

METHOD FOR CONSTRUCTING FREE DNA METHYLATION SEQUENCING LIBRARY AND USE THEREOF

Technical Field

The present application relates to biotechnology, and in particular to a method for epigenetic analysis of free DNA.

Background Art

Cell-free DNA (cfDNA) is generally a DNA fragment released from a cell into body fluid or the environment during cell metabolism, exocytosis, apoptosis, and necrosis. Plasma, serum, urine, saliva, amniotic fluid and other materials contain trace amounts of free DNA, which reflects information such as genetic variation, disease, and aging in the body. Extraction of free DNA, library construction, sequencing and analysis are widely used as an effective non-invasive/minimally invasive detection method in various aspects of clinical diagnosis and scientific research, such as cancer screening and diagnosis, prenatal diagnosis, preimplantation diagnosis, and aging research. Epigenetic information carried by free DNA, such as DNA methylomic/hydroxymethylomic characteristics and fragmentomic characteristics, are increasingly considered to be valuable molecular markers for detecting, diagnosing and/or monitoring various diseases such as cancer.

DNA methylation is a phenomenon in which methylation chemical modification occurs on DNA in organisms. The chemical essence of DNA methylation is that a hydrogen atom is replaced by a methyl group on the 5th carbon atom of a pyrimidine ring of a cytosine base to form 5-methylcytosine. This DNA modification is an epigenetic mechanism that affects many life processes such as gene expression and regulation, cell division and differentiation, cell metabolism and growth without changing the DNA sequence, ultimately affecting the growth and development, metabolic reproduction, disease occurrence and development, and aging and death of biological individuals. If high-throughput sequencing (next generation sequencing, NGS) of methylation modifications on DNA sequences is required, methylated/unmethylated cytosine must first be specifically converted into other bases to distinguish methylation modifications.

Fragmentomics information is a collection of information obtained by analyzing fragmentation

characteristics of free DNA, such as distribution characteristics of free DNA on a genome, fragment length and start and end positions, fragment end base characteristics, and fragment unevenness characteristics. As studies have shown that free DNA tends to be cleaved at certain genomic regions or elements, fragmentomics information is a type of key epigenomic information and a valuable molecular marker for monitoring and diagnosing various diseases such as cancer [1], which can reflect the states of chromatin opening and closing, nucleosome occupancy, transcription factor binding, epigenetic modification, and the like. If it is necessary to obtain the fragmentomics information during high-throughput sequencing, the integrity of DNA fragments must be preserved.

However, since free DNA is formed by strong DNA degradation inside and outside a cell, free DNA has the following characteristics that make it difficult to detect free DNA.

1. Low cfDNA content (for example, only a few nanograms of cfDNA per milliliter of plasma). This requires that the detection method must be adapted to a starting amount of cfDNA of 1 ng or even lower, and a higher library construction efficiency is required to improve the sensitivity of cfDNA detection.

2. Short cfDNA fragments (about 100 bp to 200 bp). First, short DNA fragments are easily lost during library construction due to purification operations, and cannot withstand secondary breaks caused by chemical conversion. Secondly, whole genome library construction methods such as transposase, random primers, and interruption and connecting cannot be applied to short DNA fragments, and there are also certain difficulties and limitations in applying targeted amplification library construction methods such as targeted PCR.

3. cfDNA often has severe chemical structural damage. The damage to cfDNA comes from multiple processes, including DNA damage caused by cell physiological and pathological conditions, damage caused by nuclease digestion during cell death, damage caused by continuous attack by active substances such as free nucleases in body fluid environment (such as blood circulation), damage caused by long-term storage in solutions such as plasma, and repeated freezing and thawing. Studies have found that 97% to 98% of cfDNA has gaps, which are single-strand phosphodiester bond break gaps generated on double-stranded DNA [2]. When using a single-strand library construction method, cfDNA breaks at the gaps after double-strand denaturation and dissociation, resulting in fragments [3].

Methods for constructing a DNA methylation sequencing library in the related art can be divided into the following four types: (1) a double-strand library construction method based on bisulfite conversion [4]; (2) a single-strand library construction method based on bisulfite conversion (such as Swift Biosciences kit Accel-NGS® Methyl-Seq DNA Library Kit, Cat. No. 30024); (3) a double-strand library construction method based on an enzymatic conversion method (such as New England Biolabs, Inc. EM-seq technology, WO2017075436A1 and a commercial kit E7120 thereof, TAPS technology of WO2019136413A1, and Cabemet technology of WO2021077415); and (4) a single-strand library construction method based on an enzymatic conversion method (such as CN114032287A). However, the above four methods cannot completely solve the difficulties in free DNA methylation detection.

The currently widely used DNA methylation detection method, namely the bisulfite conversion, requires DNA to react for several hours in a harsh environment of high salt (9M), high acid (pH = 5), and high temperature (50°C to 90°C). This will cause DNA denaturation and break, and most of the DNA will be degraded and lost during the conversion reaction. The large amount of DNA break and loss makes the bisulfite conversion unable to cope with the requirements of low starting amount library construction of cfDNA, and aggravates the short fragmentation and gap damage of cfDNA. Therefore, it is difficult to develop a highly sensitive free DNA methylation detection technology based on bisulfite conversion.

In recent years, DNA methylation sequencing technologies based on an enzymatic conversion method have gradually matured, such as EM-seq technology, TAPS technology, and Cabernet technology. The enzymatic conversion method abandons harsh chemical treatment processes, avoids DNA break and degradation during conversion, and has higher library construction efficiency, higher sensitivity, and higher library uniformity. However, the patents and products of enzymatic conversion methods in the related art cannot solve all the problems of free DNA methylation detection. Regarding the problem of a low starting amount of cfDNA, the minimum starting amount in EM-seq technology (WO2017075436A1) and TAPS technology (WO2019136413A1) are both 5 to 10 nanograms, which cannot perform high-sensitivity sequencing of free DNA. Cabemet technology (WO2021077415) can be used to construct a library with a low starting amount of DNA, and can even perform highly sensitive library construction on single cells, but it is impossible to use transposases on short

fragments of free DNA, especially for the problem of cfDNA gaps, for which Cabemet technology has not yet provided a solution.

Regarding the chemical damage of cfDNA, for sequencing in the related art that does not detect methylation, when double-strand library construction is used, gaps will be smoothed by an end repair process, which does not affect sequencing and has not attracted much attention. In bisulfite methylation sequencing in the related art, the chemical damage has not been taken seriously either. This is because although the use of single-strand library construction will cause cfDNA to break and be lost at the gaps, the methylation sequencing data will not be wrong. The conventional method based on a new enzymatic conversion method is double-strand library construction, and severe methylation erasure has indeed been observed, resulting in severe methylation loss. For example, data from the methylation sequencing library constructed by Erger et al. [5] using double-strand library construction showed that the average methylation level continued to decrease from about 80% to 40% at the beginning of library reads, and a general methylation level erasure was observed in subsequent Read2, but the authors did not discuss this anomaly in the article. In the DNA methylation sequencing library, a construction method, and a detection method of the DNA methylation sequencing library (CN114032287A) disclosed by Zhushi Biotechnology Co., Ltd., the authors observed the abnormality that the methylation level decreased with the library reads, but simply attributed the reason to the loss of end single strands. The authors used the single-strand library construction method to circumvent this problem, but did not really solve the problem. The break at the gaps in the single-strand state makes it impossible to accurately obtain fragmentomics information, and the single-strand ligation efficiency is not high, resulting in low library construction output. Therefore, to date, the academic community has not realized and solved the impact of the existence of gaps in the single-strand state on DNA methylation detection.

In summary, there is currently no library construction method that can comprehensively solve the problems of low template starting amount, short fragments, and serious gaps faced by free DNA methylation detection, and it is impossible to obtain a complete DNA library to accurately measure free DNA methylation and obtain fragmentomics information.

Summary of the Invention

After in-depth research on the characteristics of cfDNA and the principles of library construction, the inventors analyzed the causes of the problems existing in the cfDNA methylation detection process and made a series of targeted improvements, thereby providing a new cfDNA methylation sequencing method.

In one aspect, the present invention provides a method for pre-treating a biological sample for constructing a sequencing library, the method including:

S1: repairing a single-strand gap in a DNA fragment contained in the sample with a DNA ligase; and

S2: optionally, filling in an end gap on the DNA fragment with a DNA polymerase, in which the sequencing library is used to obtain DNA methylation information data of the biological sample or to obtain DNA methylation information and fragmentomics information data of the biological sample after sequencing.

In some embodiments, a length of the DNA fragment is between 50 bp and 1000 bp, preferably, between 100 bp and 500 bp, and more preferably, between 100 bp and 350 bp.

In some embodiments, the method further includes fragmenting DNA molecules contained in the biological sample before step S1 to produce the DNA fragment.

In some embodiments, the ligase has 3',5'-phosphodiester bond catalytic activity.

In some embodiments, the ligase is selected from the group consisting of HiFi-Taq ligase, T4 DNA ligase, Taq DNA ligase, E.coli DNA ligase, and T7 DNA ligase.

In some embodiments, the biological sample is selected from the group consisting of an animal fluid, a cell culture fluid, and a natural biological sample, which includes, but is not limited to, peripheral blood, plasma, serum, urine, feces, saliva, cerebrospinal fluid, lymph fluid, alveolar lavage fluid, amniotic fluid, blastochyle, cell culture fluid, embryo culture fluid, microbial culture medium, soil extract fluid, and bone meal extract fluid.

In some embodiments, the DNA fragment includes, but is not limited to, extracellular free DNA, extranuclear free DNA, and fragmented DNA, such as extracellular free DNA from a cancer subject.

In some embodiments, a content of the DNA fragment in the sample is not higher than 10 ng, or not higher than 1 ng, such as 0.1 ng.

In some embodiments, the DNA methylation includes, but is not limited to, methylation and/or

hydroxymethylation of cytosine.

In another aspect, the present invention provides a method for constructing a sequencing library, the method including:

1) providing a biological sample, and pre-treating the biological sample using the above pre-treatment method to obtain a DNA fragment without a 3',5'-phosphodiester bond gap; and

2) constructing the sequencing library using the DNA fragment without a 3',5'-phosphodiester bond gap, in which

the sequencing library is used to obtain DNA methylation information data of the biological sample or to obtain DNA methylation information and fragmentomics information data of the biological sample after sequencing.

In some embodiments, step 2) further includes

a) performing fill-in treatment on both ends of the DNA fragment and adding a linker sequence, preferably, part or all of cytosine in the linker sequence being methylated cytosine and/or hydroxymethylated cytosine; and

b) converting the cytosine in the DNA fragment into uracil by an enzymatic conversion method, the uracil being identified as thymine in subsequent amplification and sequencing, the methylated cytosine being identified as cytosine in subsequent amplification and sequencing.

In some embodiments, step 2) further includes

a) performing fill-in treatment on both ends of the DNA fragment and adding a linker sequence, and

b) converting the methylated cytosine in the end DNA fragment into uracil or dihydrouracil by an enzymatic conversion method, the uracil or dihydrouracil being identified as thymine in subsequent amplification and sequencing, unmethylated cytosine being identified as cytosine in subsequent amplification and sequencing.

In some embodiments, the enzymatic conversion method adopts an EM-seq conversion method or a TAPS conversion method.

In some embodiments, the method further includes adding carrier DNA to the biological sample after step a).

In some embodiments, the carrier DNA may be any DNA that does not include the linker

sequence, and preferably, a fragment size of the carrier DNA is 100 bp to 500 bp.

In another aspect, the present invention provides a sequencing library obtained by the above method.

In another aspect, the present invention provides a method for identifying a position of a single-strand gap in a DNA fragment contained in a biological sample, the method including:

- 1) dividing the biological sample into two parts, a biological sample A and a biological sample B;
- 2) treating the biological sample A using the above pretreatment method to prepare a first sequencing library, and treating the biological sample B using the above pretreatment method, without using the DNA ligase in the pre-treatment step, to prepare a second sequencing library; and
- 3) identifying the position of the single-strand gap based on a difference in DNA methylation information obtained from the first sequencing library and the second sequencing library.

In another aspect, the present invention provides a method for identifying a health condition of a subject, the method including:

- 1) providing a biological sample from the subject;
- 2) constructing a sequencing library using the above pre-treatment method; and
- 3) sequencing the sequencing library, or sequencing after hybridization capture enrichment, in which

DNA methylation information data of the biological sample or DNA methylation information data and fragmentomics information data of the biological sample are obtained according to sequencing data, and compared with normal DNA methylation information data and/or fragmentomics information data in a population to determine the health condition of the subject.

The above methods provided by the present invention can be used for epigenetic analysis of extracellular free DNA, and can also be used in the fields of genomics, medicine, diagnostics and epigenetic research.

Brief Description of Drawings

FIG. 1 is a schematic diagram of a process of a method for constructing a methylation library provided by the present invention.

FIG. 2 is a schematic diagram of a process of an EM-seq conversion method.

FIG. 3 is a schematic diagram of a process of a TAPS conversion method.

FIG. 4 shows the results of library methylation deviation analysis of methylation library construction of free DNA of human peripheral blood with DNA Ligase treatment method (solid line) and without DNA Ligase treatment method (dashed line), and along the 5' to 3' direction (horizontal axis) of each library fragment, the average value (vertical axis) of a CpG methylation rate thereof changes.

FIG. 5 shows a fragment distribution law of free DNA, in which there is a peak with a period of about 170 bp.

FIG. 6 shows the distribution of library fragments generated by different library construction methods. Capillary gel electrophoresis was performed with the same library DNA input amount to analyze the distribution of DNA fragments in the range of 1 bp to 6000 bp. The results show that the library of the method of the present invention (upper drawing) shows a periodic DNA fragment peak spectrum, which is consistent with the characteristics of free DNA of peripheral blood. However, the Swift Biosciences, Accel-NGS® Methyl-Seq DNA Library Kit (lower drawing) has only one main peak, no periodic peaks, and contains a large number of non-library impurity peaks.

Detailed Description

Unless defined otherwise, all technical and scientific terms used herein have the meanings as commonly understood by those skilled in the art.

A "DNA ligase" herein refers to an enzyme used to repair a single-strand gap in a double-stranded DNA molecule. The DNA ligase generally has 3',5'-phosphodiester bond catalytic activity. In one example, the DNA ligase can catalyze the formation of a phosphodiester bond between a 5' phosphate group at the gap and a 3' hydroxyl group. The DNA ligase generally does not add new nucleotides at the gap.

A "DNA polymerase" herein refers to an enzyme that can synthesize a complementary strand of a DNA strand by adding deoxyribonucleotide (dNTP) to a 3' end of a primer using the DNA strand as a template in the presence of the primer. The DNA polymerase generally also has 3'-5' exonuclease activity, which plays a proofreading role during synthesis, and 5'-3' exonuclease activity, which plays

an excision repair role.

A "DNA fragment" herein refers to a short DNA fragment, for example, a fragment with a length between 50 bp and 700 bp, for example, 100 bp to 500 bp, especially 100 bp to 350 bp. DNA fragments contained in a biological sample are usually heterogeneous, that is, the lengths are different, and accordingly, the above lengths may refer to an average length of these DNA fragments. These DNA fragments may have different sequences, for example, from different regions of genomes of the same organism, or even from different organisms. The DNA fragment may have a single-strand gap, or may have a blunt end or a non-blunt end (with a 3' overhang or a 5' overhang).

A "DNA fragment with a blunt end" herein refers to a double-stranded DNA fragment without 3' or 5' overhangs at the ends.

"DNA methylation" herein refers to the modification of cytosine bases in DNA molecules or DNA fragments to 5-methylcytosine (5mC). DNA methylation in vertebrates generally occurs at CpG sites (that is, sites in DNA sequences where cytosine is immediately followed by guanine), and cytosine is converted to 5-methylcytosine by DNA methyltransferase. Most CpG sites in the human genome are methylated, but in certain specific regions, such as CpG islands rich in cytosine (C) and guanine (G), CpG sites are usually not methylated. CpG methylation can affect the transcriptional activity of related genes. For example, methylation can inhibit tumor suppressor genes, while demethylation can stimulate the expression of certain oncogenes, all of which may lead to the occurrence of cancer. In addition, although the incidence is lower than 5mC, a small number of cytosine bases are modified to 5-hydroxymethylcytosine (5hmC), 5-formylation (5fC), and 5-carboxylation (5caC). When methylation is mentioned herein, it can also refer to modification to 5hmC, 5fC, and 5caC, unless the context indicates otherwise.

"DNA methylation information" herein refers to information about the methylation status in a DNA molecule or DNA fragment, including, but is not limited to, a methylation site, a methylation level, a methylation pattern (5mC or 5hmC), and the like. The "methylation level" herein can also be referred to as "methylation degree", which refers to the proportion (or frequency) of a specific methylation site in a sample that is methylated. A plurality of methods can be used to detect whether a site is methylated, and common methods include chemical or enzymatic conversion methods. During the conversion process, one of the methylated cytosine and the unmethylated cytosine is

converted to uracil (U) or a base that is substantially equivalent to uracil in terms of base pairing (for example, dihydrouracil, DHU). In subsequent amplification, the corresponding uracil is paired with adenine (A) as thymine (T), and the final result is that the cytosine or methylated cytosine at the methylation site is reflected as thymine in the test results (such as sequencing results). By comparing with a reference sequence, it is possible to determine whether the cytosine in a DNA molecule or DNA fragment is methylated. The reference sequence can be a sequence from the same sample but without the above conversion, or a corresponding sequence in a healthy population. In addition, as described below, some methods can be used to distinguish 5mC and 5hmC. Currently, DNA methylation information has been widely used in the screening and diagnosis of cancer (for example, lung cancer, breast cancer, liver cancer, and colorectal cancer), especially early screening and diagnosis.

"Fragmentomics information" herein refers to a collection of information obtained by analyzing fragmentation characteristics of free DNA in a sample, such as distribution characteristics of free DNA on a genome, fragment length and start and end positions, fragment end base characteristics, and fragment unevenness characteristics. Some literatures have recently reported the use of fragmentomics information for the screening of cancer, such as lung cancer.

An "enzymatic conversion method" herein refers to the modification of cytosine or methylated cytosine by means of the catalytic action of a specific enzyme, so that methylated and unmethylated states can be distinguished in subsequent detection, or 5mC and 5hmC can be distinguished. Some enzymatic conversion methods known in the related art include, for example, an EM-seq conversion method and a TAPS conversion method, but are not limited thereto. Enzymes are usually used in combination with some chemical reagents in the enzymatic conversion.

The "EM-seq conversion method" herein refers to a technology developed by New England Biolabs to distinguish methylated cytosine from unmethylated cytosine by enzymatic conversion, which uses dioxygenase TET (including TET1, TET2, TET3, and the like) to modify 5mC and 5hmC to 5-carboxylcytosine (5caC), and then uses cytidine deaminase to deaminate unmethylated cytosine and convert the unmethylated cytosine into uracil. The cytidine deaminase used includes members of the APOBEC protein family, such as APOBEC3A. In the case of using DNA glycosyltransferase (GT), 5mC and 5hmC can be further distinguished. The process of the EM-seq conversion method can be

seen in FIG. 2. For more detailed information about the EM-seq conversion method, see PCT application publication WO2017075436A, which is hereby incorporated by reference in its entirety.

The "TAPS (TET-assisted pyridine borane sequencing) conversion method" is somewhat similar to the EM-seq conversion method. It also modifies 5mC and 5hmC to 5-carboxylcytosine (5caC) by enzymatic methods, but does not perform a deamination step. Instead, a reducing agent pyridine borane is used to convert 5caC to dihydrouracil (DHU), while the unmethylated cytosine remains unchanged. In subsequent replication or amplification, DHU is identified as uracil by a polymerase and paired with A. In this way, sites with 5mC and 5hmC modifications are detected as T, and C without modification is still detected as C. Similarly, when DNA glycosyltransferase (GT) or KRuO_4 is used, 5mC and 5hmC can be further distinguished. The process of the TAPS conversion method can be seen in FIG. 3. For more detailed information about the TAPS conversion method, see PCT application publication WO2019136413A, which is hereby incorporated by reference in its entirety.

A "health condition" herein refers to the health status of a subject, including whether the subject has a certain disease, the risk of having a certain disease, whether the subject is suitable for a certain therapeutic agent or treatment, the prognosis of the disease, and the like.

A "subject" herein refers to an animal, such as a mammal, including, but is not limited to, humans, rodents, apes, felines, canines, equines, bovines, porcines, sheep, goats, mammalian experimental animals, mammalian farm animals, mammalian farm animals, and mammalian pets. The subject may be male or female and may be of any appropriate age, including infants, young children, adolescents, adults, and elderly subjects. In some examples, the subject is a patient. In particular examples, the subject is a human, such as a human patient. The term is often used interchangeably with "patient," "test subject," "treatment subject," and the like.

One aspect of the present invention provides a method for pre-treating a sample for sequencing library construction. The pretreatment includes repairing single-strand gaps in a DNA fragment contained in the sample, and then performing conventional end repair. A DNA ligase can be used to repair the gaps, and a DNA polymerase can be used for the end repair. The reason why the gap repair is performed before the conventional end repair is that the inventors first realized and confirmed through research that the existence of these single-strand gaps is an important reason for the loss of methylation information, especially for extracellular free DNA (cfDNA), such as circulating tumor

DNA (ctDNA). If only the conventional repair is performed, the DNA polymerase used will cause chain replacement downstream of the back gaps through its 5'-3' exonuclease activity and polymerase activity, resulting in the methylation information contained in the original strand being erased. If the ends of the DNA fragment are repaired after the gaps are repaired, there will be no strand displacement, and the original methylation information can be retained.

Another aspect of the present invention provides a method for constructing a sequencing library, which includes constructing a sequencing library after the sample is subjected to the above pretreatment. In some embodiments, the library construction also includes modifying cytosine (C) or methylated cytosine (including 5mC and/or 5hmC) therein by an enzymatic conversion method, so that methylated and unmethylated cytosine can be distinguished or 5mC and 5hmC can be distinguished in the subsequent detection step to construct a DNA methylation sequencing library. The corresponding methylation information data can be obtained after sequencing the sequencing library. Those skilled in the art could understand that by using the method of the present invention, not only methylation information data can be obtained after sequencing, but also fragmentomics information data can be obtained at the same time. This is because the library construction technology in the present invention is based on the construction of methylation NGS detection library based on cfDNA double-strand fragments, and the methylation status is distinguished by enzymatic conversion after repairing single-strand gaps, which basically does not damage the integrity of the original DNA fragments in the sample, and retains its original fragmentomics information, including fragment length distribution characteristics, fragment length, and the like.

In another aspect of the present invention, the requirement of the method of the present invention on a content of the DNA fragment in the sample can be further reduced by using carrier DNA, until the DNA content is as low as 0.1 ng, or even lower.

In another aspect of the present invention, a position of the single-strand gap in the DNA fragment contained in the biological sample is determined by a difference in methylation information caused by adding or not adding a DNA ligase. It is understandable that 5mC or 5hmC is not usually present exactly at the gap position. Therefore, in most cases, the gap position can be considered as a gap position range. For example, the gap position can be considered to be between a first position and a second position downstream of the first position. The first position is a methylation position that can

be detected with or without adding a DNA ligase, and the second position is a position where a methylation difference is detected (for example, methylation can be detected by adding a DNA ligase, but methylation cannot be detected without adding a ligase).

In another aspect of the present invention, a health condition of a subject can be judged by the obtained methylation information, and in particular, the methylation information can be combined with the fragmentomics information to more accurately judge the health condition of the subject.

It is understandable that the technical problems to be solved by the present invention include but are not limited to the following.

1. The provision of a method for constructing a high-throughput second-generation sequencing library and NGS sequencing for free DNA, which can accurately determine DNA methylation and fragmentomics information at the same time.
2. The method needs to have high sensitivity, and can efficiently construct NGS libraries for samples with a low starting amount as low as 1 ng and below.
3. The method needs to preserve the integrity of free DNA strands to ensure that fragmentomics information in sequencing libraries is not destroyed.
4. It is necessary to avoid DNA breakage and loss of methylation information caused by gap damage on free DNA.
5. After a simple change in the technical route, the determination of any methylation information such as DNA hydroxymethylation and fragmentomics information can be achieved.

The method provided by the present invention can be used for, but is not limited to, epigenetic analysis of extracellular free DNA, and can achieve the above objects. The method may include one or more of the following features.

1. A specific DNA ligase is used to avoid methylation loss caused by DNA gap damage. For DNA with gaps, in the method provided by the present invention, one or several specific DNA ligases, such as Taq DNA ligase, are used to repair phosphodiester bond gaps on double strands of template DNA to prevent the breakage and loss of DNA strands and/or methylation information from being erased in subsequent processes. After a DNA template treated with gap repair is subjected to double-strand library construction by applying an enzymatic conversion method, there will be no serious erasure of methylation information at a 3' end of the library, and the detected methylation rate will be restored

to near the average methylation rate.

2. Carrier DNA is used to assist in protecting trace DNA, so as to achieve effective library construction and highly sensitive methylation sequencing of trace DNA. For this method, which is improved based on an enzymatic conversion method, see the Cabernet technology of WO2021077415. After adding a library linker to a DNA template, carrier DNA with similar physical and chemical properties but without linker sequences is added before the enzymatic conversion. In purification and other steps where DNA loss or degradation is likely to occur, the carrier DNA becomes the main lost part, which can effectively avoid the loss of the DNA template. In addition, after the conversion, a single-strand purification step is omitted and a PCR amplification reaction system is directly added to minimize the loss of the template DNA caused by the purification process.

3. Bisulfite conversion will lead to the loss of fragmentomics information. However, according to the method provided by the present invention, enzymatic conversion and double-strand DNA library construction strategy are applied to accurately determine DNA methylation information while retaining DNA fragmentomics information, and does not affect the retention of the fragmentomics information when DNA ligases are used for treatment.

Method Overview

(1) Extraction of free DNA

Free DNA refers to DNA that is free outside cells or cell nucleus, and can be extracted from various body fluids, in vitro cell culture fluids, natural environment and other biological materials, including, but is not limited to, peripheral blood, plasma, serum, urine, feces, saliva, cerebrospinal fluid, lymph fluid, alveolar lavage fluid, amniotic fluid, blastochyle, cell culture fluid, embryo culture fluid, microbial culture medium, soil extract fluid, and bone meal extract fluid. Free DNA can be obtained by extraction and purification.

The total amount of free DNA used for downstream sequencing analysis can be as low as picogram or nanogram levels.

(2) Repair of DNA gaps

In this step, a specific ligase is used. The specific ligase may be, but is not limited to, a combination of any one or several repair reagents and/or DNA ligases such as HiFi-Taq ligase, T4

DNA Ligase, Taq DNA Ligase, T7 DNA Ligase, and E.coli DNA ligase.

The DNA ligase and/or the DNA repair reagent are used to repair gaps on double strands of template DNA to prevent the breakage and loss of DNA strands and/or methylation information from being erased in subsequent processes.

(3) Addition of library linker sequence to two ends of DNA

The gaps at the ends of the double strands are filled using a polymerase or the like. In some examples, A (adenine) bases are added to the ends to facilitate the connection of linkers.

In this step, a DNA ligase is used to connect a DNA linker with a specific sequence containing partial or all cytosine methylation modification or hydroxymethylation modification to two ends of the template DNA.

In some examples, the DNA linker sequence can be labeled with biotin to facilitate subsequent purification.

In some examples, sample barcodes and single DNA molecule barcodes can be introduced into the DNA linker sequence to label the source of the DNA template, increase the detection accuracy, and allow samples from multiple sources to be mixed in subsequent steps.

(4) Addition of Carrier DNA

In order to prevent the loss of trace amounts of template DNA in subsequent steps, carrier DNA with similar properties that is several times the amount of the template DNA is added to the template DNA solution. The carrier DNA can become the main lost part in steps such as purification where DNA loss or degradation is likely to occur, thereby effectively avoiding the loss of the template DNA.

Since carrier DNA does not carry a DNA linker sequence with a specific sequence, it will not be detected in subsequent library amplification and sequencing.

(5) Methylation conversion

In this step, enzymatic conversion is used to specifically convert cytosine, methylated cytosine and/or hydroxymethylated cytosine on DNA, so that base pairing rules of cytosine and methylated and/or hydroxymethylated cytosine are changed, corresponding to different bases in subsequent amplification and sequencing, thereby being distinguished.

In this step, the "EM-seq" technology invented by New England Biolabs can be used. In another example, the TAPS technology can be used.

(6) Library amplification and sequencing

During methylation conversion, DNA undergoes thermal denaturation and changes in base pairing rules, becoming single-stranded or partially single-stranded. Single-stranded DNA has weak affinity during the purification process, resulting in a large amount of loss. This method can omit the single-strand purification step and directly add a PCR amplification reaction system to minimize the loss of template DNA during the purification process.

The library is subjected to PCR amplification reaction, and an amplification primer contains sequences complementary to the above linker sequences and linker sequences of a sequencer. Sample barcodes can also be introduced. The primer can complement the linker sequences added to two ends of the template DNA to form the desired sequencing library during amplification.

After purification, the sequencing library is quantified and quality checked, and then subjected to sequencing by the sequencer.

(7) Analysis of sequencing results

In the example of methylation conversion using EM-seq technology, cytosine is converted to thymine, and methylated cytosine is retained as cytosine. In the example of methylation conversion using TAPS technology, methylated cytosine is converted to thymine, and cytosine remains unchanged. DNA methylation information can be obtained by aligning with a reference genome. Since the library linker is added before methylation conversion and when free DNA remains in a double-stranded state in the method provided by the present invention, the DNA fragment information is retained. By aligning with the reference genome, the original DNA fragment length, fragment distribution, upstream and downstream break positions, end base pattern and other fragmentomics information of the free DNA can be obtained. For the above analysis process, see FIG. 1.

Other Use of Method of Present Invention

DNA hydroxymethylation and fragmentomics sequencing of free DNA

In mammals, a content of DNA hydroxymethylation is lower than that of DNA methylation, but it has different biological significance.

In the above methylation conversion process, DNA methylation and hydroxymethylation are not distinguished. If only hydroxymethylation needs to be measured, 5mC and 5hmC can be distinguished,

for example, it is only necessary not to add TET oxidase during the conversion process and retain other enzymatic reactions.

Targeted PCR-based enrichment method

Enriching target fragments of interest can save sequencing costs and improve the detection sensitivity of the target gene.

In step (6) of the above method overview, PCR is performed using primers complementary to the linkers at the two ends of the template DNA to amplify the entire template DNA. If it is necessary to enrich the target gene fragment, a targeted PCR method can be used for amplification. Targeting primers containing sequences complementary to a targeted gene fragment after methylation conversion are designed and introduced to the PCR reaction system to amplify the targeted gene sequence.

The targeting primers may or may not contain part or all of the linker sequences that are compatible with the sequencer.

Only an upstream or downstream primer may be introduced to obtain a fragment having a targeting primer at one end and the linker sequence added in step (3) of the above method overview at the other end.

The upstream or downstream primer can also be connected at the ends and constructed on a piece of DNA to form a locked primer.

A default linker sequence primer can be added first, and then the targeting primer can be added after amplification. The default linker primer and the targeting primer can also be added at the same time. It is also possible not to add the default linker sequence primer and only use the targeting primer.

Hybridization capture-based enrichment method

In step (6) of the above method overview, hybridization capture enrichment can be performed before or after amplification. It is necessary to design a probe containing sequences complementary to the targeted gene fragment after methylation conversion.

In the hybridization capture enrichment, a single-stranded DNA/RNA probe labeled with biotin or the like is mixed with the library. After thermal denaturation and renaturation, the library complementary to the target sequences will hybridize with the probe. The hybridized probe can be captured using a method such as magnetic beads with streptavidin to obtain an enriched library.

Obtaining gap position information without gap repair

If the ligase treatment in step (2) of the above method overview is not performed, and a suitable polymerase is used for repair, or a modified nucleotide is used, the methylation information downstream of a gap will be changed. The position of the gap can be obtained by analyzing the sequence where methylation is erased/changed.

The present invention discloses a highly sensitive DNA methylation sequencing technology that can be used for DNA with single-strand gaps. The technology first repairs DNA gap damage to ensure that DNA methylation signals are not lost or misinterpreted. Secondly, the technology uses double-stranded DNA as a template to construct a library, overcoming the low efficiency of single-strand library construction. After that, a greatly optimized and adjusted enzymatic conversion method is used to perform conversion of base pairing rules for methylation information, and the template DNA is protected by carrier DNA during the reaction and purification process. Finally, a high-quality, highly sensitive methylation information sequencing library is obtained using an extremely low starting amount of DNA as a template, while retaining the fragmentomics information of the DNA.

The advantages of the method of the present invention include but are not limited to the following.

1. The minimum starting amount of DNA is as low as 0.1 ng or even lower.
2. The library yield can be increased by 4 times.
3. The library alignment rate is increased from 40% to 70%.
4. The length of DNA methylation loss at the 3' end of the library is reduced from 170 bp to less than 40 bp.

The method of the present invention can be applied to various liquid biopsy diagnostic products such as early cancer screening, cancer gene detection, prenatal diagnosis, preimplantation diagnosis, and genetic diagnosis, and also provides strong technical support for scientific research in related fields.

The present invention is further illustrated by specific examples below.

Example 1:

This example describes a sequencing library preparation process using peripheral blood cfDNA as an example.

Description

1.1 Extraction of cfDNA

The extraction of cfDNA can be performed by any standard means in the related art.

A 5 mL EDTA venipuncture vacuum blood collection tube (purple cap) was used to collect 5 mL of peripheral blood, and plasma separation was performed within 4 hours by low-speed centrifuging the sample at room temperature at 1600g using a horizontal rotor for 10 minutes, carefully removing the upper-layer plasma and transferring it to a 1.5 mL tube, performing high-speed centrifuging at 4°C at 6000g for 10 minutes, and transferring the supernatant to a fresh 1.5 mL tube, and storing it in a refrigerator at -80°C or on dry ice. A nucleic acid extraction or purification kit (QIAGEN, QIAamp® Circulating Nucleic Acid, catalog number: 55114) was used to extract cfDNA from the separated plasma, and 5 ng to 20 ng cfDNA was usually obtained from 1 ml plasma.

1.2 cfDNA gap repair

A ligase reaction system was configured. Here Taq DNA Ligase was used as an example. The reagents in Table 1 below were added to the purified cfDNA solution. The Sonicated Spike-in DNA was an equal volume mixture of whole genome CpG methylated pUC19 DNA (NEB E7122) and whole genome CpG non-methylated lambda DNA (NEB E7123), and was ultrasonically sheared to 300 bp and then diluted to 0.2 ng/μL for later use.

Table 1 Ligase premix composition

Ligase premix	Volume μL
H ₂ O	6.3
Ligase 10x reaction buffer	2.5
Sonicated Spike-in DNA	0.2
ligase	1
	10

After brief mixing and centrifugation, the reaction was carried out on a PCR instrument at 37°C to 60°C for 10 minutes to 30 minutes.

1.3 End repair and linker connection

An End Prep reaction system was configured. To the repaired product from the previous step, the reagents of a NEBNext Ultra II End Repair/dA-Tailing Module (NEB E7546L) kit were added in order (Table 2).

Description

Table 2 End repair reagents

Reagent name	Volume/ μ L
End Prep reaction buffer	3.5
End Prep Enzyme	1.5
Total volume μ L	5

After brief mixing and centrifugation, the reaction was carried out on the PCR instrument at 20°C for 30 minutes and 65°C for 30 minutes.

After the reaction, linker sequences modified with all cytosines methylated were added. To complement the primers for subsequent library construction, a pair of recommended linker sequences were 5'-A[5mC]A[5mC]T[5mC]T T T[5mC][5mC][5mC]T A[5mC]A[5mC]G A[5mC]G[5mC]T[5mC]T T[5mC][5mC]G A T[5mC]T-3' (SEQ ID NO: 1) and 5'-phosphate-G A T[5mC]G G A A G A G[5mC]A[5mC]A[5mC]G T[5mC]T G A A[5mC]T[5mC][5mC]A G T[5mC]A-3' (SEQ ID NO: 2). The primers were diluted to a total concentration of 15 μ M, and 1.25 μ L was added and mixed.

Then, the following reagents of the kit NEBNext Ultra II Ligation Module (NEB E7595L) were added in order (Table 3).

Table 3 Addition of linker sequence reagents

Reagent name	Volume/ μ L
Ligation enhancer	0.5
Ultra II Ligation	15
Total volume μ L	15.5

After brief mixing and centrifugation, the reaction was carried out on the PCR instrument at 20°C for 15 minutes.

1.4 Addition of carrier DNA and purification

To the connected product from the previous step, 1 μ L of a carrier DNA was added and then mixed briefly and centrifuged.

The carrier DNA was prepared by ultrasonically shearing lambda DNA (NEB N3011) to 300 bp and then diluting it to 25 ng/ μ L for later use.

For DNA purification, 1.8 volumes of SPRI or Ampure XP magnetic beads were used, and the

purified DNA was eluted with water.

1.5 Enzymatic conversion of DNA methylation modification

The purified DNA obtained in the previous step was subjected to enzymatic conversion using the NEBNext® Enzymatic Methyl-seq Conversion Module (NEB, E7125L) kit. This includes TET oxidation reaction and glycosylation protection of DNA, DNA thermal denaturation, and APOBEC deamination reaction. After the APOBEC deamination reaction, the product was directly subjected to the next step without purification.

1.6 Library PCR amplification

PCR primers with library linkers and index tags were added to the liquid after the APOBEC deamination reaction in the previous step. For example, 5 µL of a pair of primers of NEBNext® Multiplex Oligos for Illumina was added. Then, an equal volume of 2x Q5U master mix (NEB M0597L) was added to the liquid. After brief mixing and centrifugation, the following was carried out on the PCR instrument.

98°C for 30 seconds; 7 cycles of 98°C for 10 seconds, 62°C for 30 seconds, and 72°C for 90 seconds; and 65°C for 5 minutes. The product was stored at 4°C.

For DNA purification, 1.1 volumes of SPRI or Ampure XP magnetic beads were used, and the purified DNA was eluted with 1x TE buffer and can be stored at -20°C.

1.7 Library sequencing and analysis

After the concentration of the constructed library was quantified and the fragment distribution was checked, double-end sequencing was performed on an Illumina sequencer. The sequencing data was aligned and analyzed using Bismark software (to avoid bases with uneven methylation rate distribution, the methylation rates of the first 10 bases of Read1 and the first 40 bases of Read2 were not included in the methylation analysis results) to obtain the whole genome methylation data. After bioinformatics analysis, the distribution characteristics of the library on the genome, the fragment length and start and end positions, and the 5' end base distribution characteristics are fragmentomics information.

Table 4 shows the sequencing results using the following reads as an example.

Read number	A00682:655:HKFCVDSX2:2:1205:32524:18490_1:N:0:CTCAGAA+A ACTTGC
-------------	---

Description

Sequences obtained by sequencing:	TTTTAAAAATATTTAAGTAAAGTAGAGAATGTAGAAATGTTATA GATTATATTTTTTGATTATGATATAATAAAATTAGAAATTATAGT ATGGAAATTTAAAAGTTTTTTTTTAAATAATTTTATGTTAAAAA GAAATTTAGGTCGG GT (SEQ ID NO: 3) and ATGTTTGTAATTTTAGTATTTTGGGAGGTTAAGGTGGGTAGGTT ATTTGAGGTTAGTAGTTTAAGATTAGTTTCGTTAATATGGTAAT ATTTTGTTTTTATTAAAAATATAAAAATTAGTTGGGTTTGGTGG TTTATGTTTGTAATTT TA (SEQ ID NO: 4)
Methylation characteristics:	hhh.....h.h...h.....h.x.....h.x.....h.h.h...h.....x.h.....h...h.h.....h.....h.....hx...xZ.... and ...hx.....hx...h.....hh.....x...h.h.....x.x...h...hx..hh.Z..h.h..... h.hhx...h.h..h.....h.....x...hx.....h.h...hx.....hhx Here, "." represents a non-cytosine base, z represents the cytosine in the CpG site, h and x represent the cytosine in the non-CpG site, lowercase letters represent that the cytosine is not methylated, and uppercase letters represent that the cytosine is methylated. Methylation was detected at all CpG sites in this read.
Starting position aligned to the genome:	Position 100915 on chromosome 1 (reference genome hg38 was used).
End position aligned to the genome:	Position 101074 on chromosome 1 (reference genome hg38 was used).
Fragment length:	159 bp
5' end base:	CCC

The sequencing results include tens to billions of pieces of the above information, and the sequencing volume varies depending on the selection of sequencing instruments and chips.

Example 2 Comparison of DNA gap repair

To evaluate whether the method of the present invention can overcome the methylation loss defect of the enzymatic conversion library construction method in the related art, a parallel control comparison test was performed.

Free DNA was extracted from the plasma of a healthy person and divided into two groups for experiments:

Experimental group: treated with a DNA ligase (see step 1.2: DNA gap repair reaction in Example

1)

Control group: not treated with a DNA ligase (as a control group, water was used instead of the DNA ligase)

The two groups of free DNA were subjected to the same double-strand library construction steps:

The reagent NEBNe.t Ultra II End Repair/dA-Tailing Module (NEB E7546L) was used for end modification, and the reagent NEBNe.t Ultra II Ligation Module (NEB E7595L) and the reagent NEBNe.t® EM-seqTM Adaptor (NEB E7165) were used for linker connection.

After adding the linkers, the template DNA was purified by magnetic beads and enzymatically converted using the NEBNe.t® Enzymatic Methyl-seq Conversion Module (NEB, E7125L) kit. The converted DNA was subjected to PCR library amplification using 2.Q5U master mi. (NEB M0597L).

The constructed library was subjected to double-end 150 bp sequencing on the Illumina sequencer, and Bismark software was used for aligning and analysis. The library methylation bias (M-bias) was statistically analyzed, and the average methylation rate of each reading position was plotted along the 5' to 3' direction of the library fragment to evaluate whether the methylation level of the library was average. The results are shown in FIG. 4.

As illustrated, no gap repair was performed on free DNA. After methylation double-strand library construction, the average methylation level of the DNA in the sequencing library started from about 130 bp of Read1 and extended along the 5' to 3' direction of the library fragment to Read2, resulting in increasingly significant methylation loss (dashed line as illustrated), which was particularly obvious in the Read2 sequencing data, and the 40% methylation rate with real value was lost on average. The experimental group was treated with DNA ligase repair (solid line as illustrated), the methylation loss on the Read2 was repaired (the decrease in the methylation rate of 30 bp at the 3' end of the library is due to the loss of the 3' end of the free DNA double-stranded template).

This data shows that the method provided by the present invention can overcome the loss of methylation caused by free DNA gap damage.

Example 3 Fragmentomics information can be obtained while obtaining methylation information

To evaluate the superiority of the method of the present invention over the methods in the related

Description

art, the technology of the present invention and a commercial kit were subjected to a control test.

Experimental group: Technical method of the present invention. See Example 1 for specific operations.

Control group: Swift Biosciences (kit manufacturer and catalog number: 33024 Accel-NGS® Methyl-Seq DNA Library Kit). See the kit operation manual for specific operations.

After extracting free DNA from the plasma of a healthy person, two different methods were used to construct libraries with the same free DNA starting amount (3 ng) and the same number of PCR amplification rounds (7 rounds). After sequencing and analyzing the libraries, the Bismark software was used to evaluate the alignment efficiency. The results are shown in Table 5.

Table 5 Comparison of library construction between the technical method of the present invention and the Swift method

	Technical method of the present invention	Swift method
Library yield	83.4 ng	21.15 ng
Library alignment rate	70%	44%
Library duplication rate	15%	22%
Fragment distribution	With nucleosome cycle peak	No nucleosome cycle peak

Specifically, the DNA output after library construction is quantified by Qubit. Compared with the Swift kit, the library yield of the technical method of the present invention is 83.4 ng, which is much higher than the library yield of the Swift method. The alignment rate of the sequencing library of the technical method of the present invention is 70%, which is nearly twice as high as the 44% of the swift library, indicating that the library is effective. The duplication rate of the sequencing library constructed by the method of the present invention is 15%, which is significantly lower than the duplication rate of 22% of the Swift library, indicating that the library contains fewer duplicate fragments and has better quality. According to the method of the present invention, the fragmentomics information is retained (see FIGS. 5 and 6). The Agilent Fem to Pulse automated electrophoresis fragment analysis system was used to analyze the DNA fragments before library construction, and the results are shown in FIG. 5. Free DNA itself has a characteristic peak spectrum with a period of about 170bp, which is a typical fragmentation characteristic caused by nucleosome occupancy. The Agilent 5200 Fragment Analyzer System was used to analyze the fragments after library construction,

and the results are shown in FIG. 6. Compared with the library constructed by the Swift kit, the sequencing library constructed by the library construction and sequencing method provided by the present invention contains fragmentation distribution of free DNA characteristics, which can better analyze and obtain fragmentomics data.

The present invention points out for the first time that the serious methylation loss problem caused by double-strand library construction is due to single-strand gaps and end single-strand loss. The end single-strand loss is the loss of original information and cannot be recovered, while single-strand gaps cause the methylation of most reads to be erased. The principle is that in the double-strand library construction step, the double-stranded DNA template needs to be end-repaired first to form double-stranded DNA with a blunt end or a 3'-A sticky end, and then a linker sequence is connected. The DNA synthase used in the end repair has the activity of extension from a gap. Starting from the gap position, the DNA synthase displaces or enzymatically hydrolyzes the original downstream DNA template chain along the 5' to 3' direction of the DNA, and synthesizes a replacement chain with artificially provided unmodified bases, resulting in the erasure of continuous DNA methylation information. This methylation erasure cannot be restored by bioinformatics and other means in the later stage. Due to the randomness of the gap position, it is impossible to use consistent standards to accurately determine the effective methylation information part. The pre-treatment method provided by the present invention repairs the gap before library construction, solving the problem of large-scale methylation loss (FIG. 4).

References:

1. Lo, Y.M.D., Han, D.S.C., Jiang, P.&Chiu, R.W.K.Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science* 372, (2021).
2. Sanchez, C., Roch, B., Mazard, T., Blache, P., Dache, Z., Pastor, B., Pisareva, E., Tanos, R.,&Thierry, A.R.(2021).Circulating nuclear DNA structural features, origins, and complete size profile revealed by fragmentomics. *JCI insight*, 6(7), e144561. <https://doi.org/10.1172/jci.insight.144561>.
3. (Sanchez, C., Snyder, M.W., Tanos, R., Shendure, J.,&Thierry, A.R.(2018).New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA

analysis. NPJ genomic medicine, 3, 31. <https://doi.org/10.1038/s41525-018-0069-0>).

4. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 2008;133:523-36.

5. Erger, F., Nörling, D., Borchert, D., Leenen, E., Habbig, S., Wiesener, M.S., Bartram, M.P., Wenzel, A., Becker, C., Toliat, M.R., Nürnberg, P., Beck, B.B., & Altmüller, J. (2020). cfNOME-A single assay for comprehensive epigenetic analyses of cell-free DNA. Genome Medicine, 12(1), 54. <https://doi.org/10.1186/s13073-020-00750-5>.

Claims

1. A method for pre-treating a biological sample for constructing a sequencing library, the method comprising:

S1: repairing a single-strand gap in a DNA fragment contained in the sample with a DNA ligase; and

S2: filling in an end gap on the DNA fragment by a DNA polymerase, wherein

the sequencing library is used to obtain data of DNA methylation information of the biological sample or to obtain data of both DNA methylation information and fragmentomics information of the biological sample after sequencing, and

the single-strand gap in step S1 is a 3',5'-phosphodiester bond gap.

2. The method according to claim 1, wherein the length of the DNA fragment is between 50 bp and 1,000 bp.

3. The method according to claim 1, further comprising: fragmenting DNA molecules contained in the biological sample before step S1 to produce the DNA fragment.

4. The method according to claim 1, wherein the ligase has 3',5'-phosphodiester bonding catalytic activity.

5. The method according to claim 1 or 4, wherein the ligase is selected from the group consisting of HiFi-Taq ligase, T4 DNA ligase, Taq DNA ligase, *E.coli* DNA ligase, and T7 DNA ligase.

6. The method according to claim 1, wherein the biological sample is selected from the group consisting of a cell culture fluid and a natural biological sample.

7. The method according to claim 1, wherein the biological sample is selected from the group consisting of peripheral blood, plasma, serum, urine, feces, saliva, cerebrospinal fluid, lymph fluid, alveolar lavage fluid, amniotic fluid, blastochyle, cell culture fluid, embryo culture fluid, microbial culture medium, soil extract fluid, and bone meal extract fluid.

8. The method according to claim 7, wherein the DNA fragment is selected from the group consisting of extracellular free DNA and fragmented DNA.

9. The method according to claim 1, wherein the content of the DNA fragment in the sample is not higher than 10 ng.

10. The method according to claim 1, wherein the DNA methylation includes methylation and/or hydroxymethylation of cytosine.

11. A method for constructing a sequencing library, the method comprising:

1) providing a biological sample, and pre-treating the biological sample by the method according to any one of claims 1 to 10 to obtain a DNA fragment without a 3',5'-phosphodiester bond gap; and

2) constructing the sequencing library by using the DNA fragment, wherein the sequencing library is used to obtain data of DNA methylation information of the biological sample or to obtain data of both DNA methylation information and fragmentomics information of the biological sample after sequencing.

12. The method according to claim 11, wherein step 2) further includes

a) performing fill-in treatment on both ends of the DNA fragment, and adding a linker sequence in which a part or all of cytosines are methylated cytosines and/or hydroxymethylated cytosines, and

b) converting the cytosine in the DNA fragment into uracil by an enzymatic conversion method, wherein the uracil is identified as thymine in subsequent amplification and sequencing, and the methylated cytosine is identified as cytosine in subsequent amplification and sequencing, or

a) performing fill-in treatment on both ends of the DNA fragment and adding a linker sequence, and

b) converting the methylated cytosine in the DNA fragment into uracil or dihydrouracil by an enzymatic conversion method, wherein the uracil or dihydrouracil is identified as thymine in subsequent amplification and sequencing, and unmethylated cytosine is identified as cytosine in subsequent amplification and sequencing.

13. The method according to claim 12, wherein the enzymatic conversion method includes an EM-seq conversion method or a TAPS conversion method.

14. The method according to claim 12, further comprising: adding a carrier DNA to the biological sample after step a).

15. The method according to claim 14, wherein the carrier DNA does not contain the linker sequence, and a fragment size of the carrier DNA is 100 bp to 500 bp.

16. A method for identifying the position of a single-strand gap in a DNA fragment contained in a biological sample, the method comprising:

1) dividing the biological sample into two parts, biological sample A and biological sample B;

2) treating biological sample A by the method according to any one of claims 1 to 10 to prepare a first sequencing library, and treating biological sample B by the method according to any one of claims 1 to 10 without using the DNA ligase in step S1 to prepare a second sequencing library; and

3) identifying the position of the single-strand gap based on a difference in DNA methylation information obtained from the first sequencing library and the second sequencing library, wherein the single-strand gap is a single-strand 3',5'-phosphodiester bond gap.