDNA, such that it covalently attaches the oligonucleotide probe to an end of the broken strand at a gap or SSB site in the dsDNA.

In some embodiments, the oligonucleotide probe is ligated to the 3' end of the broken strand of the dsDNA at the gap or SSB site. A suitable oligonucleotide probe for attachment to a 3' end may comprise a 5'
40 phosphate group and a 3' tag. Following tagmentation with the transposase, the second sequencing adapter

may be located at the 5' end of the isolated DNA strands and the tag may be located at the 3' end of the isolated DNA strands.

5 In other embodiments, the oligonucleotide probe is ligated to the 5' end of a strand of the broken strand of the dsDNA at the gap or SSB site. A suitable oligonucleotide probe for attachment to a 5' end may comprise a 5' tag and a 3' hydroxyl group. Following tagmentation with the transposase, the second sequencing adaptor may be located at the 3' end of the isolated DNA strands and the tag is located at the 5' end of the isolated DNA strands.

10 Following ligation of the oligonucleotide probe to the broken strand at an SSB site, the dsDNA is fragmented with a transposase to produce a population of dsDNA fragments with tagged with a terminal second sequencing adapter.

15 In some embodiments, the transposase may be an activatable transposase. An activatable transposase may be converted from an inactive state to an active state by modification of the conditions. For example, a magnesium-activated transposase may be converted into an active state by increasing the concentration of magnesium ions (Mg^{2+}) or manganese ions (Mn^{2+}), for example to a concentration within range of about 0.1 mM to about 10 mM.

20 Suitable activatable transposases are well known in the art and include Tn5, Tn7, Mu, IS5 and IS91 (for example US9005935B2, Mizuuchi, K., Cell, 35: 785, 1983; Savilahti, H, et al, EMBO J., 14: 4893, 1995, Goryshin and Reznikoff, J. Biol. Chem, 273:7367 (1998) and WO2022056309A1). In some preferred embodiments, the transposase may be Tn5 transposase. Tn5 transposase may be obtained from commercial suppliers (e.g. Tagmentase™, Diagenode™).

25 The transposase may be loaded with one or more second sequencing adapters. The transposase may be pre-loaded with the second sequencing primers or the method may comprise loading the transposase with the second sequencing primers. The second sequencing adapters may be non-covalently bound to the transposase to form a complex comprising the transposase and the one or more second sequencing
30 adapters. Suitable methods for loading transposase with second sequencing adapters are well-known in the art. For example, the second sequencing adapters may be incubated with the transposase at room temperature for 1 hour.

Sequencing adapters are described above. Second sequencing adaptors suitable for attachment to a
35 transposase may be double stranded oligonucleotides that comprise a transposase recognition sequence and a primer recognition sequence. A transposase recognition sequence is a sequence that is targeted for transposition by a transposase. For example, inverted transposase recognition sequences may bracket the transposon sequence that is inserted by a transposase. The sequence of the transposase recognition sequence may depend on the transposase. Suitable transposase recognition sequences for different
40 transposases are well known in the art. For example, suitable transposase recognition sequences for Tn5 may include 19-mer end sequences, such as outside end (OE) sequences, such as 5'-CTG ACT CTT ATA

CAC AAG T - 3' and inside end (IE) sequences, such as 5'-CTG TCT CTT GAT CAG ATC T - 3'; and mosaic end (ME) sequences, such as 5'-CTG TCT CTT ATA CAC ATC T - 3'. In some preferred embodiments, 19-mer ME sequences may be employed. The second sequencing adaptor may comprise a transposase recognition site and a second primer recognition sequence. Suitable sequencing adaptors are well known in the art (e.g. N7 Nextera™ adaptor).

Example of a suitable second sequence adaptor is shown in Table 4 which details N7_top and N7_bottom sequences.

10 The transposase cleaves the sample of dsDNA to generate a population of DNA fragments. The DNA fragments in the population are tagged at one or both ends with the second sequencing adaptor. DNA fragments in the population that do not include the site of a gap or SSB are tagged at both ends with the second sequencing adaptor. DNA strands from the population of fragments that include the site of a gap or SSB are tagged at one end with the second sequencing adaptor and at the other end with oligonucleotide probe which comprises the first sequencing adaptor.

DNA strands tagged with both the first and second sequencing adaptors may be isolated from the population using the tag. The fragments may be denatured using any suitable technique, for example NaOH treatment, to produce DNA strands. DNA strands tagged with both the first and second sequencing adaptors may then be isolated using a binding member that binds to the tag of the oligonucleotide probe.

A binding member is a molecule that binds specifically to a target molecule or ligand, such as a tag. The binding member may bind covalently or more preferably non-covalently to the tag of the oligonucleotide probe. For example, when the tag is biotin; the binding member may specifically bind to biotin. Suitable binding members for binding to a biotin tag may include anti-biotin antibodies, avidin and streptavidin. A binding member that specifically binds to a tag may not show any significant binding to molecules other than the tag. In particular, the binding member may show no significant binding to proteins, DNA or other antigens that may be present in a cell or cell extract. Generally, an antibody or other binding member which specifically binds to a tag may have a binding affinity (K_a) greater than about 10^5 moles/liter (e.g. 10^6 or more, 10^7 or more, 10^8 or more, 10^9 or more, 10^{10} or more, 10^{11} or more, or 10^{12} or more moles/liter). Suitable binding members are well-known in the art and may be produced using standard techniques or obtained from commercial sources.

In preferred embodiments, the tag is biotin and the binding member is streptavidin. In some embodiments, the binding member may be immobilised on a solid support, such as a bead.

Washing steps may be performed as required to remove unbound reagents during any step of the methods described herein. For example, an oligonucleotide probe that is not bound to the dsDNA after the contacting step may be removed by washing. Similarly, DNA ligase and/or transposase may also be removed by washing. Suitable methods of washing are well known in the art. For example, washing may be performed using a buffer, such as 20 mM HEPES pH7.5, 150 mM NaCl and 0.5 mM spermidine, as described herein.

- Isolated DNA strands produced as described above may be amplified to produce amplification products for sequencing. The isolated DNA strands may be amplified using primers that hybridise to the primer recognition sequences of the sequencing adaptors. Suitable amplification methods are well established in the art and include polymerase chain reaction (PCR) (reviewed for instance in "PCR protocols; A Guide to Methods and Applications", Eds. Innis et al, 1990, Academic Press, New York, Mullis et al, Cold Spring Harbor Symp. Quant. Biol., 51:263, (1987), Ehrlich (ed), PCR technology, Stockton Press, NY, 1989, and Ehrlich et al, Science, 252:1643-1650, (1991)).
- Amplification may generate a population or library of amplification products for sequencing. Each amplification product in the population or library may contain the nucleic acid sequence at the site of an SSB in the dsDNA. A population or library of amplification products generated as described herein may be purified before sequencing using standard techniques, including spin-column chromatography (e.g. Ampure XP™ beads).
- The sequence of the isolated strands or amplification products may be determined by any suitable technique. Suitable techniques include sequencing and hybridisation-based techniques, preferably sequence-specific amplification, such as qPCR.
- In preferred embodiments, the sequence of the isolated strands or amplification products may be determined by sequencing the strands or products.
- The isolated DNA strands or amplification products may be sequenced using standard sequencing techniques. Suitable techniques include using any convenient low or high throughput sequencing technique or platform, including Sanger sequencing, Solexa-Illumina sequencing, ligation-based sequencing (SOLiD™), pyrosequencing; single molecule real-time sequencing (SMRT™); PacBioscience sequencing; and semiconductor array sequencing (Ion Torrent™). Preferably, sequencing is performed by a next-generation sequencing technique. Suitable protocols, reagents and apparatus for nucleic acid sequencing are well-known in the art and are available commercially.
- The sequences of the isolated DNA strands or amplification products may be indicative of the nucleotide sequences of SSB sites in the dsDNA. For example, the sequences of the isolated DNA strands or amplification products may include the nucleotide sequence at the site in a strand at which the SSB occurred. The position of SSBs within the dsDNA may be identified or mapped from the sequences of the isolated DNA strands or amplification products. For example, the position of SSBs within the genome of a cell, such as a eukaryotic, mammalian or human cell, may be identified or mapped by a method described herein. The sequences of the isolated strands or amplification products may be indicative of frequency at which an SSB occurs at a site in the dsDNA, or the probability that a SSB will occur at a site.
- In some embodiments, a method described herein may comprise mapping SSB sites within a first dsDNA and a second dsDNA and identifying the SSB sites that are present in the first dsDNA and not in the second

dsDNA or present in the second dsDNA and not in the first dsDNA. A method described herein may comprise mapping the frequency of SSBs at sites within a first dsDNA and a second dsDNA and identifying sites with increased frequency of SSB occurrence in the first dsDNA relative to the second dsDNA or in the second dsDNA relative to the first dsDNA.

5

The first dsDNA may be a test sample and the second dsDNA may be a reference or control sample. In some embodiments, the first dsDNA may be from a disease cell, for example a cell obtained from an individual with a disease, such as a cancer cell, and the second dsDNA may be from a normal or healthy cell.

10 The frequency or distribution of SSBs within the first dsDNA relative to the second dsDNA may be indicative of the diagnosis, severity, penetrance grade or prognosis of a disease, such as cancer, in the individual from whom the first dsDNA is obtained. The disease or cancer may be caused or characterised by a defect in a DNA repair pathway.

15 In other embodiments, the first dsDNA may be from a cell at a specific developmental stage, for example a cell obtained from an embryo, and the second dsDNA may be from an adult or mature cell. In other embodiments, the first dsDNA may be from an age-associated cell, for example a cell obtained from an individual of a specific age, such as an elderly person, and the second dsDNA may be from a control individual. In other embodiments, the first dsDNA may be from a treated cell and the second dsDNA may be
20 from an untreated control or reference cell. For example, the first dsDNA may be from a cell treated with reprogramming factors, such as Oct3/4, Sox2, Klf4 and c-Myc (the "Yamanaka factors" or "OSKM"), radiation, chemical compounds, and drugs, such as chemotherapy agents.

In some embodiments, the methods described herein may be used to stratify cells according to their age
25 and/or disease status e.g. cells obtained from a test subject e.g. a human subject.

In some embodiments, the methods described herein can be used in an assay to detect the therapeutic effects of agents used to treat infection e.g. wherein said subject, such as a human subject, is infected with a virus or bacteria. In such an assay viral or bacterial dsDNA is obtained from a subject who has been treated
30 with an anti-viral or anti-bacterial agent, and said dsDNA is then subjected to the methods described herein to detect SSBs. The test results can be compared to control samples, comprising the same viral or bacterial DNA that has not been exposed to any anti-viral/antibacterial agent. Such an assay can be used to determine the effectiveness of anti-infective agents such as antivirals or anti-bacterial agents (e.g. antibiotics).

35

Preparation of dsDNA can be done using standard methods known in the art

In some embodiments, the preparation of dsDNA from any of the types of cells described herein can comprise one or more of the following steps which can facilitate introduction of the oligonucleotide probe:

(i) cell samples comprising dsDNA can be subjected to a wash and lysis step, for example this can be done in a vessel e.g. tubes, vials or coverslips, with a small volume e.g. volume of about 0.1ml to about 2.0ml, e.g. a volume of about 0.2ml, or about 0.5ml or about 1.5ml.

(ii) samples comprising dsDNA can then be subjected to a buffer exchange step e.g. by
5 centrifugation of the sample or by dilution.

Buffer exchanges described above can be achieved by use for example of the following method:

Firstly, cells are suspended in PBS (or a similar iso-tonic buffer). To isolate nuclei from cells, they are centrifuged, the PBS is removed, and the cells are resuspended in a lysis (or nuclear isolation) buffer which
10 is hypotonic and contains a mild detergent resulting in solubilisation of the cell membrane but leaving the nuclear membrane intact. Next cells are spun down again and resuspended in a second lysis buffer containing a stronger detergent (SDS) which solubilises protein, thus removing the chromatin protein from the DNA (a step known as nucleosome depletion). After this, nuclei are spun down again and resuspended in a buffer without SDS, this stage is repeated a few times in order to wash the nuclei and ensure SDS is
15 removed (as presence of SDS can inhibit downstream reactions).

In some embodiments the preparation of dsDNA from cells can comprise the following step to facilitate dsDNA purification;

after contacting the samples of dsDNA with the DNA ligase according to the methods detailed herein
20 said samples can then be purified e.g. by proteinase digestion, for example by Proteinase K digestion e.g. by incubation of the sample with Proteinase K for a period of about 2hrs to about 30 minutes, e.g. by incubation for about 30 minutes.

In some embodiments, the isolation of the DNA strands that comprise the first and second sequencing
25 adaptors can be performed using magnetic beads, wherein samples comprising dsDNA are mixed with magnetic beads and incubated for a period of about one hour to about 2 hours e.g. about 30 mins.

In some embodiments, the disclosure provides a method for detecting SSB breaks in dsDNA obtained from cells (e.g. any of the cells detailed herein) comprising the following steps:

- 30 (i) optionally fixing said cells,
(ii) subjecting said cells to a wash and lysis step, e.g. in tubes/vials/coverslips with a volume of about 0.1 – about 0.3ml e.g. about 0.2ml, and then
(iii) subjecting samples comprising dsDNA to buffer exchange, e.g. by centrifugation of the sample or by dilution,
35 (iv) contacting a sample of dsDNA with an oligonucleotide probe comprising a first sequencing adaptor and a tag,
(v) contacting the sample of dsDNA with DNA ligase, such that the DNA ligase covalently links the oligonucleotide probe to a strand of the dsDNA at the site of a single strand break in the dsDNA,
40 (vi) purifying the dsDNA e.g. by proteinase digestion, for example by use of Proteinase K, e.g. by incubation of the sample with Proteinase K for a period of about 2hrs to about 30 mins,

- (vii) contacting the sample of dsDNA with a transposase loaded with a second sequencing adaptor, such that the transposase cleaves the dsDNA to generate a population of fragments comprising a terminal second sequencing adaptor,
- (viii) isolating from said population, DNA strands that comprise the first and second sequencing adaptors, e.g. by incubation with magnetic beads, e.g. for an incubation period of about 5 mins to about 2 hours e.g. for about 30 mins, and
- (ix) determining the sequence of the isolated DNA strands.

In some embodiments of the above, the oligonucleotide incorporation and ligation steps can be performed directly on purified DNA or on intact nuclei.

The number of cells in step (i) or (ii) above, can range from about 10,000 to about 1 million e.g. about 25,000, about 50,000, about 75,000, about 100,000, about 150,000, about 200,000.

- Proteinase digestion can be carried out at temperatures ranging from about 37°C to about 60°C, e.g. at about 37°C when a thermolabile Proteinase K is used.

The above steps (i) to (ix) can allow the methods for detecting SSBs to be performed using fewer cells and also enable them to be performed in a shorter time period.

High throughput methods: the methods for detecting SSBs described herein may be adapted to obtain high throughput methods. In order to scale the method to higher numbers of samples, cells can be collected and processed in 96 or 384 well PCR plates (or by using a microfluidic device). The oligonucleotide can be redesigned to contain a barcode sequence which is specific to each sample (i.e. 96 distinct barcodes used for a 96w plate of samples). After ligating the barcoded oligo, all samples can be pooled together (e.g. rapidly using a centrifuge into a v-block - clickbio CBVBLOK200) for purification and all other downstream processing. After sequencing and data processing, samples can be identified via the barcode sequence. Use of an automated liquid handler can also be used to make the process more efficient. Removal of buffer exchanges, instead of adding new buffers to the old ones and diluting out the previous components, can also be employed so reducing pipetting steps required.

Single cell methods: the methods for detecting SSBs described herein can be performed on single cells, including for example when the dsDNA is obtained using the above methods for optimising introduction of the oligonucleotide probe and purification of dsDNA, and can generally be performed as detailed below:

a sample containing dissociated, fixed cells is flow-sorted in order to isolate single-cells into individual wells of a 96 or 384 well plate, which contains the lysis buffer required for nuclei isolation and nucleosome depletion. As with the high-throughput method described above, the subsequent steps are performed in the PCR plate and a barcoded oligo can be employed to allow pooling of samples. Removal of the buffer exchanges, instead using buffer additions may optimally be employed for this since cell losses are observed with centrifugation/buffer exchange. When buffer additions are employed, cells are collected into lysis buffer after lysis incubation and another buffer is then added which can dilute or inactivate the lysis

reagents. After this step oligo and ligation reagents can be added, ensuring that the concentration of salts and detergent is appropriate for ligase activity.

Methods described herein may be useful for example in cytotoxicity assays. A cytotoxicity assay may
5 comprise subjecting a cell to a stimulus and determining the frequency or distribution of gaps or SSBs in dsDNA from the cell. An increase in the frequency or distribution of gaps or SSBs in dsDNA from the cell relative to dsDNA from a control cell not subjected to the stimulus is indicative that the stimulus is cytotoxic. Suitable stimuli include exposure to chemical compounds and radiation.

10 In other embodiments, the first dsDNA may be from a cell treated with a nuclease such as a CRISPR nuclease (eg, Cas9 or Cpf1) and the second dsDNA may be from an untreated control cell. The frequency or distribution of SSBs within the first dsDNA relative to the second dsDNA may be indicative of off-target effects of the nuclease in the treated cell.

15 In other embodiments, the first dsDNA may be from an egg or sperm cell obtained from a donor and the second dsDNA may be from a control egg or sperm cell. The frequency or distribution of SSBs within the first dsDNA relative to the second dsDNA may be indicative of the quality of the donor egg or sperm cell. This may be useful for example in assisted reproductive technology (ART) methods.

20 A set of sequence reads of amplified or isolated DNA strands may be generated by sequencing. For example 1000 or more, 10,000 or more, 100,000 or more, 1,000,000, 10,000,000 or more, or 100,000,000 or more, or 1,000,000,000 or more sequence reads may be generated. The sequence reads may be analysed by routine bioinformatic techniques. Suitable techniques are well-known in the art. For example, duplicate reads, low quality sequence reads and reads arising only from sequencing adaptors may be removed.

25 In some embodiments, the sequence reads in the set may be analysed for the presence of SSB sites, the frequency of SSBs occurring at sites and/or patterns of SSB sites within the dsDNA. The sequence reads in the set may further be analysed for the presence of other features, such as mutations, epigenetic modifications, or sequence motifs, that are associated with SSBs.

30 The fragment sequence reads in the set may be mapped to one or more locations in a reference genome. Suitable reference genomes are available in the art. For example, human nucleic acid fragment sequence reads in the set may be mapped to locations in the sequence of the human genome. In some embodiments, the reference genome may be matched to the gender, ethnicity and/or other characteristics of the individual
35 from whom the sample is obtained.

The fragment sequence reads in the set may be mapped by aligning sequence reads in the set with the sequence of the reference genome, for example the human genome. The location of the sequence reads within the reference genome may be identified. Suitable software tools for mapping populations of sequence
40 reads within the genome are readily available in the art.

The distribution of the fragment sequence reads in the set within the genome or at set of sites or loci within the genome (i.e. the number of sequence reads that map to each location within the genome or set of sites or loci) may be determined from the locations of the sequence reads in the set. Optionally, the distribution of the fragment sequence reads may be subjected to mathematical transformation.

5

A sample SSB profile may be generated from the distribution or transformed distribution of the nucleic acid fragment sequence read. The sample SSB profile may comprise a set of scores or values indicative of the number or density of nucleic acid fragment sequence reads that map to each location or position within the genome (i.e. a genome wide plot) or set of sites or loci within the genome. The number or density of fragment sequence reads that map to a location or position in the genome may be indicative of the frequency of SSBs at that location or position. The sample SSB profile may therefore reflect the distribution of SSBs in the genome or in target loci within the genome. The sample SSB profile may be expressed in any convenient format, for example numerically or graphically.

10

15 In some embodiments, the sample SSB profile may be used to identify sites of SSBs that are associated with biological responses to a drug or other compound, such as cytotoxic responses, and may be useful in determining or predicting the response of an individual to treatment with the drug or other compound. The sample SSB profile may be used for therapeutic stratification, the optimization of combination therapies, diagnosis of a disease condition, determination of whether cells in a subject (e.g. a human subject) are less
20 resilient and can be selected for treatments that can stimulate repair of SSBs and /or rejuvenate and/or reprogram the cells e.g. by exposing them to agents (chemical and/or biological) that can stimulate repair of SSBs, the determination of side-effects of treatment with a drug or other compound, cytotoxicity assays, the determination of the off-target effects of nucleases, assisted reproductive technology methods, or the study of ageing or development. In some embodiments the SSB profile can be used to distinguish between
25 different cell types, e.g. for cell type identification, the method can be used for cell profiling or in screening to discover cell-type specific differences in genomic distribution of DNA gaps and also to examine the effect of transcriptional perturbation on DNA gap formation dynamics.

30

A marker (target) site is a position in the genome at which the presence or frequency of SSBs varies between different sources i.e. an SSB at the marker site is source specific, for example from different individuals, tissues, cell-types, developmental stages or ages. A set of marker sites in a sample SSB profile may provide a signature that is characteristic of the source.

35

In some embodiments, suitable marker sites may be identified by determining the presence of SSBs at a plurality of candidate sites in reference SSB profiles from a set of reference sources. The reference SSB profiles may comprise a set of scores or values indicative of SSBs at the set of marker sites in genomic DNA from a known source, for example, a specific known tissue or cell type. The reference SSB profile may reflect the presence of SSBs at the locations in genomic DNA from the known source. Suitable reference SSB profiles may be obtained or generated by routine experimentation using known tissues or cell type or
40 produced from publicly available data sources, such as databases of genomic information.

A candidate site may be identified as an SSB marker site if the frequency of an SSB occurring at the site is higher or lower for one reference source in the set than for the other sources in the set. For example, the frequency of SSB occurrence at the site may be higher or lower than the mean frequency of SSB occurrence in the other reference sources in the set. In some embodiments, the frequency of SSB occurrence at the site in one reference source may be above or below a predetermined threshold value relative to the mean frequency of SSB occurrence at the corresponding site in the other reference sources in the set.

In other embodiments, suitable marker sites may be identified by providing a first set of sample SSB profiles from control individuals, for example healthy individuals, and a second set of sample SSB profiles from test individuals, for example individuals with a disease condition, such as a specific cancer, individuals known to be responsive or non-responsive to a drug, or individuals of a specific age. The frequency of SSBs at a plurality of candidate sites in the first and second sets of sample SSB profiles may be compared. A candidate site may be identified as a marker site if the frequency of SSB occurrence at the site is higher or lower in the first set of sample SSB profiles than the second set of sample SSB profiles. For example, the mean frequency of SSB occurrence at the site may be higher or lower in the first set than the second set. In some embodiments, the mean frequency of SSB occurrence at the site in the first set may be above or below a predetermined threshold value relative to the mean frequency of SSB occurrence at the site in the second set.

A candidate site may also be identified as a marker site if it is found to be in linkage disequilibrium with a site at which the frequency of SSB occurrence is higher or lower for one reference source in the set than for the other sources in the set.

A sample SSB profile from an individual may be used to determine the presence of SSBs or the frequency of SSB occurrence at a set of marker sites in the individual. This may be useful in determining the effect of a drug or other compound on the individual or the responsiveness of the individual to the drug or other compound, for example, to determine the effectiveness of treatment with the drug or other compound in the individual or to determine the suitability of the individual for treatment with the drug or other compound. A sample SSB profile from an individual may also be used to demonstrate the mechanism of effect of the drug or other compound; as part of a clinical trial; or as evidence for off-target toxicity, where the toxicity is caused by SSBs at additional/alternative locations within the dsDNA.

The methods described herein may be used in combination with other analytical techniques, for example in a multiomic genome analysis. In some embodiments, methods described herein may further comprise determining the presence of epigenetic marks at or near the sites of SSBs in dsDNA. Suitable epigenetic marks include modified nucleic acid bases, such as methylated DNA cytosine. For example, isolated DNA strands produced as described herein may be treated with bisulphite or in combination with other chemical modifications (e.g. conversion of existing modifications to 5caC so to discriminate 5hmC/5mC) and sequenced. The sequences of the isolated DNA strands may be compared with a reference genome sequence and cytosine residues that are modified in the isolated DNA strands identified.

The methods described herein may substitute dsRNA or a hybrid of DNA/RNA wherein one strand is DNA, and the other strand is RNA, for dsDNA. When dsRNA is used in the methods described herein the method can be modified as follows: an RNA ligase (e.g. T4 RNA ligase) can optionally be used instead of a DNA ligase, and RNA oligonucleotides may optionally be employed instead of DNA oligonucleotides. A reverse transcriptase may be employed to transcribe the RNA to DNA after performing the ligation step.

The dsRNA can be for example a siRNA, mRNA, or tRNA. The methods described herein may be used, for example, as part of quality control, e.g. to assess stability of dsRNA such as therapeutic siRNA, including stability of such RNA molecules when kept at different conditions e.g. when present in different formulations, kept in different storage conditions, such as storage at different temperatures and/or for different time periods. The methods when applied to dsRNA may also be used to identify RNA base modifications on the RNA strand bearing the modification.

Screening methods: the methods described herein to detect SSBs may be used in screening methods e.g. to identify agents (such as a chemical and/or biological agent, e.g. a small molecule and/or a peptide and/or protein therapeutic) that can enable repair of SSBs in dsDNA and/or reprogram cells.

Hence, also provided is a method of screening for such agents e.g. a small molecule and/or a peptide and/or protein therapeutic, that can enable repair of SSBs in dsDNA and/or reprogram cells, which comprises the following steps:

- (a) detecting single strand break (SSB) sites in a control sample of double stranded (ds) DNA that has not been exposed to an agent (such as a chemical and/or biological agent e.g. a small molecule and/or a peptide and/or protein therapeutic) and also detecting single strand break (SSB) sites in a test version of the same sample of dsDNA that has been exposed to an agent (such as a chemical and/or biological agent) by;
 - (i) contacting a sample of both the test and control dsDNA with an oligonucleotide probe comprising a first sequencing adaptor and a tag,
 - (ii) contacting the sample of test and control dsDNA with DNA ligase, such that the DNA ligase covalently links the oligonucleotide probe to a strand of the dsDNA at the site of a single strand break in the dsDNA,
 - (iii) contacting the sample of test and control dsDNA with a transposase loaded with a second sequencing adaptor, such that the transposase cleaves the dsDNA to generate a population of fragments comprising a terminal second sequencing adaptor,
- (b) isolating from said population in both test and control samples, DNA strands that comprise the first and second sequencing adaptors;
- (c) determining the sequence of the isolated DNA strands from the test and control samples,
- (d) comparing the profile of SSBs in both the test and control samples,
- (e) isolating the chemical and/or biological agents that are associated with a reduction in SSBs in test samples compared to control samples, and
- (f) formulating said isolated chemical and/or biological agents as pharmaceutical compositions e.g. by adding appropriate pharmaceutically acceptable excipient(s).

The above pharmaceutical compositions can be used to treat subjects (such as human subjects) to promote repair of SSBs in cells and/or to reprogram said cells in said subjects.

- 5 The screening method for SSBs detailed above may be applied to cells such as human cells in test subjects and can also further comprise one or more of the following steps e.g. to facilitate introduction of the oligonucleotide probe:

(i) cell samples comprising dsDNA can be subjected to wash and lysis for example in tubes or vials or coverslips with a small volume e.g. about 0.1- about 0.3ml, e.g. about 0.2ml,

- 10 (ii) samples comprising dsDNA can then be subjected to buffer exchange, e.g. by centrifugation of the sample or by dilution,

- In some embodiments, the screening methods above can comprise the following step to facilitate dsDNA purification: after contacting the samples of dsDNA with the DNA ligase the samples can be purified, e.g. by
15 proteinase digestion, for example by Proteinase K digestion, e.g. by incubation of the sample with Proteinase K for a period of about 2hrs to about 30 mins.

- In some embodiments, the isolation of the DNA strands that comprise the first and second sequencing adaptors can be performed using magnetic beads e.g. samples comprising dsDNA can be incubated with
20 magnetic beads for an incubation period of about 5 minutes to about 2 hours e.g. about 30 mins.

Proteinase digestion can be carried out at temperatures ranging from about 37°C to about 60°C, e.g. at about 37°C when a thermolabile Proteinase K is used.

- 25 The above screening methods can be used to identify agents/compounds that can reprogram cells and/or repair SSBs in dsDNA present in cells. The screening methods can also be used to follow the time course of cell reprogramming and/or repair, e.g. by performing the methods on ds DNA obtained at different time points following incubation of the agent/compound with the dsDNA. The screening methods above can also be adapted to make and isolate dsDNA in which gaps have been introduced by design, by substituting
30 agents that can enable repair of SSBs with agents that promote gap formation, and then testing the dsDNA exposed to such agents to confirm presence of gaps by comparison to control samples that have not been exposed to agents that promote gap formation.

- Also provided are kits for using the methods described herein and the use of such kits in such methods. A
35 kit may comprise;

an oligonucleotide probe comprising a first sequencing adaptor and a tag,
a DNA ligase,
a transposase and
a second sequencing adaptor loaded or loadable onto the transposase.

The kit may further comprise a binding member that specifically binds to the tag; sequencing primers and/or other reagents. The kit may further comprise suitable buffers and washing solutions; fixing agents and permeabilising agents.

- 5 The kit may further comprise a solid support. Suitable solid supports may comprise a capture molecule that binds to eukaryotic cells, such as a lectin; and/or the binding member.

The kit may further comprise nucleic acid extraction and purification reagents. Suitable reagents are well-known in the art and include spin-chromatography columns.

10

The kit may further comprise amplification reagents. Suitable reagents are well-known in the art and include primers, dNTPs, and thermostable polymerases. In some embodiments, a kit may comprise amplification primers for amplification of one or more regions or loci of interest in a genome.

- 15 The kit may include instructions for use in a method described herein.

Other aspects and embodiments of the invention provide the aspects and embodiments described above with the term “comprising” replaced by the term “consisting of” and the aspects and embodiments described above with the term “comprising” replaced by the term “consisting essentially of”.

20

It is to be understood that the application discloses all combinations of any of the above aspects and embodiments described above with each other, unless the context demands otherwise. Similarly, the application discloses all combinations of the preferred and/or optional features either singly or together with any of the other aspects, unless the context demands otherwise.

25

Modifications of the above embodiments, further embodiments and modifications thereof will be apparent to the skilled person on reading this disclosure, and as such, these are within the scope of the present invention.

- 30 All documents and sequence database entries mentioned in this specification are incorporated herein by reference in their entirety for all purposes.

“and/or” where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. For example “A and/or B” is to be taken as specific disclosure of each of (i) A, (ii) B and (iii) A and B, just as if each is set out individually herein.

35

Experimental

Example 1:

Materials and Methods

- 40 Examples of the GAPS-Seq method are shown schematically in Figure 1. These examples comprise 1) cross linking cell/tissues using a cross-linking agent to fix DNA bearing single strand DNA gaps or SSBs; 2) using

detergents to permeabilise cells and denature chromatin proteins thereby allowing access to genomic DNA containing SSBs, 3) directly ligating to sites of DNA gaps a DNA oligonucleotide probe comprising a random hybridisation region, a first sequencing adaptor and a tag; 4) nuclei lysed to isolate genomic DNA; 5) adding a second adaptor to the isolated DNA using tagmentation; 6) capturing DNA gap bound, tagged adaptors; and 6) separating, amplifying and sequencing the DNA strand containing the gap. Briefly cells were dissociated into single cells using TrypLE Express Enzyme then fixed in 2% formaldehyde for 10 minutes at room temperature. Nuclei were isolated using 0.2% Triton-X100 on ice for 15 minutes then chromatin proteins/DNA bound proteins denatured using SDS. Nuclei were then pelleted using centrifugation and resuspended in biotinylated GAPS-seq oligo which contains an Illumina Read 1 sequence. T4 DNA ligase was then added and the suspension incubated overnight at 16°C with shaking (at 800rpm). Proteinase K was then used to reverse crosslinks and digest protein, following which the DNA was purified on columns. About 40ng of the purified DNA was tagmented with Illumina Read 2 then GAPS-seq fragments are captured on streptavidin beads prior to indexed PCR.

15 *Sample Preparation*

1M cells were dissociated into single-cells and fixed in 2% formaldehyde for 10min at room temperature. The reaction was quenched by adding 2 M glycine to a final concentration of 125 mM and the supernatant removed by spinning at 100g for 5min at 4°C. The cells were then resuspended twice in 1ml Phosphate Buffered Saline (PBS) and the supernatant removed by spinning at 100g for 5min at 4°C to wash. The cells were resuspended in 1mL ice cold LB1 (10mM Tris pH8.0, 10mM NaCl, 1mM EDTA, 0.2% Triton) and incubated for 15 mins on ice. The supernatant was then removed by spinning at 300g for 5 min at room temp. The nuclei were then resuspended in 1ml LB2 (10mM Tris pH8.0, 150mM NaCl, 0.3% SDS) pre-warmed to 37°C and then incubated at 37°C in a thermal mixer for 60 min, gently shaking at 400 rpm and spun at 300g for 5 min at room temp to remove the supernatant. Fresh CTSX buffer was prepared (100ul 10x Cutsmart, 10ul 10% Triton X-100, 890ul nuclease free water) and warmed to 37°C and used to resuspend the nuclei. The nuclei were checked by trypan blue staining and viewing under microscope to ensure they were 100% trypan blue positive and be intact and round in shape. The nuclei were then spun at 300g for 5 min at room temp, the supernatant removed and the CTSX wash repeated.

30 *Sample preparation for etoposide treatment*

Etoposide treatment was carried out to validate the method as etoposide is known to induce both single and double strand DNA gaps (Muslimovic et al., 2009). For this experiment, passage 14 HEK293 cells were treated with 30uM etoposide added to the cell culture media and incubated at 37°C, in a 5% CO2 incubator for 4 hours. Control samples were treated with DMSO to the same concentration as in the etoposide treated cells and also incubated for 4 hours, in a 5% CO2 incubator at 37°C. Dissociation and fixation were performed as above.

Sample preparation for restriction enzyme experiment

1M cells (6pg x 1M = 6ug dsDNA) were fixed and processed exactly as above. After washing in CTSX buffer, the nuclei were re-suspended in 100ul of CTSX buffer and 2ul of the appropriate restriction enzyme was added. Samples were then incubated overnight at the recommended temperature (37°C for all enzymes

except TspRI, which required 65°C) with shaking at 400rpm. A control sample which did not receive any restriction enzyme was processed in parallel and this was also incubated at 37°C, with shaking at 400rpm.

Ligation of GAPS-seq oligo

- 5 The nuclei were then resuspended in 4ul of 100uM ssBreak oligo and 96ul of ligation mix (formulated as shown below in Table 1) added on ice

water	82 ul
10x T4 ligase buffer (NEB)	10 ul
T4 ligase (2000u/ul)	4 ul

Table 1

- 10 The nuclei were incubated with the T4 ligase overnight on shaker at 16°C and 400 rpm

Reverse Cross linkers and DNA Purification

10ul of proteinase K (NEB) was added to the ligation reaction and incubated at 55°C for 2 hours. DNA from the nuclei was then purified using the zymo Quick-DNA Microprep kit, 500ul DNA lysis buffer added, and the
15 nuclei vortexed and spun at 10,000g for 1 min. The supernatant was transferred to a column and spun at 10,000g for 1 min, 200ul prewash buffer was added to the column and the column was spun at 10,000g for 1 min, then 500ul gDNA wash buffer was added to the column and the column spun at 10,000g for 1min.

- 20 20ul water was then added to the column and it was placed in a 1.5ml Eppendorf tube and spun at max speed for 5 min at room temperature. DNA was then quantified using qubit and the results recorded. A tagmentation reaction was then set up as shown below in Table 2;

TD buffer (Illumina)	21 ul
N7 Tn5 12.5uM	1 ul
DNA	20 ul

Table 2

The reaction was incubated at 55°C for 10min and then 10.5ul of 0.2% SDS was added.

25

Biotin Pull down of SSB Samples

Streptavidin beads were prepared. 5ul C1 beads per sample were washed twice in 2x B&W (10mM Tris, 1mM EDTA, 2M NaCl) and resuspended in 50ul of 2x B&W. 50ul sample was added to 50ul beads and mixed on a shaker at room temp for 1 hour. The supernatant was removed using a magnet and the beads
30 washed twice in 0.1M NaOH and twice in water. The beads were resuspended in PCR mix as shown below in Table 3:

KAPA HiFi 2x readymix	10 ul
Bst3.0 (NEB)	1 ul
10uM i7 index primer	1 ul

10uM i5 index primer	1 ul
water	7 ul

Table 3

The DNA on the beads was amplified as follows; 65°C 5min; 98°C 45s, then 20 cycles of; 98°C 15s, 60°C 30s, and then finally 72°C 1 min. The PCR products were then purified with 0.65x AMPure XP beads (13ul) and QC'd on tapestation D5000 high sensitivity tape.

5

Cell number dilution experiment

In order to titrate the input requirements, a dilution series was prepared comprising 1 million, 500,000, 100,000, 10,000, 1,000, 500 and 100 cells. All steps remain the same except for the number of PCR cycles which was increased by 2 (100,000 cells), 3 (10,000 cells), 5 (1,000 cells) or 8 (500 and 100 cells) extra cycles.

10

Sequencing and data processing

Libraries were pooled in equimolar amounts then sequenced using an Illumina NextSeq2000 instrument and a P1 flowcell following the manufacturer's instructions with 50nt paired end reads and 8nt i7 and i5 index reads. Sequencing reads were de-multiplexed using bcl2fastq2 (Illumina) to produce 2 fastq files corresponding to read 1 and read 2. Fastq files were processed using the Nextflow chipseq pipeline (Ewels, et al. 2020) to produce BAM files containing aligned reads. Briefly this entails raw read QC with FastQC (v0.11.9)(Andrews 2010), then trimming using Trim Galore (v0.6.7) (Krueger et al 2023) and cutadapt (v3.4) (Martin et al 2011) to remove Illumina primer sequences from the 3' end and additionally the first 6nt at the 5' end, which corresponds to the random priming segment. Trimmed reads are then aligned to the GRCh38 version of the human genome using BWA (v0.7.17-r1188) (Li et al 2010) and filtered to remove duplicates and reads mapping to blacklisted regions using Samtools (v1.15.1) (Danecek et al 2021) and picard (v2.27.4).

15

20

25

30

For the analysis of promoter regions, BAM files were loaded into Seqmonk v1.48.1 and a probe trend plot generated using default parameters. Data was exported as a text file and imported into R (v4.1.2) (R Core Team 2022) to enable plotting using ggplot2 (v3.4.2) (Wickham 2009). For the analysis of restriction enzyme sites, we first located all instances of a given restriction enzyme motif in the GRCh38 genome using the BSgenome.Hsapiens.UCSC.hg38 (v1.4.4) package and the matchPattern function from biostrings (v2.64.1) in R. This was imported as an annotation track into Seqmonk and a probe trend plot generated as with promoters above.

Results

The GAP-seq method was used to determine the relative frequency of single-strand gaps at gene promoters (Figure 2). This data corresponded to existing knowledge about the distribution of breaks in the genome and confirmed that the GAP-seq method can successfully detect SSBs.

35

ssDNA gaps were profiled in HEK293 cells treated and untreated with etoposide. Etoposide treatment is known to increase endogenous DNA gaps by inhibiting topoisomerase. GAPS-seq was found to detect more

ssDNA gaps in etoposide treated vs DMSO only treated controls (Figure 3). Interestingly, compared to surrounding regions, GAPS-seq also showed the presence of more ssDNA gaps in enhancer sequences, providing indication that regulatory regions undergo dynamic cycles of ssDNA gap and repair.

- 5 DNA GAP-seq was found to detect ssDNA gaps in libraries obtained from different input samples, ranging from 1 million, to 100 cells (Figure 4).

HEK293 cells were treated with restriction enzymes to cut dsDNA at distinct motifs leaving either 3' or 5' overhangs. Using the GAP-seq method to targets the 3' end of broken strand, cut sites with a 5' overhang (BstEII-HF; EcoR1-HF) were successfully mapped, but not blunt end cut sites (SrfI) or cut sites with a 3' overhang (Apa1) (Figure 5). The plots show an average signal over all known restriction sites in the human genome.

15 The GAPS-Seq method described herein was found to be a highly versatile and quantitative method for the genome-wide detection of DNA gaps at base-resolution.

Example 2: enhanced protocol - GAPS-seq v2

Further optimisations of the above protocol now allow the GAPS-seq method to be completed in a shorter time frame (24 hours instead of 48 hours) and with fewer steps. This updated protocol (GAPS-seq v2) additionally allows a reduced input in terms of cell number (e.g. about 100,000 instead of about 1 million).

The following describes the modifications from GAPS-seq v1 (in the example above) to v2:

After fixation, cells are transferred into 0.2ml PCR tubes for the subsequent wash and lysis steps. Use of the smaller tubes reduces cell losses associated with nuclei 'sticking' to the walls of tubes.

25 Gap-seqv2 also allows:

Reduction of all centrifugation times from 5 minutes to 2 minutes.

A simplified gDNA purification following the overnight ligation. Proteinase K digestion has been shortened from 2 hours to 30 minutes and the use of magnetic beads instead of columns that allows processing of many more samples, with less input material.

30

Specifically, the protocol was performed as follows:

- (i) Following overnight ligation, nuclei are pelleted by centrifugation at 500g for 2 minutes.
- (ii) Supernatant is removed by pipetting, leaving approximately 10ul of liquid in the tube.
- (iii) 1 ul of 8U/ul Proteinase K (NEB P8107S) is added.
- 35 (iv) Samples are incubated at 55c for 15 minutes followed by 80c for 15 minutes
- (v) 35ul of RLT plus buffer (Qiagen, 1053393) is added.
- (vi) 45ul of SPRI select beads (Beckman, B23318) are added, vortexed thoroughly and incubated at room temperature for 10 minutes.
- (vii) Samples are placed on a magnet and supernatant removed once the beads pellet.
- 40 (viii) Beads are washed twice with 200ul of 80% ethanol with the tubes on the magnet

- (ix) To elute gDNA, 25ul of EB buffer (Qiagen, 19086) is added to the beads off the magnet and vortexed to mix
- (x) Tubes are placed on a thermomixer at 50c and 2000rpm for 10 minutes
- (xi) Tubes are then placed on a magnet and once the beads pellet, the liquid is transferred to clean PCR tubes
- (xii) Optimised tagmentation reaction: 4ul of 12.5uM N7 loaded Tn5 (obtained from Diagenode - cat C01070010-10) is now used instead of 1ul. This results in greater library complexity.
- (xiii) Samples are conjugated to streptavidin beads for about 15 minutes instead of 1 hour.

This above optimized protocol has been performed on multiple cell lines and examples of those are detailed in the below examples. The numbers of cells used ranged from 10,000 cells to 1,000,000 cells.

Example 3: Nickase and exonuclease experiment

GAPS-seq is designed to profile single-strand gaps in gDNA but not nicks, as nicks are much more abundant and physiologically less impactful. To determine whether GAPS-seq is sensitive to nicks versus gaps the following experiment was designed and carried out. BJ cells [Cell type: human foreskin fibroblast; Source: ATCC: Cat. no: CRL-2522 Lot number: 70027151) were prepared following the GAPS-seq v2 cell fixation and lysis protocol described above in example 2. Nuclei were then treated with 0.1U/ul of the nicking endonuclease, Nt.BbvCI (NEB R0632S) in 1x rCutSmart Buffer (NEB, B6004) buffer for 1 hour at 37°C. Next, nuclei were pelleted by centrifugation (as detailed above, at 300g for 2 minutes) and washed twice in CTSX buffer before being resuspended in 1x rCutSmart Buffer (NEB, B6004) with 1U/ul of Exonuclease III (NEB, M0206S) and incubated for 1 hour at 37°C. Nuclei were once again washed twice in CTSX buffer then the 5-prime GAPS-seq v2 protocol was continued from the step in which the oligo and then the ligation mastermix is added. Controls included (a) nuclei that received only the nickase and not the exonuclease and (b) nuclei that were not treated with either enzyme. Plotting at the Nt.BbvCI recognition site (CCTCAGC) was performed as with the restriction enzyme experiment above. The combination of the two enzymes creates a single-strand gap with a 5' end at the Nt.BbvCI recognition site, whereas treatment with the nickase only creates nicks in the DNA for which GAPS-seq is designed to ignore. Thus this experiment demonstrates that GAPS-seq is specific for single-strand DNA gaps and does not profile nicks in the DNA (Figure 6).

Example 4: Transcription start site profiles stratified by gene expression level

Both double-strand and single-strand gaps are known to be enriched at the transcription start sites of genes in a manner that is dependent on expression levels of the corresponding gene. To confirm that GAPS-seq can detect this phenomenon, GAPS-seq v2 was performed on HEK293 cells as described above. For plotting, GAPS-seq BAM files (i.e. the mapped sequencing reads) were read into R using the Rsubread package (<https://bioconductor.org/packages/release/bioc/html/Rsubread.html>) and overlapped with 10kb regions surrounding the transcription start site of all genes. These were then stratified into 3 groups based on the expression level of the corresponding gene, which was derived from published RNA-seq of HEK293 cells (<https://doi.org/10.1093/nar/gky861>): 1) highly expressed genes (500 or more raw counts); 2) middling expressed genes (raw counts between 10 and 500) and 3) lowly expressed genes (raw counts less than 10).

GAPS-seq signal was then averaged over 400bp windows then averaged across all genes (for each group separately) and plotted as a line graph (Figure 7). As the level of transcription is found to correlate with levels of basal DNA breaks, it shows this method could e.g. be used to examine the effect of transcriptional perturbation on DNA gap formation dynamics and possibly vice versa.

5 Example 5: GAPS-seq can discriminate between cell types

GAPS-seq has the potential to reveal cell type specific DNA gaps and be used for cell type discrimination. To demonstrate this, GAPS-seq was performed on (1) HEK293 cells (2) BJ fibroblast cells and (3) human ESCs and NPCs derived in culture from ESCs. Library preparation, sequencing and data processing was performed as described above in Gap seq version 2. To plot a dendrogram, the GAPS-seq signal was
10 quantified at promoter regions (+/- 2kb from the TSS of all protein coding genes) then hierarchical clustering was performed using the dist and hclust functions within R and plotted using the gg dendro package [obtainable from <https://cran.r-project.org/web/packages/ggdendro/index.html>]. Hierarchical clustering clearly separates samples by the cell line of origin, thereby demonstrating GAPS-seq utility in cell type discrimination (Figure 8). Results show GAPS-seq can clearly distinguish between different cell types, which
15 could be useful for cell type identification and indicates that the method can be used to discover cell type specific differences in genomic distribution of DNA gaps.

Example 6: GAPS-seq profiles are distinct from double-strand break profiling method sBLISS

Methods for profiling double-strand breaks, exist e.g. sBLISS.. To test that GAPS-seq is profiling a distinct genomic feature to double strand breaks sBLISS was performed on the same batch of HEK293 cells as
20 GAPS-seq described above. The sBLISS protocol was followed as described in Bouwman et al., Nature Protocols 15, 3894-3941 (2020) . To compare the signal generated by the two methods we overlapped the signal with promoter regions (+/- 2kb from the TSS) then performed principal component analysis using the prcomp function within R (v4.1.2) (R Core Team 2022) <https://cran.r-project.org/>. The results are shown in Figure 9, which clearly shows the separation of the samples by the method used to produce the library and
25 therefore demonstrates that the GAPS-seq method is measuring a distinct aspect of the genome from double-strand break profiling methods such as sBLISS.

References

Lindahl T, Nyberg B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry*. 1972 Sep 12;11(19):3610-8. doi: 10.1021/bi00769a018. PMID: 4626532

Lindahl T, Barnes DE. Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol*. 2000;65:127-33. doi: 10.1101/sqb.2000.65.127. PMID: 12760027.

Amente S, Scala G, Majello B, Azmoun S, Tempest HG, Premi S, Cooke MS. Genome-wide mapping of genomic DNA damage: methods and implications. *Cell Mol Life Sci*. 2021 Nov;78(21-22):6745-6762. doi: 10.1007/s00018-021-03923-6. Epub 2021 Aug 31. PMID: 34463773; PMCID: PMC8558167.

Yan, W., Mirzazadeh, R., Gamberone, S. et al. BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat Commun* 8, 15058 (2017). <https://doi.org/10.1038/ncomms15058>

Wu W, Hill SE, Nathan WJ, Paiano J, Callen E, Wang D, Shinoda K, van Wietmarschen N, Colón-Mercado JM, Zong D, De Pace R, Shih HY, Coon S, Parsadanian M, Pavani R, Hanzlikova H, Park S, Jung SK, McHugh PJ, Canela A, Chen C, Casellas R, Caldecott KW, Ward ME, Nussenzweig A. Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature*. 2021 May;593(7859):440-444. doi:

Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc Natl Acad Sci U S A*. 2003 Oct 28;100(22):12871-6. doi: 10.1073/pnas.2135498100. Epub 2003 Oct 17. PMID: 14566050; PMCID: PMC240711

Cao et al *Nature Communications* (2019) 10 5799

Muslimović A, Nyström S, Gao Y, Hammarsten O. Numerical analysis of etoposide induced DNA breaks. *PLoS One*. 2009 Jun 10;4(6):e5859. doi: 10.1371/journal.pone.0005859. Erratum in: *PLoS One*. 2009;4(6). doi: 10.1371/annotation/290cebfd-d5dc-4bd2-99b4-f4cf0be6c838. PMID: 19516899; PMCID: PMC2689654.

Baranello et al *Int J Mol Sci* (2014) 15 13111-1322

Baranello, L. et al. (2018). Mapping DNA Breaks by Next-Generation Sequencing. In: Muzi-Falconi, M., Brown, G. (eds) *Genome Instability. Methods in Molecular Biology*, vol 1672. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7306-4_13

Hu et al *Genes & Dev*. 2015. 29: 948-960

Sriramachandran et al *Mol Cell* (2020)78 975-985

- Ewels, P.A., Peltzer, A., Fillinger, S. et al. Nat Biotechnol 38, 276–278 (2020).
<https://doi.org/10.1038/s41587-020-0439-x>
- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available
5 online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Felix Krueger; Frankie James; Phil Ewels; Ebrahim Afyounian; Michael Weinstein; Benjamin Schuster-
Boeckler; Gert Hulselmans; scamons [Online] <https://doi.org/10.5281/zenodo.7598955>
- 10 MARTIN, Marcel. Cutadapt removes adapter sequences from high-throughput sequencing reads.
EMBnet.journal, [S.l.], v. 17, n. 1, p. pp. 10-12, may 2011. ISSN 2226-6089.
doi:<https://doi.org/10.14806/ej.17.1.200>.
- Heng Li , Richard Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform,
15 Bioinformatics, Volume 26, Issue 5, March 2010, Pages 589–595,
<https://doi.org/10.1093/bioinformatics/btp698>
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical
Computing, Vienna, Austria. URL <https://www.R-project.org/>
- 20 Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-
319-24277-4, <https://ggplot2.tidyverse.org>.

Table 4: Sequences

Name	Sequence (5' to 3')	notes
ssBreak_3prime	/5Phos/NNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3BioTEG/	
ssBreak_5prime	/5BioTEG/TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNN	
i7 index	CAAGCAGAAGACGGCATACGAGAT-[i7]-GTCTCGTGGGCTCGGAGATGT*G	i7 = 8nt index sequence
i5 index (Truseq - for 3' method)	AATGATACGGCGACCACCGAGATCTACAC-[i5]-ACACTCTTTCCCTACACGA*C	i5 = 8nt index sequence
i5 index (Nextera - for 5' method)	AATGATACGGCGACCACCGAGATCTACAC-[i5]-TCGTCGGCAGCGT*C	i5 = 8nt index sequence
N7_top	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	N7_top is annealed to N7_bottom
N7_bottom	/5Phos/CTGTCTCTTATACACATCT	N7_top is annealed to N7_bottom

Claims

1. A method of detecting single strand break (SSB) sites in double stranded (ds) DNA comprising;
(i) contacting a sample of dsDNA with an oligonucleotide probe comprising a first sequencing
5 adaptor and a tag,
(ii) contacting the sample of dsDNA with DNA ligase, such that the DNA ligase covalently links the oligonucleotide probe to a strand of the dsDNA at the site of a single strand break in the dsDNA,
(iii) contacting the sample of dsDNA with a transposase loaded with a second sequencing adaptor, such that the transposase cleaves the dsDNA to generate a population of fragments comprising a terminal
10 second sequencing adaptor,
(iv) isolating from said population, DNA strands that comprise the first and second sequencing adaptors and;
(v) determining the sequence of the isolated DNA strands.
- 15 2. A method according to claim 1 wherein the sequences of the isolated DNA strands are indicative of single strand breaks (SSBs) sites within the dsDNA.
3. A method according to any one of the preceding claims wherein the dsDNA is genomic DNA.
- 20 4. A method according to any one of the preceding claims wherein the dsDNA is in a cell nucleus or extract thereof.
5. A method according to any one of the preceding claims wherein the dsDNA is from a diseased cell; an embryonic cell; an age-associated cell; or a treated cell.
- 25 6. A method according to claim 5 wherein the treated cell has undergone a treatment selected from;
exposure to one or more compounds,
exposure to light or irradiation,
exposure to cell culture conditions,
30 exposure to a nuclease, optionally Cas9 or Cpf1, or
exposure to reprogramming factors, optionally Oct3/4, Sox2, Klf4 and c-Myc.
7. A method according to any one of the preceding claims wherein the sample of dsDNA is produced by a method comprising extracting the nucleus from a eukaryotic cell.
- 35 8. A method according to claim 7 wherein the eukaryotic cell is fixed and permeabilised.
9. A method according to any one of the preceding claims, wherein the sample of dsDNA can comprise one or more of the following steps prior to contacting the sample with the oligonucleotide probe:
40 (i) samples (e.g. cells) comprising dsDNA are subjected to wash and lysis,

(ii) said samples comprising dsDNA are subjected to buffer exchange e.g. by centrifugation of the sample or by dilution.

10. A method according to any one of the preceding claims, wherein the sample of dsDNA is purified by
5 proteinase digestion after contacting the dsDNA with the ligase.

11. A method according to any one of the preceding claims, wherein the isolation of the DNA strands that comprise the first and second sequencing adaptors is performed using magnetic beads.

10 12. A method according to any one of the preceding claims wherein the tag is biotin.

13. A method according to any one of the preceding claims wherein the probe further comprises a hybridisation region that anneals to the dsDNA at SSB sites.

15 14. A method according to claim 13 the hybridisation sequence comprises comprise 5-10 nucleotides.

15. A method according to any one of claims 13 or 14 wherein the hybridisation sequence is diverse.

16. A method according to claim 15 wherein the hybridisation sequence is random.
20

17. A method according to any one of claims 13 or 14 wherein the hybridisation sequence is complementary to a target sequence in the dsDNA

18. A method according to claim 14 wherein the target sequence is a promoter.
25

19. A method according to any one of the preceding claims wherein the oligonucleotide probe is ligated to the 3' end of the broken strand of the dsDNA at the SSB site.

20. A method according to claim 19 wherein the tag is located at the 3' end of the oligonucleotide probe
30 and the oligonucleotide probe further comprises a 5' phosphate group.

21. A method according to any one of claims 1 to 18 wherein the oligonucleotide probe is ligated to the 5' end of the broken strand of the dsDNA at the SSB site.

35 22. A method according to claim 21 wherein the tag is located at the 5' end of the oligonucleotide probe and the oligonucleotide probe further comprises a 3' OH group.

23. A method according to any one of the preceding claims wherein the transposase is Tn5 transposase.

40 24. A method according to any one of the preceding claims wherein fragments comprising the first and second sequencing adaptors are isolated by contacting the tag with an immobilised binding member.

25. A method according to claim 24 wherein the tag is biotin and the binding member is streptavidin.
26. A method according to claim 24 or claim 25 wherein the binding member is immobilised on a bead.
- 5 27. A method according to any one of the preceding claims comprising treating the isolated DNA strands with bisulfite.
- 10 28. A method according to claim 27 wherein the sequences of the bisulfite treated strands are compared to a reference genome to identify methylated cytosine residues in the isolated DNA strands.
29. A method according to any one of the preceding claims comprising amplifying the isolated DNA strands using sequencing primers to produce amplification products.
- 15 30. A method according to any one of the preceding claims comprising sequencing the isolated DNA strands or amplification products.
31. A method according to any one of the preceding claims further comprising generating a set of sequence reads of the isolated DNA strands or amplification products.
- 20 32. A method according to claim 31 comprising mapping the sequence reads in the set to one or more sites in a reference genome.
33. A method according to claim 32 comprising determining the frequency of SSBs at a site in the reference genome from the numbers of sequence reads mapped to the site.
- 25 34. A method according to claim 33 comprising mapping the frequency of SSBs at sites within a first dsDNA and a second dsDNA and identifying sites with increased frequency of SSB occurrence in the first dsDNA relative to the second dsDNA or in the second dsDNA relative to the first dsDNA.
- 30 35. A method according to claim 34 wherein first dsDNA is a test sample, and the second dsDNA is a reference or control sample.
36. A method according to claim 35 wherein the first dsDNA is from a disease cell; an embryonic cell; an age-associated cell; or a treated cell.
- 35 37. A method according to claim 36 wherein the treated cell has undergone a treatment selected from; exposure to one or more compounds, exposure to light or irradiation, exposure to cell culture conditions, exposure to a nuclease, optionally Cas9 or Cpf1, or
- 40

exposure to Oct3/4, Sox2, Klf4 and c-Myc.

38. A kit for detecting sites of single strand breaks (SSBs) in double stranded (ds) DNA; the kit comprising;
- 5 an oligonucleotide probe comprising a first sequencing adaptor and a tag,
a DNA ligase,
a transposase and
a second sequencing adaptor loaded or loadable onto the transposase.
- 10 39. A kit according to claim 38 comprising instructions for use in a method according to any one of claims 1 to 37.
40. A method of screening for an agent, such as a chemical and/or biological agent (e.g. a small molecule and/or a peptide and/or protein therapeutic) that can enable repair of SSBs in dsDNA, wherein said
- 15 screening method comprises the following steps:
- (a) detecting single strand break (SSB) sites in a control sample of double stranded (ds) DNA that has not been exposed to said agent and also detecting single strand break (SSB) sites in a test version of the same sample of the dsDNA that has been exposed to said agent by the steps of;
- (i) contacting a sample of the test and control dsDNA with an oligonucleotide probe
- 20 comprising a first sequencing adaptor and a tag,
- (ii) contacting the sample of test and control dsDNA with DNA ligase, such that the DNA ligase covalently links the oligonucleotide probe to a strand of the dsDNA at the site of a single strand break in the dsDNA,
- (iii) contacting the sample of test and control dsDNA with a transposase loaded with a
- 25 second sequencing adaptor, such that the transposase cleaves the dsDNA to generate a population of fragments comprising a terminal second sequencing adaptor, and then
- (b) isolating from said population, DNA strands that comprise the first and second sequencing adaptors,
- (c) determining the sequence of the isolated DNA strands from the test and control samples,
- 30 (d) comparing the profile of SSBs in both the test and control samples,
- (e) isolating the agent e.g. chemical and/or biological agent, that is associated with a reduction in SSBs in test sample compared to control samples, and
- (f) formulating said isolated agent obtained from (e) with pharmaceutically acceptable excipient(s).