(19) DANMARK

(10) DK/EP 3090068 T3

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: *C 12 Q 1/68 (2006.01)*

(45) Oversættelsen bekendtgjort den: **2019-08-26**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2019-05-29**

(86) Europæisk ansøgning nr.: **14877282.5**

(86) Europæisk indleveringsdag: **2014-12-16**

(87) Den europæiske ansøgnings publiceringsdag: **2016-11-09**

(86) International ansøgning nr.: **CN2014093940**

(87) Internationalt publikationsnr.: **WO2015101175**

(30) Prioritet: **2013-12-31 CN 201310756037**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **Berry Genomics Co., Ltd., Room 801, Floor 8, Building No. 5, Yard 4 , Shengmingyuan Road , Science Park , Changping District, Beijing 102206, Kina**

(72) Opfinder: **WANG, Lina, No. 9 Building, No. 6 Jingshun East Street, Chaoyang District, Beijing 100015, Kina**
RU, Lanlan, No. 9 Building, No. 6 Jingshun East Street, Chaoyang District, Beijing 100015, Kina
LI, Tiancheng, No. 9 Building, No. 6 Jingshun East Street, Chaoyang District, Beijing 100015, Kina
ZHANG, Yaxi, No. 9 Building, No. 6 Jingshun East Street, Chaoyang District, Beijing 100015, Kina
ZHANG, Jianguang, No. 9 Building, No. 6 Jingshun East Street, Chaoyang District, Beijing 100015, Kina

(74) Fuldmægtig i Danmark: **Plougmann Vingtoft A/S, Strandvejen 70, 2900 Hellerup, Danmark**

(54) Benævnelse: **FREMGANGSMÅDE OG KIT TIL IKKE-INVASIV DETEKTERING AF EGFR-GENMUTATIONER**

(56) Fremdragne publikationer:
WO-A1-2013/138510
CN-A- 101 921 830
CN-A- 101 967 684
CN-A- 102 181 533
CN-A- 102 851 375
CN-A- 103 298 955
US-A1- 2013 309 724
MARK T. GREGORY ET AL: "Targeted single molecule mutation detection with massively parallel sequencing", NUCLEIC ACIDS RESEARCH, 17 September 2015 (2015-09-17), page gkv915, XP055241840, ISSN: 0305-1048, DOI: 10.1093/nar/gkv915

DESCRIPTION

Field of Invention

[0001] The present invention relates to genetic diagnosis field. More specifically, the present invention is directed to a method for detecting Epidermal Growth Factor Receptor (EGFR) gene mutations, as well as a kit for use in the detection of EGFR gene mutations.

Background of Invention

[0002] EGFR (Epidermal Growth Factor Receptor) is normally embedded in the cell membrane on the cell surface. EGFR gene is 118kb in length, includes 28 exons, and encodes a 170kD glycoprotein comprised of 1186 amino acids. It is a membrane protein that plays an important role in regulating proliferation, growth, repair and survival of tumor cells. Currently, targeted therapy has become an important tool in the clinical treatment of Non Small Cell Lung Cancer (NSCLC). Iressa (Gefitinib, AstraZeneca) and Tarceva (Erlotinib, Roche), which function as EGFR tyrosine kinase inhibitors (TKI), are the main drugs approved by FDA for NSCLC targeted therapy. However, clinical experiments showed that Iressa and Tarceva only had significant therapeutic effects on 10-30% NSCLC patients. Further studies indicated that EGFR gene mutations relate to the therapeutic effects of NSCLC targeted therapy, and most of the patients carrying EGFR gene mutations showed significant therapeutic effects. A large number of research documents indicated that EGFR gene mutations are mainly located in the tyrosine kinase coding domain (exons 18-21), wherein deletion in exon 19 (746-753) accounts for about 45% of all mutations, and substitution in exon 21 (mainly L858R) accounts for about 40% of all mutations. At present, it is generally believed that these two hot mutations can enhance the sensitivity of tumor cells against TKI, and can be used as an effective index to predict TKI treatment. Therefore, the detection of EGFR gene mutations has an important reference value for guiding clinical administration in patients with NSCLC. Methods for detecting EGFR currently used in clinical treatment include: 1) traditional sequencing. This method has high accuracy. However, high requirement on the sample source, long sequencing time, the need for sequence analysis, and high cost of this method limit its use in clinic. 2) Polymerase chain reaction-single strand conformation polymorphism (PCR-SSCP). It is a classical method for detecting gene mutations, and can detect unknown mutations. It has advantages of simple operation and high sensitivity. However, it also has an obvious disadvantage, which is the requirement for parallel standard control. Additionally, this method has a high false positive: the detection rate is 75-95% when the tested PCR product is less than 200bp, while the detection rate is only about 50% when the tested PCR sample is more than 400bp. 3) Mutation enriched PCR: a two-step PCR using a restriction endonuclease to selectively digest the wild type EGFR gene. After the first PCR, the wild type EGFR is digested selectively and the mutated EGFR genes are enriched, and then the second PCR is conducted. The PCR product is detected by electrophoresis, and whether EGFR is mutated is determined based on the detection results of

PCR product. This method is highly sensitive, and can detect one mutation among 10^3-10^4 wild type EGFRs. However, this method needs twice PCR and enzyme digestion, thus is complex and time-consuming. In addition, there are technologies like AMRS and micro digital PCR, but their application in clinic still needs time. WO2013138510 detects mutations in EGFR, using plasma DNA (example 1, page 28 and following). WO2013138510 further uses tagged primers for a first amplification and a further amplification (see also Figure 1). These tagged amplicons were then used in massive parallel sequencing to analyse the mutation.

[0003] Thus, there is an urgent need for a fast and efficient method for detecting EGFR gene mutations clinically. The inventor found, during the research of fragment DNA detection, a new method for detecting DNA fragments, including cyclizing and then amplifying DNA fragments. Based on this discovery and in combination with the second generation high throughput sequencing technology, the inventor improved the aforementioned method and designed optimized primers specifically based on the EGFR genes, and developed a method and a kit thereof for sequencing and analyzing EGFR gene mutations in plasma DNAs.

Summary of Invention

[0004] In one aspect, the present invention provides a method for non-invasively detecting EGFR gene mutations in subjects, comprising the following steps:

--designing primers according to EGFR gene exons;

--extracting plasma DNAs from the subjects;

--connecting the extracted plasma DNAs with tagging linkers;

--PCR pre-amplifying the tagging linkers connected plasma DNAs;

--cyclizing the amplified DNAs to obtain cyclised DNAs;

--PCR amplifying the cyclised DNAs using the designed primers; and

--high throughput sequencing the PCR amplified product and analyzing the EGFR gene mutations.

[0005] Further, the primers are a pair of primers that are adjacent and backward extended.

[0006] Further, primers of the backward extended primer pair are located on 5' or 3'end of the EGFR gene exons.

[0007] Further, space of the backward extended primer pair is 0-1/2 of the total base pairs of the fragment DNAs.

[0008] Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes.

[0009] Further, 5' end of the backward extended primers contains linker sequences for high throughput sequencing library.

[0010] Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes, and sequences of the primers are as follows:

E18-1-3F: CCCAGCTTGTGGAGCCTC

E18-1-3R: GACAAGAACACAGAGACAAGGGT

E18-2-F: GCAGGGCCTCTCATGGTC

E18-2-R: CCTGTGCCAGGGACCTTAC

E19-1-F: ACGTCTTCCTTCTCTCTCTGTCAT

E19-1-R: GTGAGATGGTGCCACATGCT

E19-2-F: GTCCATGGCTCTGAACCTCA

E19-2-R: CCACACAGCAAAGCAGAAAC

E20-1-F: CACACTGACGTGCCTCTCC

E20-1-R: CTTCGCATGGTGGCCAGA

E20-2-1F: CCTCCCCGTATCTCCCT

E20-2-1R: GGAGATAAGGAGCCAGGAT

E21-1-F: AGCAGGGTCTTCTCTGTTTCA

E21-1-R: GAGGGACAGATCATCATGGG

E21-2-F: TTTCCTGACACCAGGGACC

E21-2-R: TGACCTAAAGCCACCTCCTT.

[0011] Further, the cyclization is a splint mediated single strand DNA cyclization.

[0012] Further, the EGFR genes in plasma DNAs have insertion, deletion, substitution or gene fusion mutations.

[0013] Further, the high throughput sequencing technologies are selected from Roche/454

FLX, Illumina/Hiseq/Miseq, Applied Biosystems SOLID and life Technologies/Ion Torrent/Proton.

[0014] In another aspect, the present invention provides a kit for non-invasively detecting EGFR gene mutations, comprising: reagents for extracting plasma DNAs, a DNA cyclase, primers and reagents for amplifying target DNAs.

[0015] Further, the kit comprises primers and reagents for pre-amplifying regions to be tested in the EGFR genes.

[0016] Further, the kit comprises reagents for high throughput sequencing.

[0017] Further, the reagents for high throughput sequencing are applicable to the following high throughput sequencing technologies: Roche/454 FLX, Illumina /Hiseq/Miseq, Applied Biosystems SOLID and life Technologies/Ion Torrent/Proton.

[0018] Further, the primers for amplifying regions to be tested in the EGFR genes are a pair of primers that are adjacent and backward extended.

[0019] Further, primers of the backward extended primer pair are located on 5' or 3' end of the sites or regions to be tested in the EGFR genes.

[0020] Further, space of the backward extended primer pair is 0-1/2 of total base pairs of the plasma DNAs.

[0021] Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes.

[0022] Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes, and sequences of the primers are as follows:

E18-1-3F: CCCAGCTTGTGGAGCCTC

E18-1-3R: GACAAGAACACAGAGACAAGGGT

E18-2-F: GCAGGGCCTCTCATGGTC

E18-2-R: CCTGTGCCAGGGACCTTAC

E19-1-F: ACGTCTTCCTTCTCTCTCTGTCAT

E19-1-R: GTGAGATGGTGCCACATGCT

E19-2-F: GTCCATGGCTCTGAACCTCA

E19-2-R: CCACACAGCAAAGCAGAAAC

E20-1-F: CACACTGACGTGCCTCTCC

E20-1-R: CTTCGCATGGTGGCCAGA

E20-2-1F: CCTCCCCGTATCTCCCT

E20-2-1R: GGAGATAAGGAGCCAGGAT

E21-1-F: AGCAGGGTCTTCTCTGTTTCA

E21-1-R: GAGGGACAGATCATCATGGG

E21-2-F: TTTCCTGACACCAGGGACC

E21-2-R: TGACCTAAAGCCACCTCCTT.

[0023] Further, the plasma DNA linkers contain tagging sequences.

[0024] Further, the plasma DNAs are pre-amplified before they are cyclised.

[0025] Further, the cyclization is a splint mediated single strand DNA cyclization.

[0026] Further, the sites or regions to be tested in the EGFR genes have insertion, deletion, substitution or gene fusion mutations.

[0027] In yet another aspect, the present invention provides a use of primers against EGFR gene exons in the preparation of diagnosing reagents or kits for non-invasively detecting EGFR gene mutations in subjects, characterized in that the diagnosing reagents or kits are applicable to a method for non-invasively detecting EGFR gene mutations in subjects comprising the following steps:

--extracting plasma DNAs from the subjects;

--connecting the extracted plasma DNAs with tagging linkers;

--PCR pre-amplifying the tagging linkers connected plasma DNAs;

--cyclizing the amplified DNAs to obtain cyclised DNAs;

--PCR amplifying the cyclised DNAs using the designed primers; and

--high throughput sequencing the PCR amplified product and analyzing the EGFR gene mutations.

[0028] Further, the primers are a pair of primers that are adjacent and backward extended.

[0029] Further, primers of the backward extended primer pair are located on 5' or 3' end of the EGFR gene exons.

[0030] Further, space of the backward extended primer pair is 0-1/2 of total base pairs of the fragment DNAs.

[0031] Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes.

[0032] Further, 5' end of the backward extended primers contains linker sequences for high throughput sequencing library.

[0033] Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes, and sequences of the primers are as follows:

E18-1-3F: CCCAGCTTGTGGAGCCTC

E18-1-3R: GACAAGAACACAGAGACAAGGGT

E18-2-F: GCAGGGCCTCTCATGGTC

E18-2-R: CCTGTGCCAGGGACCTTAC

E19-1-F: ACGTCTTCCTTCTCTCTCTGTCAT

E19-1-R: GTGAGATGGTGCCACATGCT

E19-2-F: GTCCATGGCTCTGAACCTCA

E19-2-R: CCACACAGCAAAGCAGAAAC

E20-1-F: CACACTGACGTGCCTCTCC

E20-1-R: CTTCGCATGGTGGCCAGA

E20-2-1F: CCTCCCCGTATCTCCCT

E20-2-1R: GGAGATAAGGAGCCAGGAT

E21-1-F: AGCAGGGTCTTCTCTGTTTCA

E21-1-R: GAGGGACAGATCATCATGGG

E21-2-F: TTTCCTGACACCAGGGACC

E21-2-R: TGACCTAAAGCCACCTCCTT.

[0034] Further, the cyclization is a splint mediated single strand DNA cyclization.

[0035] Further, the EGFR genes in the plasma DNAs have insertion, deletion, substitution or gene fusion mutations.

[0036] Further, the high throughput sequencing is selected from Roche/454 FLX, Illumina/Hiseq/Miseq, Applied Biosystems SOLID and life Technologies/Ion Torrent/Proton.

Brief Description of Figures

[0037]

Fig. 1 shows experimental design of the present invention. The plasma DNA is first pre-amplified to increase yield of the detection product, and then the pre-amplified DNA is cyclised to increase available templates for detection. The EGFR gene is amplified to prepare a library for high throughput sequencing.

Fig. 2 shows design principle of the cyclization of the present invention. The pre-amplified DNA is a double strand DNA with universal linkers on both ends. Splint with reverse complimentary sequences is designed based on the sequences of the universal linkers on both ends, and a partial double strand is formed after annealing and hybridization. Single strand cyclised product is finally obtained by ligation using Tap ligase. Fig. 2A shows a general scheme, and Fig. 2B shows the cyclization principle.

Fig. 3 shows cyclization rate of the present invention. This figure shows the imaging results of 6% denaturing gel. 1: a PCR product before cyclization; 2: a product after cyclization (not digested by exonuclease), M: a marker; 3: an exonuclease digested product after cyclization (not purified); 4: an exonuclease digested and purified product after cyclization.

Fig. 4 shows the final EGFR sample library. This figure shows the results of 2% agarose gel. 1: LC2; 2: LC3; 3: LC113; 4: LC314; 5: LC320; 6: NTC(a negative control). The sample is a Lung Cancer (LC) sample.

Embodiments

[0038] With improvements in the sequencing technology, the traditional Sanger sequencing has been unable to fully meet the requirement of research. Thus, the second generation sequencing technology with lower cost, higher throughput, faster speed, and having capability for sequencing the whole genome emerges. The main principle of the second generation sequencing technology is high throughput sequencing by synthesis, namely, determining DNA sequences by capturing newly synthesized end labelling. The available technologic platforms

mainly include Roche/454 FLX, Illumina/Genome Analyzer/Hiseq/Miseq, Applied Biosystems SOLID, life Technologies/Ion Torrent, and the like. Taking Illumina product as an example, HiSeq 2000 can reach a sequencing throughput of 30X coverage in 6 human genomes per run, i.e., about 600 G/run, and the operation time is reduced to 30 minutes. Furthermore, with the maturity of the second generation sequencing technology, the investigation on its clinical application has developed quickly. Research shows that fetus genetic health can be judged by sequencing maternal plasma DNAs, and that sequencing plasma DNAs in subjects can be used for early cancer screening, which will have a wide application in the future.

[0039] Plasma DNAs, also known as circulating DNAs, are extracellular DNAs in the blood and is tens to hundreds of nucleotides in length (a main peak of about 167bp). It presents in a form of DNA-protein complex, or as free DNA fragments. Normally, plasma DNAs are derived from DNA release of a small amount of senescent and dead cells. Under healthy condition, the generation and removal of the circulating DNAs are in a dynamic equilibrium and are maintained at a relatively steady low level. 1mL plasma from a normal person contains about 2000 genomic DNAs. The circulating DNAs can reflect the metabolic condition of cells in human body, thus is an important index for judging health. The change of quantity and quality of circulating DNAs in peripheral blood is closely related to several diseases (including tumour, complex severe traumata, organ transplantation, pregnancy-related diseases, infectious diseases, organ failure, and the like). As a non-invasive detection index, it is expected to be an important molecular marker for early diagnosing some diseases, monitoring the conditions, and evaluating the therapeutic effects and prognosis of the diseases. For example, researches show that EGFR regulates the cell cycle progress, repair and survival of tumorcells, and at the same time relates to tumor metastasis. Recently, molecular targeted therapy using EGFR as the therapy target has received widespread attention from cancer communities both at home and abroad, and an EGFR tyrosine kinase inhibitor, Iressa (Genfitinib), has been approved by America Food and Drug Administration (FDA) for treating advanced NSCLC. The prominent feature of molecular targeted drugs is that its therapeutic effect strongly depends on the target: the therapeutic effect is significant strong in patients with the "target", while the therapeutic effect is weak or none in patients without the "target",which thus delays other treatments and makes the conditions worse. Therefore, blind administrations without target detection not only may result in high economic loss, but also may delay the valuable timing for treatment, or even aggravate the conditions. It is curtail to judgequickly and accurately whether the patient has the specific target for the targeted drug treatment.The traditional EGFR detection mainly detects the lesion tissue section by FISHor qPCR. However, it is found that there are more free DNAs in the plasma of NSCLC patients, about 10 times of those of normal people. A large quantity of free DNAs in plasma derives from DNAs release of senescent and dead tumorcells. They are similar to tumor genomic DNAs in genetic characteristics, and mutations thereof include deletion, point mutation, and increased copy number. EGFR gene mutations can be detected by examining plasma DNAs in NSCLC patients, and thus make it possible to detect EGFR expression non-invasively. The present invention detects EGFR expressions and mutations in plasma DNAs by the second generation sequencing technology quickly, accurately, non-invasively and with high sensitivity, and thus provides various diagnosis basis for patients.

[0040] In view of the clinical significance of non-invasive detection by plasma DNAs sequencing and the rapid development of the second generation high throughput sequencing, the inventor found that sequencing plasma DNAs in large-scale can detect EGFR gene expressions and mutations more quickly, accurately, and non-invasively. It is applicable to a variety of second generation high throughput sequencing platforms, including but not limited to, Roche/454 FLX, Illumina/Genome Analyzer/Hiseq/Miseq, Applied Biosystems SOLID, life Technologies/Ion Torrent, and the like.

[0041] The present invention is based on the following two facts: 1) plasma free DNAsin patients are similar to genomic DNAs in genetic characteristics. The plasma free DNAs of patients are higher in content than those of normal people, and often contain lots of mutations, while each mutation may be of low frequency; 2)The second generation high throughput sequencing can obtain the information of plasma free DNAs quickly, accurately, and with high throughput. Combining these two facts enables the non-invasive large scale application of the detection in genome specific regions. Researches show that plasma DNAs exist as fragments with low amount (1mL plasma contains about 2000 genomes) and short length (mainly about 167bp), which make it difficult for the traditional PCR to enrich mutations effectively using plasma DNAs as templates, resulting in rapid decrease in detection sensitivity. The present invention differs from the traditional methods in that the DNA fragments connected with sequence tagging linkers are amplified and single strand cyclized, then by means of back-to-back primer amplification, the templates are used maximally and the library is sequenced by high throughput paired-end sequencing. The original amplified templates are assembled based on the original sequencing analysis, and the tagging sequences are recorded. Sequences with the same position on the genome and the same tagging sequences are calculated as one template. The number of templates amplified by every primer pair is calculated and the number of mutated template is counted and recorded. The present invention improves the cyclization method and optimizes primers for the EGFR genes. The unique design of tagging sequences reduces background and prevents contamination effectively. The unique template set is counted by restoring templates in systems accurately, and thus a single molecule detection with high accuracy is finally achieved.

[0042] According to one specific embodiment of the present invention, it provides a method for non-invasively detecting EGFR gene mutations in subjects, comprising the following steps: (1) designing primers according to EGFR gene exons; (2)extracting plasma DNAs from the subjects; (3)connecting the extracted plasma DNAs with tagging linkers; (4)PCR pre-amplifying the tagging linkers connected plasma DNA; (5)cyclizing the amplified DNAs to obtain cyclised DNAs; (6)PCR amplifying the cyclised DNAs using the designed primers; and (7)high throughput sequencing the PCR amplified product and analyzing the EGFR gene mutations. "Non-invasive detection" in the present invention means that in comparison with that routine histology detection methods such as surgery, tissue biopsy and the like, which are directly against cancer tissues, will result in body damages in subjects, the present invention only detects blood sample from the subjects. Traditional methods for detecting DNAs or gene fragments require PCR amplification of the regions to be tested before detection, and thus the DNAs or gene fragments to be tested should be complete. However, most of the plasma DNA

fragments are incomplete, and thus the DNA fragments that can be used as templates in PCR amplification is few in number and is difficult to be detected by a routine PCR. Therefore, PCR amplification in the present invention adopts DNA cyclllization technology to transform fragment DNAs into cyclic DNAs using linker sequences and enzymes. Primers based on the regions to be tested are designed; sequencing library is amplified and constructed, and then is sequenced by a high throughput sequencing technology; and the EGFR gene mutations are analyzed. Figure 1 shows principle of the present invention. Plasma DNA is first extracted and connected with tagging linkers, which is in Y-shape. The double strand region contains tagging sequences that are distinguishable between different templates. To increase the amount of amplificable DNA fragments, pre-amplification can be conducted, universal sequence on the primers and linkers can be annealed, and then the pre-amplified DNAs are cyclised, for example, using a splint. Figure2 shows DNA cyclization principle of the present invention. Splint sequence and the pre-amplified primer region are complementary. Primers according to EGFR exons, for example, exon 18, 19, 20 and 21, are designed, and specific PCR amplification according to EGFR genes is conducted to generate a sequencing library, which is then sequenced by the current high throughput sequencing technology to obtain the specific regions in the tested EGFR genes, namely, gene sequences of the exons, and then gene mutations in the specific region are analyzed. The high throughput sequencing of the present invention refers to the available second generation sequencing techonology, such as Roche/454 FLX, Illumina/Genome Analyzer/Hiseq/Miseq, Applied Biosystems SOLID, life Technologies/Ion Torrent and the like. Different platforms differ in sequencing principles. Taking IlluminaMiseq as an example, the generated sequencing library is denatured and high throughput sequenced in a Miseq sequencer. The read length is 250bp on both ends, and every primer site produces about 100, 000 sequences. The double-end sequences are combined to a single end sequence based on terminal repeated regions. Linker regions in the single end sequence are searched, and the sequence is restored to the original plasma DNA fragments, while the tagging sequence in every sequence is recorded. The restored DNA fragments are compared with human genome, with the start and terminal coordinates of the sequence on the genome recorded. A unique template sequence is obtained after redundancy removal using the start and terminal coordinates and tagging sequences as standards. The type and ratio of mutations are determined by counting the template sequences in the target zone.

[0043] According to another further specific embodiment of the present invention, primers of the backward extended primer pair are located on 5' or 3' end of the EGFR gene exons. Further, space of the backward extended primer pair is 0-1/2 of total base pairs of the fragment DNAs. Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes. Further, 5' end of the backward extended primers contains linker sequences for high throughput sequencing library. Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of EGFR genes, and sequences of the primers are as follows:

E18-1-3F: CCCAGCTTGTGGAGCCTC

E18-1-3R: GACAAGAACACAGAGACAAGGGT

E18-2-F: GCAGGGCCTCTCATGGTC

E18-2-R: CCTGTGCCAGGGACCTTAC

E19-1-F: ACGTCTTCCTTCTCTCTCTGTCAT

E19-1-R: GTGAGATGGTGCCACATGCT

E19-2-F: GTCCATGGCTCTGAACCTCA

E19-2-R: CCACACAGCAAAGCAGAAAC

E20-1-F: CACACTGACGTGCCTCTCC

E20-1-R: CTTCGCATGGTGGCCAGA

E20-2-1F: CCTCCCCGTATCTCCCT

E20-2-1R: GGAGATAAGGAGCCAGGAT

E21-1-F: AGCAGGGTCTTCTCTGTTTCA

E21-1-R: GAGGGACAGATCATCATGGG

E21-2-F: TTTCCTGACACCAGGGACC

E21-2-R: TGACCTAAAGCCACCTCCTT.

[0044] Further, the cyclization is a splint mediated single strand DNA cyclization.

[0045] Further, the high throughput sequencing technologies are selected from Roche/454 FLX, Illumina /Hiseq/Miseq, Applied Biosystems SOLID and life Technologies/Ion Torrent/Proton. Illumina technology is used in the present invention.

[0046] According to another specific embodiment of the present invention, it provides a kit for non-invasively detecting EGFR gene mutations, comprising: reagents for extracting plasma DNAs, a DNA cyclase, primers and reagents for amplifying target DNAs. Conventional reagents or commercially available kits can be used in the extraction of plasma DNAs. Further, the amplification primers of the regions to be tested in the EGFR genes are a pair of primers that are adjacent and backward extended. Further, primers of the backward extended primer pair are located on 5' or 3' end of the sites or regions to be tested in theEGFR genes. Further, space of the backward extended primer pair is 0-1/2 of total base pairs of the plasma DNAs. Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes. Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes, and sequences of the primers are as follows:

E18-1-3F: CCCAGCTTGTGGAGCCTC

E18-1-3R: GACAAGAACACAGAGACAAGGGT

E18-2-F: GCAGGGCCTCTCATGGTC

E18-2-R: CCTGTGCCAGGGACCTTAC

E19-1-F: ACGTCTTCCTTCTCTCTCTGTCAT

E19-1-R: GTGAGATGGTGCCACATGCT

E19-2-F: GTCCATGGCTCTGAACCTCA

E19-2-R: CCACACAGCAAAGCAGAAAC

E20-1-F: CACACTGACGTGCCTCTCC

E20-1-R: CTTCGCATGGTGGCCAGA

E20-2-1F: CCTCCCCGTATCTCCCT

E20-2-1R: GGAGATAAGGAGCCAGGAT

E21-1-F: AGCAGGGTCTTCTCTGTTTCA

E21-1-R: GAGGGACAGATCATCATGGG

E21-2-F: TTTCCTGACACCAGGGACC

E21-2-R: TGACCTAAAGCCACCTCCTT.

[0047] Further, the kit comprises primers and reagents for pre-amplifying the regions to be tested in the EGFR genes. Specifically, the reagents and primers for pre-amplification include Taq DNA polymerase and its buffer, and primers for pre-amplification that are complementary to the Y-shape linkers.

[0048] Further, the kit comprises reagents for high throughput sequencing. Further, the reagents for high throughput sequencing are applicable to the following high throughput sequencing technologies: Roche/454 FLX, Illumina /Hiseq/Miseq, Applied Biosystems SOLID and life Technologies/Ion Torrent/Proton. Further, the plasma DNA connection linkers contain tagging sequences. Further, the plasma DNAs are pre-amplified before they are cyclised. Further, the cyclization is a splint mediated single strand DNA cyclization.

[0049] According to yet another specific embodiment of the present invention, it provides a use of primers according to EGFR gene exons in the preparation of diagnosing reagents or kits for non-invasively detecting EGFR gene mutations in subjects, characterized in that the diagnosing reagents or kits are applicable to a method for non-invasively detecting EGFR gene mutations in subjects comprising the following steps:

--extracting plasma DNAs in subjects;

--connecting the extracted plasma DNAs with tagging linkers;

--PCR pre-amplifying the tagging linkers connected plasma DNAs;

--cyclizing the amplified DNAs to obtain cyclised DNAs;

--PCR amplifying the cyclised DNAs using the designed primers; and

--high throughput sequencing the PCR amplified product and analyzing the EGFR gene mutations.

[0050] Further, the primers are a pair of primers that are adjacent and backward extended. Further, primers of the backward extended primer pair are located on 5' or 3' end of the EGFR gene exons. Further, space of the backward extended primer pair is 0-1/2 of total base pairs of the fragment DNAs. Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes. Further, 5' end of the backward extended primers contains linker sequences for high throughput sequencing library. Further, the backward extended primers aim at exon 18, exon 19, exon 20, or exon 21 of the EGFR genes, and sequences of the primers are as follows:

E18-1-3F: CCCAGCTTGTGGAGCCTC

E18-1-3R: GACAAGAACACAGAGACAAGGGT

E18-2-F: GCAGGGCCTCTCATGGTC

E18-2-R: CCTGTGCCAGGGACCTTAC

E19-1-F: ACGTCTTCCTTCTCTCTCTGTCAT

E19-1-R: GTGAGATGGTGCCACATGCT

E19-2-F: GTCCATGGCTCTGAACCTCA

E19-2-R: CCACACAGCAAAGCAGAAAC

E20-1-F: CACACTGACGTGCCTCTCC

E20-1-R: CTTCGCATGGTGGCCAGA

E20-2-1F: CCTCCCCGTATCTCCCT

E20-2-1R: GGAGATAAGGAGCCAGGAT

E21-1-F: AGCAGGGTCTTCTCTGTTTCA

E21-1-R: GAGGGACAGATCATCATGGG

E21-2-F: TTTCCTGACACCAGGGACC

E21-2-R: TGACCTAAAGCCACCTCCTT.

[0051] Further, the cyclization is a splint mediated single strand DNA cyclization. Further, the high throughput sequencing technologies are selected from Roche/454 FLX, Illumina /Hiseq/Miseq, Applied Biosystems SOLID and life Technologies/Ion Torrent/Proton.

Examples

Example 1

[0052] The plasma DNA template was amplified using self-designed linkers and according to a method for constructing a plasma DNA high throughput sequencing library (that is, PCR pre-amplification using phosphorylated primers after linker connection). The PCR product was purified by a gel cutting and cyclised by a splint connection. The cyclised product was digested by Exo III, purified, and screened by multiplex PCR with 8 pairs of back-to-back primers (the primers contain universal sequences for constructing a sequencing library). The mutation sites should be close to the forward primer or the reverse primer. The library was finally obtained by purifying the amplification product amplified by the universal primers.

1. Linker design. It is annealed to a double strand, wherein X is a tagging sequence:

ssCycADT-1:	GTCTCATCCCTGCGTGXXXXT
ssCycADT-2:	pXXXXCACGCAGGGTACGTGT

The structure of connection product:

Top:	GTCTCATCCCTGCGTGXXXXTNNN AXXXXCACGCAGGGTACGTGT
Bottom:	TGTGCATGGGACGCACXXXXA NNN TXXXXGTGCGTCCCTACTCTG

Primers:

ssCycUniprimer-F:	pGTCTCATCCCTGCGTG
ssCycUniprimer-R:	pACACGTACCCTGCGTG

The library structure after pre-amplification:

pGTCTCATCCCTGCGTGXXXXTNNN	AXXXXCACGCAGGGTACGTGT
CAGAGTAGGGACGCACXXXXA NNN	TXXXXGTGCGTCCCATGCACAp

Back-to-back primers for amplification in target zones:

EX0N18 (123bp)

CAAGTGCCGTGTCCTGGCACCCAAGCCCATGCCGTGGCTGCTGGTCCCCCTGCTGGGCCATGTCTGGCACTGCTTTCCAGCATGGTG

AGGGCTGAGGTGACCCTTGTCTCTGTGTTCTTGTCCCCCCCAG*CTTGTGGAGCCTCTTACACCCAGTGGAGAAGCTCCCAACCAAGC*

*TCTCTTGAGGATCTTGAAGGAAACTGAATTCAAAAAGATCAAAGTGCTGGGCTCCGGTGCGTTCGGCACGGTGTATAAG*GTAAGGTC

CCTGGCACAGGCCTCTGGGCTGGGCCGCAGGGCCTCTCATGGTCTGGTGGGGAGCCCAGAGTCCTTGCAAGCTGTATATTTCCATCA
TCTACTTTACTCTTTGTTTCACTGAGTGTTTGG

E18-1-3F:	CCCAGCTTGTGGAGCCTC
E18-1-3R:	GACAAGAACACAGAGACAAGGGT
E18-2-F:	GCAGGGCCTCTCATGGTC
E18-2-R:	CCTGTGCCAGGGACCTTAC

EX0N19 (99bp)

GCAATATCAGCCTTAGGTGCGGCTCCACAGCCCCAGTGTCCCTCACCTTCGGGGTGCATCGCTGGTAACATCCACCCAGATCACTGG
GCAGCATGTGGCACCATCTCACAATTGCCAGTTAACGTCTTCCTTCTCTCTCTGTCATAG*GGACTCTGGATCCCAGAAGGTGAGAAA*
*GTTAAAATTCCCGTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAACAAGGAAATCCTCGATG*TGAGTTTCTGCTTT
GCTGTGTGGGGGTCCATGGCTCTGAACCTCAGGCCCACCTTTTCTCATGTCTGGCAGCTGCTCTGCTCTAGACCCTGCTCATCTCCA
CATCCTAAATGTTCACTTTCTATG

E19-1-F:	ACGTCTTCCTTCTCTCTCTGTCAT
E19-1-R:	GTGAGATGGTGCCACATGCT
E19-2-F:	GTCCATGGCTCTGAACCTCA
E19-2-R:	CCACACAGCAAAGCAGAAAC

EX0N20 (186bp)

CCATGAGTACGTATTTTGAAACTCAAGATCGCATTCATGCGTCTTCACCTGGAAGGGGTCCATGTGCCCCTCCTTCTGGCCACCATG
CGAAGCCACACTGACGTGCCTCTCCCTCCCTCCAG*GAAGCCTACGTGATGGCCAGCGTGGACAACCCCCACGTGTGCCGCCTGCTGG*
GCATCTGCCTCACCTCCACCGTGCAGCTCATCACGCAGCTCATGCCCTTCGGCTGCCTCCTGGACTATGTCCGGGAACACAAAGACA
*ATATTGGCTCCCAGTACCTGCTCAACTGGTGTGTGCAGATCGCAAAG*GTAATCAGGGAAGGGAGATACGGGGAGGGGAGATAAGGAG
CCAGGATCCTCACATGCGGTCTGCGCTCCTGGGATAGCAAGAGTTTGCCATGGGGATATG

E20-1-F:	CACACTGACGTGCCTCTCC
E20-1-R:	CTTCGCATGGTGGCCAGA
E20-2-1F:	CCTCCCCGTATCTCCCT
E20-2-1R:	GGAGATAAGGAGCCAGGAT

EX0N21 (156bp)

CTAACGTTCGCCAGCCATAAGTCCTCGACGTCGAGAGGCTCAGAGCCTGGCATGAACATGACCCTGAATTCGGATGCAGAGCTTCTT
CCCATGATGATCTGTCCCTCACAGCAGGGTCTTCTCTGTTTCAG*GGCATGAACTACTTGGAGGACCGTCGCTTGGTGCACCGCGACC*
TGGCAGCCAGGAACGTACTGGTGAAAACACCGCAGCATGTCAAGATCACAGATTTTGGGCTGGCCAAACTGCTGGGTGCGGAAGAGA
*AAGAATACCATGCAGAAGGAGGCAAAG*TAAGGAGGTGGCTTTAGGTCAGCCAGCATTTTCCTGACACCAGGGACCAGGCTGCCTTCC
CACTAGCTGTATTGTTTAACACATGCAGGGGAGGATGCTCTCCAGACATTCTGGGTGAGCTCGCAGC

E21-1-F:	AGCAGGGTCTTCTCTGTTTCA
E21-1-R:	GAGGGACAGATCATCATGGG
E21-2-F:	TTTCCTGACACCAGGGACC
E21-2-R:	TGACCTAAAGCCACCTCCTT

2. Plasma free DNAs were extracted from 2mL plasma.

3. End-filling:

The reaction mixture was prepared as follows:

Table 1

PlasmaDNA solution	38.5 µl
T4 DNA phosphorylation buffer (10X)	5 µl
10 mM dNTP mixture	2 µl
T4 DNApolymerase	2 µl
T4 DNA phosphorylase	2 µl
Klenow enzyme	0.5µl
SterileH_2O	0 µl
Total volume	50 µl

The mixture was placed in a 20°C warm bath for 30min. The DNA sample was purified on a purification column and eluted by 42µl sterile dH_2O or an elution buffer.

4. Adding poly-adenine tail on the 3' end of the DNA fragments:

The reaction mixture was prepared as follows:

Table 2

End-filled DNA	32µl
Klenow reaction buffer (10X)	5µl
dATP solution	10 µl
klenow ex-enzyme (lacking 3'-5' exonuclease activity)	3µl
SterileH_2O	0µl
Total volume	50µl

The mixture was placed in a 37°C warm bath for 30min. The DNA sample was purified on a column and eluted by 25µl sterile dH_2O or an elution buffer.

5. Connecting linkers to the DNA fragments

The reaction mixture was prepared as follows:

Table 3

End-filled dA-tailed DNA	33µl
Reaction buffer for quick connection (5X)	10µl
5 µm DNA linker	2 µl
Quick T4 DNA ligase (NEB)	5µl
Total volume	50µl

The mixture was placed in a 20°C warm bath for 15min. The DNA sample was purified on a Qiagen column and eluted by 25µl sterile dH_2O or an elution buffer.

6. Enriching the linker-modified DNA fragments by PCR pre-amplification

The PCR reaction mixture was prepared as follows:

Buffers EB	14 uL
10X Taq ligase buffer	5 uL
Split Oligo (10 µ M)	4 uL

Pre-lib	25 uL
Taq ligase	2 uL
Total volume	50 uL

PCR programs:

95°C	30s	30cycles
50°C	2min	
4°C	stop	

Table 4

DNA	12.5 μl
Phusion DNA polymerase (Phusion DNA polymerase mixture)	25 μl
PCR primer mixture	2 μl
Ultrapure water	10.5μl
Total volume	50μl

Amplification using the following PCR programs:

1. a. 98°C 30s;
2. b. 18 cycles as follows:
   98°C 10s, 65°C 30s, 72°C 30s;
3. c. 72°C 5min;
4. d. maintained at 4°Co

6. PCR product was analyzed by electrophoresis on 2% agarose gel, and the results were shown in Figure 4. The gel with target band of 200bp was cut and recycled using a Qiagen Kit (Figure 4), and eluted by 20μl elution buffer.

7. Cyclization

The cyclization system was prepared as follows(Table 5)

DNA template	12ul
Circ Ligase II 10X reaction buffer	2 ul
50 mM MnCl2	1 ul
5 M Betaine (optional):	4 ul
Circ Ligase II ssDNA ligase (100 U)	1 ul
Total volume	20 ul

Reaction conditions

60°C	1h
80°C	10min
4°C	stop

8. Enzyme digestion:

All the cyclization products were digested by Exo III, and the digestion system was

10 × NE buffer	1.2ul
ExoIII	1ul
Cyclization product	10ul

The digestion system was placed in a PCR machine, and reacted for 30 min at 37°C.

The digested product was purified on a purification column and dissolved in 30ul EB buffer. The concentration was measured by Qubit, and the results were as follows:

Sample Number	Concentration
2	4ng/ul
3	0.3 ng/ul
4	2.33 ng/ul
5	5.74ng/ul

Examples of the present invention used a splint-mediated cyclization, which has a high rate of cyclization. Detailed tracking and detecting results of every step during the cyclization process were shown in Figure 3. Lane 3 shows the cyclization product after digestion by exonuclease, which is the target product. The cyclization rate was about 20%.

9. Reverse PCR screening of the target zone using back-to-back primers

PCR reaction system was prepared as follows (Table 6)

dd H_2O	13 or 18 ul
AmpliTaq Gold® 360 Master Mix (2 ×)	25 ul
CycEGFR18-F1	1ul
CycEGFR18-R1	1ul
DNA	10 or 5 ul
Total volume	50 ul
Note: the controls were P (with primers and without template) and N (template is un-cyclized ssCyc Lib)	

PCR reaction conditions (Table 7)

95°C	10min	1 cycle
95°C	30s	30 cycles
55°C	30s	
72°C	30s	
72°C	5min	1 cycle

10. The second round PCR

The second round PCR was performed using products of the reverse PCR as templates. System (Table 8)

Phusion PCR Master Mix (2 ×)	25 uL
P5-B1-F (10mM)	1 uL
Primer 2 -index 1 -2 (10uM)	1 uL

Products of reverse PCR	5uL
ddH_2O	add to 50 uL

Programs:

98°C	30s	1 cycle
98°C	30s	12 cycles
65°C	30s	
72°C	30s	
72°C	5min	1ycle

10 uL PCR product from the second round PCR was analyzed by electrophoresis on 2% agarose gel, and the results were shown in Figure 4.

11. The remaining 40 uL PCR product from the second round PCR was purified on a QIAGEN column, and dissolved in 20 uL EB buffer to generate the final library.

12. After quality control, the generated library was 250bp double-end sequenced by IlluminaMiseq.

13. Every high throughput sequenced double-end sequences were assembled to one sequence based on repeated regions. Linkers were removed, and the sequence was restored to the original template sequence, which was then compared to the human genome (hg19). The unique template sequence set was counted by comparing the start and terminal coordinates and tagging sequences of the template sequence on the genome. Using the unique template sequence, genome coverage was then calculated, which can be used for evaluating the specificity of the library and calculating the somatic cell mutations in the EGFR region.

[0053] The results were compared with an EGFR gene mutation detection kit from AmoyDx(directed to the same cancer tissue) and a digital PCR. The comparison results were:

Sample Number	ARMS Results (AmoyDx kit)	ddPCR Results (positive oil droplets/total oil droplets)	Sequencing Results (positive templates/total templates)
LC113	19-del	19-del (2/723)	19-del c.2239_2251>C (7/1192)
LC314	WT	WT (del:0/34; 1858r:0/178; t790:0/208)	WT (0/876)
LC320	19-del	19-del (1/161)	19-del c.2236_2250del15 (9/909)
LC2	WT	n/a	WT (0/668)
LC3	L858R	L858R (589/858)	c.2573T>G; p.L858R (1599/2510)

[0054] 5 samples tested by the method of the present invention showed highly consistent results with those tested by other methods. AMRS-PCR (amplification refractory mutation

system) Taq DNA polymerase lacks 3'-5' exonuclease activity. Under certain conditions, effective amplification occurs only when the last base on 3' end of PCR primers is complementary to the template DNA. Mutated genes and wild type genes can be directly distinguished by PCR using suitable primers directed to different known mutations. This method is mainly for biopsy and FFPE samples. 5 mL peripheral blood before surgery and FFPE samples after surgery from the same patient were provided. The mutation type of FFPE samples on the known sites was detected by ARMS-PCR, and then plasma DNA was detected by digital PCR and the method of the present invention, respectively.

Example 2

[0055] Reliability of the present method was verified by detecting known cancer mutation sites. The method was the same as that disclosed in Example 1.

[0056] Cancer cell line DNA: cell DNAs containing hybrid c.2235_2249del15 (exon 19) mutation were broken by ultrasonication, 166 ± 10bp fragments were recycled and mixed with plasma DNAs from a normal person at certain ratio. The sensitivity and stability of the method were examined.

[0057] The detection results of c.2235_2249del15 were as follows:

Sample Nos.	Prediction	Nos. of Total templates	Nos. of Non-Del templates	Nos. of Del templates	ratio
Sample 1	0% mutation	862	862	0	0.00%
Sample 2	0.1 mutation	438	433	5	1.09%
Sample 3	1% mutation	905	843	62	6.78%
Sample 4	5% mutation	880	786	94	10.67%
Sample 5	25% mutation	1631	1017	614	37.65%

SEQUENCE LISTING

[0058]

<110> BERRY GENOMICS CO., LTD

<120> A method and a kit for non-invasively detecting EGFR gene mutations

<130> CIE140060PCT

<160> 20

<170> PatentIn version 3.2

<210> 1
<211> 18
<212> DNA
<213> Artificial

<400> 1
cccagcttgt ggagcctc 18

<210> 2
<211> 23
<212> DNA
<213> Artificial

<400> 2
gacaagaaca cagagacaag ggt 23

<210> 3
<211> 18
<212> DNA
<213> Artificial

<400> 3
gcagggcctc tcatggtc 18

<210> 4
<211> 19
<212> DNA
<213> Artificial

<400> 4
cctgtgccag ggaccttac 19

<210> 5
<211> 24
<212> DNA
<213> Artificial

<400> 5
acgtcttcct tctctctctg teat 24

<210> 6
<211> 20

<212> DNA
<213> Artificial

<400> 6
gtgagatggt gccacatgct 20

<210> 7
<211> 20
<212> DNA
<213> Artificial

<400> 7
gtccatggct ctgaacctca 20

<210> 8
<211> 20
<212> DNA
<213> Artificial

<400> 8
ccacacagca aagcagaaac 20

<210> 9
<211> 19
<212> DNA
<213> Artificial

<400> 9
cacactgacg tgcctctcc 19

<210> 10
<211> 18
<212> DNA
<213> Artificial

<400> 10
cttcgcatgg tggccaga 18

<210> 11
<211> 17
<212> DNA
<213> Artificial

<400> 11
cctccccgta tctccct 17

<210> 12
<211> 19

<212> DNA
<213> Artificial

<400> 12
ggagataagg agccaggat 19

<210> 13
<211> 21
<212> DNA
<213> Artificial

<400> 13
agcagggtct tctctgtttc a 21

<210> 14
<211> 20
<212> DNA
<213> Artificial

<400> 14
gagggacaga tcatcatggg 20

<210> 15
<211> 19
<212> DNA
<213> Artificial

<400> 15
tttcctgaca ccagggacc 19

<210> 16
<211> 20
<212> DNA
<213> Artificial

<400> 16
tgacctaaag ccacctcctt 20

<210> 17
<211> 381
<212> DNA
<213> Human being

<400> 17
caagtgccgt gtcctggcac ccaagcccat gccgtggctg ctggtccccc tgctgggcca 60

tgtctggcac tgctttccag catggtgagg gctgaggtga cccttgtctc tgtgttcttg 120

tccccccccag cttgtggagc ctcttacacc cagtggagaa gctcccaacc aagctctctt 180

gaggatcttg aaggaaactg aattcaaaaa gatcaaagtg ctgggctccg gtgcgttcgg 240

cacggtgtat aaggtaaggt ccctggcaca ggcctctggg ctgggccgca gggcctctca 300

tggtctggtg gggagcccag agtccttgca agctgtatat ttccatcatc tactttactc 360

tttgtttcac tgagtgtttg g 381

<210> 18
<211> 372
<212> DNA
<213> Human being

<400> 18
gcaatatcag ccttaggtgc ggctccacag ccccagtgtc cctcaccttc ggggtgcatc 60

gctggtaaca tccacccaga tcactgggca gcatgtggca ccatctcaca attgccagtt 120

aacgtcttcc ttctctctct gtcataggga ctctggatcc cagaaggtga gaaagttaaa 180

attcccgtcg ctatcaagga attaagagaa gcaacatctc cgaaagccaa caaggaaatc 240

ctcgatgtga gtttctgctt tgctgtgtgg gggtccatgg ctctgaacct caggcccacc 300

ttttctcatg tctggcagct gctctgctct agaccctgct catctccaca tcctaaatgt 360

tcactttcta tg 372

<210> 19
<211> 408
<212> DNA
<213> Human being

<400> 19
ccatgagtac gtattttgaa actcaagatc gcattcatgc gtcttcacct ggaaggggtc 60

catgtgcccc tccttctggc caccatgcga agccacactg acgtgcctct ccctccctcc 120

aggaagccta cgtgatggcc agcgtggaca acccccacgt gtgccgcctg ctgggcatct 180

gcctcacctc caccgtgcag ctcatcacgc agctcatgcc cttcggctgc ctcctggact 240

atgtccggga acacaaagac aatattggct cccagtacct gctcaactgg tgtgtgcaga 300

tcgcaaaggt aatcagggaa gggagatacg gggaggggag ataaggagcc aggatcctca 360

catgcggtct gcgctcctgg gatagcaaga gtttgccatg gggatatg 408

<210> 20
<211> 415
<212> DNA
<213> Human being

<400> 20
ctaacgttcg ccagccataa gtcctcgacg tggagaggct cagagcctgg catgaacatg 60

accctgaatt cggatgcaga gcttcttccc atgatgatct gtccctcaca gcagggtctt 120

ctctgtttca gggcatgaac tacttggagg accgtcgctt ggtgcaccgc gacctggcag 180

ccaggaacgt actggtgaaa acaccgcagc atgtcaagat cacagatttt gggctggcca 240

```
aactgctggg tgcggaagag aaagaatacc atgcagaagg aggcaaagta aggaggtggc    300
tttaggtcag ccagcatttt cctgacacca gggaccaggc tgccttccca ctagctgtat    360
tgtttaacac atgcagggga ggatgctctc cagacattct gggtgagctc gcagc         415
```

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO2013138510A [0002] [0002]

Patentkrav

1. Fremgangsmåde til ikke-invasiv detektering af EGFR-genmutationer hos et individ, omfattende de følgende trin:

at udforme primere ifølge EGFR-gen exoner;
at ekstrahere plasma DNA'er i en plasmaprøve opnået fra individet;
at forbinde de ekstraherede plasma DNA'er med mærkningslinkere;
at PCR-præ-amplificere de mærkningslinkede forbundne plasma DNA'er;
at cyklisere de præ-amplificerede DNA'er for at opnå cykliserede DNA'er;
at PCR-amplificere de cykliserede DNA'er under anvendelse af de designede primere; og
high-throughput sekvensering af det PCR-amplificerede produkt og analysering af EGFR-genmutationerne.

2. Fremgangsmåde ifølge krav 1, **kendetegnet ved, at** de designede primere er et par af primere, der er tilstødende og bagud forlængede.

3. Fremgangsmåde ifølge krav 2, **kendetegnet ved, at** primere fra det bagudrettede forlængede primerpar er placeret på 5'- eller 3'-enden af EGFR-gen exonerne.

4. Fremgangsmåde ifølge krav 2, **kendetegnet ved, at** område af det bagudrettede forlængede primerpar er 0-1/2 af totalen af basepar af DNA'erne.

5. Fremgangsmåde ifølge krav 3, **kendetegnet ved, at** de bagudrettede forlængede primere er rettet mod exon 18, exon 19, exon 20 eller exon 21 af EGFR-generne.

6. Fremgangsmåde ifølge krav 3, **kendetegnet ved, at** 5'-enden af de bagudrettede forlængede primere indeholder linkersekvenser for high-throughput sekvenseringsbibliotek.

7. Fremgangsmåde ifølge krav 3, **kendetegnet ved, at** de bagudrettede forlængede primere er rettet mod exon 18, exon 19, exon 20 eller exon 21 af EGFR-generne, og sekvenserne af primerne er som følger:

E18-1-3F: CCCAGCTTGTGGAGCCTC
E18-1-3R: GACAAGAACACAGAGACAAGGGT
E18-2-F: GCAGGGCCTCTCATGGTC
E18-2-R: CCTGTGCCAGGGACCTTAC
E19-1-F: ACGTCTTCCTTCTCTCTCTGTCAT
E19-1-R: GTGAGATGGTGCCACATGCT
E19-2-F: GTCCATGGCTCTGAACCTCA
E19-2-R: CCACACAGCAAAGCAGAAAC
E20-1-F: CACACTGACGTGCCTCTCC
E20-1-R: CTTCGCATGGTGGCCAGA
E20-2-1F: CCTCCCCGTATCTCCCT
E20-2-1R: GGAGATAAGGAGCCAGGAT
E21-1-F: AGCAGGGTCTTCTCTGTTTCA
E21-1-R: GAGGGACAGATCATCATGGG
E21-2-F: TTTCCTGACACCAGGGACC
E21-2-R: TGACCTAAAGCCACCTCCTT.

8. Fremgangsmåde ifølge krav 1, **kendetegnet ved, at** cykliseringen er en splint-medieret enkeltstrenget DNA-cyklisering.

9. Fremgangsmåde ifølge krav 1, **kendetegnet ved, at** EGFR-generne i plasma DNA'er har indsætning, sletning, substitution eller genfusionsmutationer.

10. Fremgangsmåde ifølge krav 1, **kendetegnet ved, at** high-throughput sekvenseringsteknologierne er valgt fra Roche/454 FLX, Illumina /Hiseq/Miseq, Applied Biosystems SOLID og life Technologies/Ion Torrent/Proton.

11. Kit til ikke-invasiv detektering af EGFR-genmutationer, omfattende:
primere til præ-amplificering af regioner der skal testes i EGFR-generne,
en DNA-cyklase, og
primere til amplificering af mål-DNA'er, hvor primerne til amplificering af mål-DNA'er er primere designet ifølge EGFR-gen exoner.

12. Kit ifølge krav 11, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** primerene til amplificering af mål-DNA'er er et par af

primere, der er tilstødende og bagud forlængede.

13. Kit ifølge krav 11, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** primere af det bagudrettede forlængede primerpar er placeret på 5'- eller 3'-enden af stederne eller regionerne der skal testes i EGFR-generne.

14. Kit ifølge krav 11, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** område af det bagudrettede forlængede primerpar er 0-1/2 af totalen af basepar af plasma DNA'erne.

15. Kit ifølge krav 11, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** de bagudrettede forlængede primere er rettet mod exon 18, exon 19, exon 20 eller exon 21 af EGFR-generne.

16. Kit ifølge krav 14, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** de bagudrettede forlængede primere er rettet mod exon 18, exon 19, exon 20 eller exon 21 af EGFR-generne, og sekvenser af primerne er som følger:

E18-1-3F: CCCAGCTTGTGGAGCCTC
E18-1-3R: GACAAGAACACAGAGACAAGGGT
E18-2-F: GCAGGGCCTCTCATGGTC
E18-2-R: CCTGTGCCAGGGACCTTAC
E19-1-F: ACGTCTTCCTTCTCTCTCTGTCAT
E19-1-R: GTGAGATGGTGCCACATGCT
E19-2-F: GTCCATGGCTCTGAACCTCA
E19-2-R: CCACACAGCAAAGCAGAAAC
E20-1-F: CACACTGACGTGCCTCTCC
E20-1-R: CTTCGCATGGTGGCCAGA
E20-2-1F: CCTCCCCGTATCTCCCT
E20-2-1R: GGAGATAAGGAGCCAGGAT
E21-1-F: AGCAGGGTCTTCTCTGTTTCA
E21-1-R: GAGGGACAGATCATCATGGG
E21-2-F: TTTCCTGACACCAGGGACC

E21-2-R: TGACCTAAAGCCACCTCCTT.

17. Kit ifølge krav 11, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** 5'-enden af de bagudrettede forlængede primere indeholder linkersekvenser for high-throughput sekvenseringsbibliotek.

18. Kit ifølge krav 11, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** linkerne indeholder mærkningssekvenser.

19. Kit ifølge krav 11, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** plasma DNA'erne er præ-amplificerede, inden de cykliseres.

20. Kit ifølge krav 11, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** cykliseringen er en splint-medieret enkeltstrenget DNA-cyklisering.

21. Kit ifølge krav 11, til ikke-invasiv detektering af EGFR-genmutationer, **kendetegnet ved, at** steder eller regioner der skal testes i EGFR-generne har indsætning, sletning, substitution eller genfusionsmutationer.

DRAWINGS

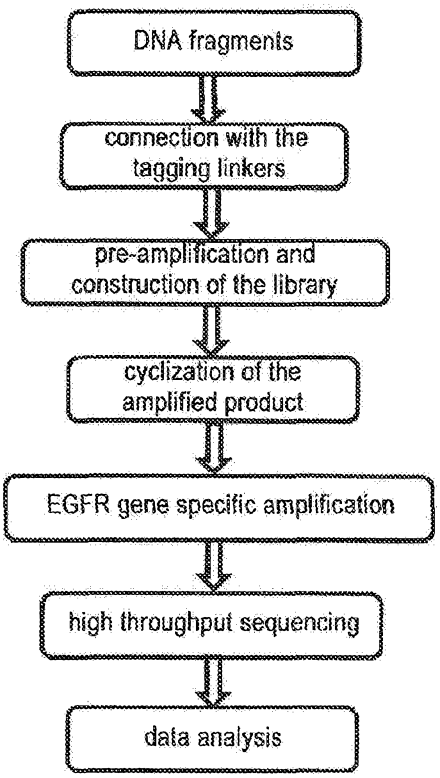


Fig. 1

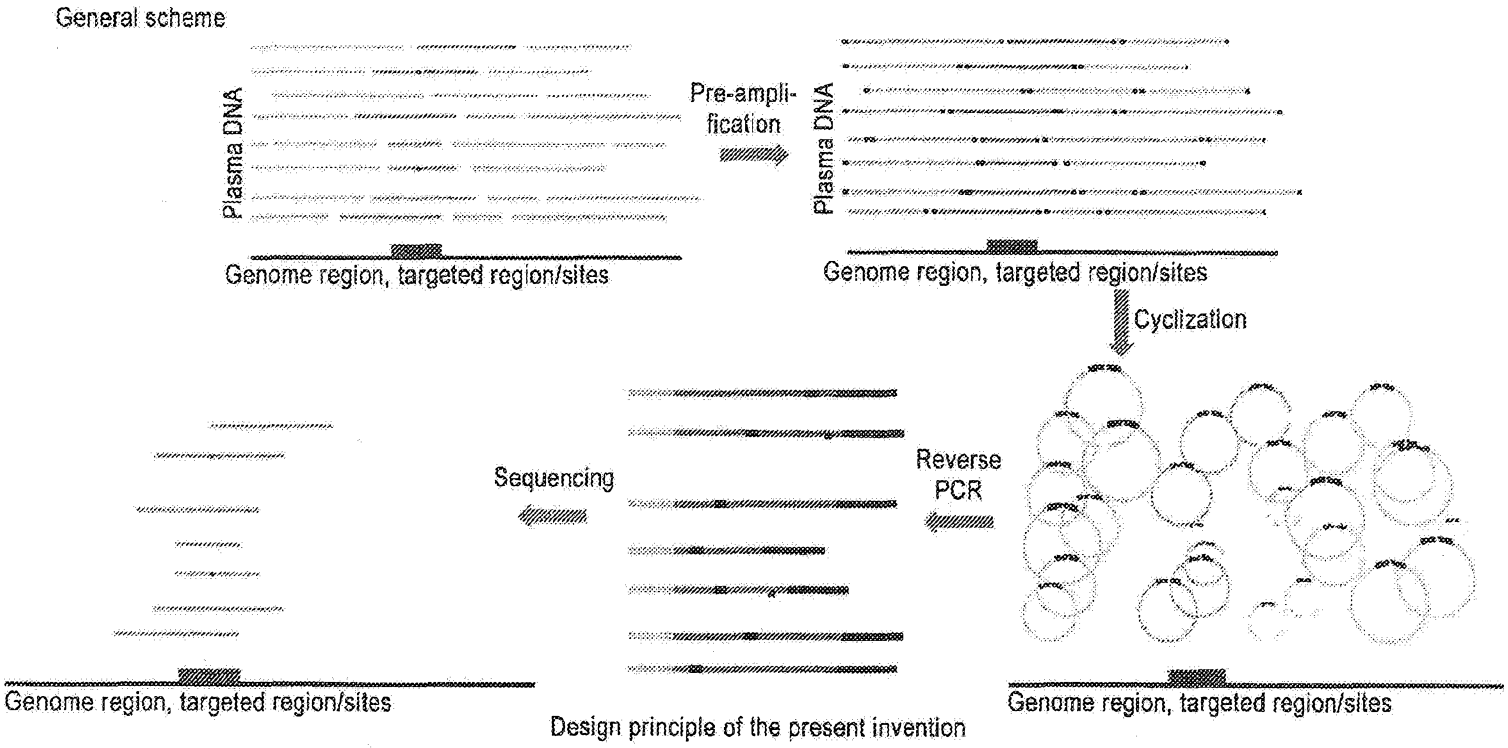


Fig. 2A

DK/EP 3090068 T3

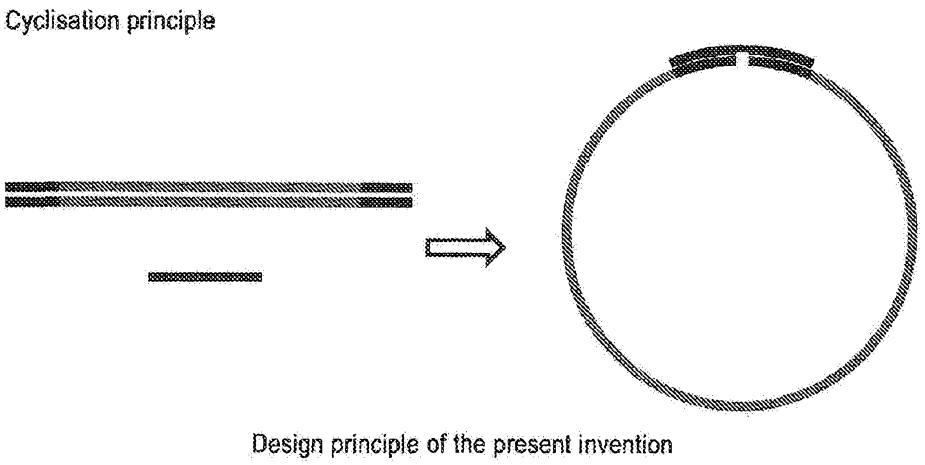
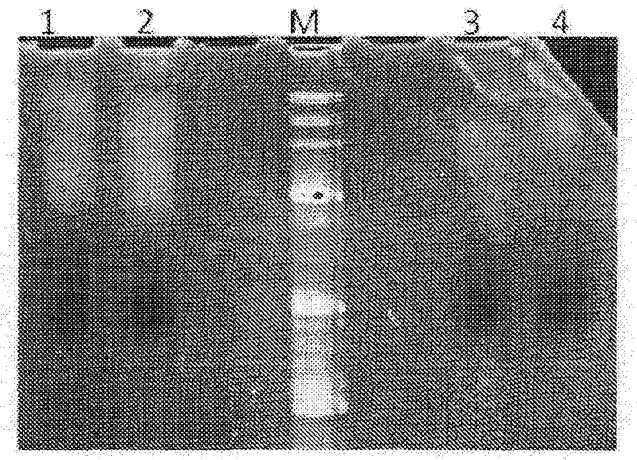

Cyclisation principle
Design principle of the present invention

Fig. 2B



1: PCR product before cyclization

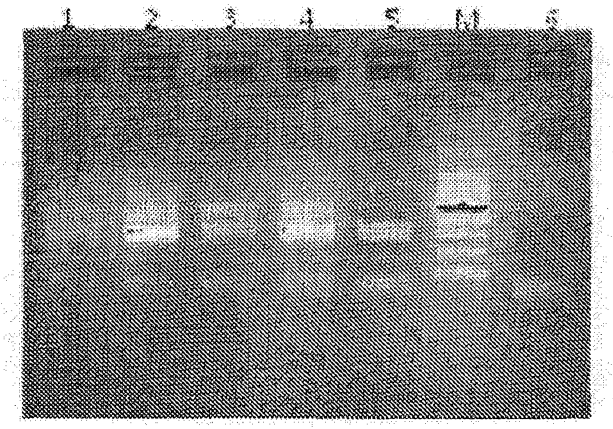
2: Product after cyclization (not digested by an exonuclease) M: marker

3: Exonuclease digested product after cyclization (not purified)

4: Exonuclease digested and purified product after cyclization

Cyclization rate of the present invention (6% denaturing gel)

Fig. 3



Final library of EGFR sample (2% agarose gel)

Fig. 4

1: LC2; 2: LC3; 3: LC113; 4: LC314; 5: LC320; 6: negative control;

M: 100bp DNA ladder (TaKaRa).