(12) UK Patent Application (19) GB (11) 2 382 137 (13) A

(43) Date of A Publication 21.05.2003

(21) Application No 0127704.5

(22) Date of Filing 20.11.2001

(71) Applicant(s)
Mats Gullberg
Glimmervagen 11b, SE-752 41 Uppsala, Sweden

Ulf Landegren
Norra Parkvagen 29, SE-756 45 Uppsala, Sweden

(72) Inventor(s)
Mats Gullberg
Ulf Landegren

(74) Agent and/or Address for Service
Haseltine Lake & Co
Imperial House, 15-19 Kingsway, LONDON, WC2B 6UD, United Kingdom

(51) INT CL^7
C12Q 1/68

(52) UK CL (Edition V)
G1B BAC B200 B201 B222 B223 B224

(56) Documents Cited
WO 2002/057486 A **WO 1999/049079 A**
US 6221603 B
Nucl Acids Res; Vol 29, pp e116 (2001). Qi et al.
Nucl Acids Res; Vol 28, pp e58 (2000). Antson et al.

(58) Field of Search
INT CL^7 **C12Q** 1/68
Other: **ONLINE: WPI, EPODOC, JAPIO, CAPLUS, MEDLINE, EMBASE, BIOSIS, SCISEARCH**

(54) Abstract Title
Nucleic acid enrichment

(57) A method of enriching a preselected target nucleic acid segment which comprises a specific variant from a mixture of sequences comprising: cleaving the nucleic acid sequences within the mixture to provide a fragment comprising the pre-selective nucleic acid segment; providing a template oligonucleotide, one end of which hybridises to a sequence of the segment at or close to the variant position, the other end of which hybridises to the end of a protecting sequence; hybridising the template to the segment such that the variant position and the protecting sequence are brought into close proximity of each other; ligating the end of the protecting sequence to the end of the variant sequence; and enriching for ligated product. The enrichment step preferably consists of rolling circle amplification.

GB 2 382 137 A

Figure 1

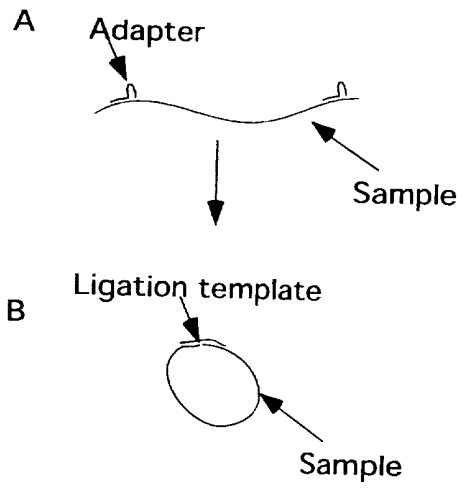


Figure 2

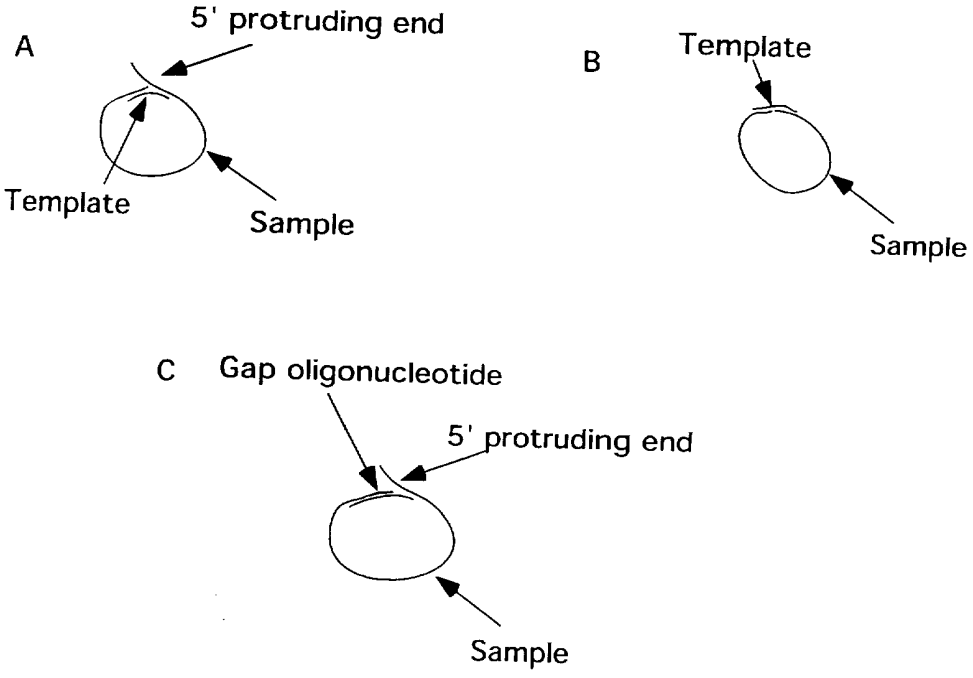


Figure 3

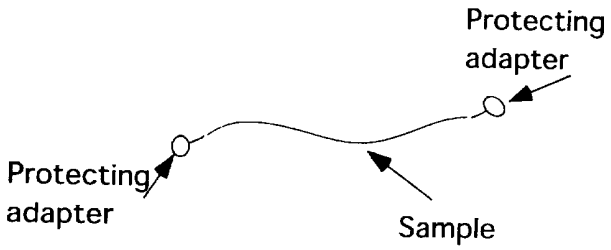


Figure 4

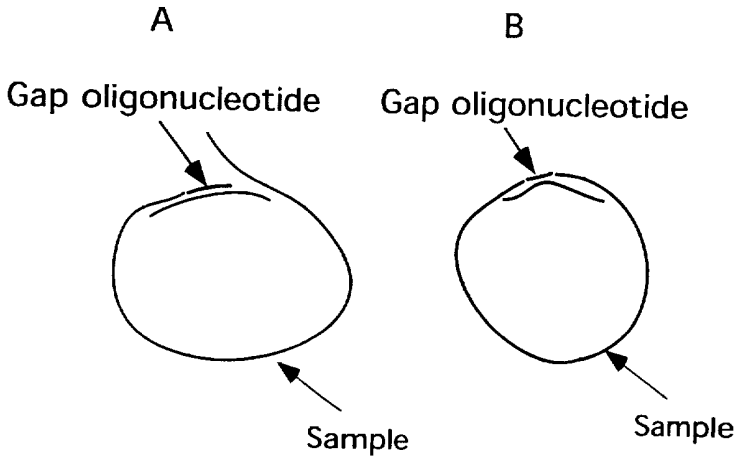


Figure 5

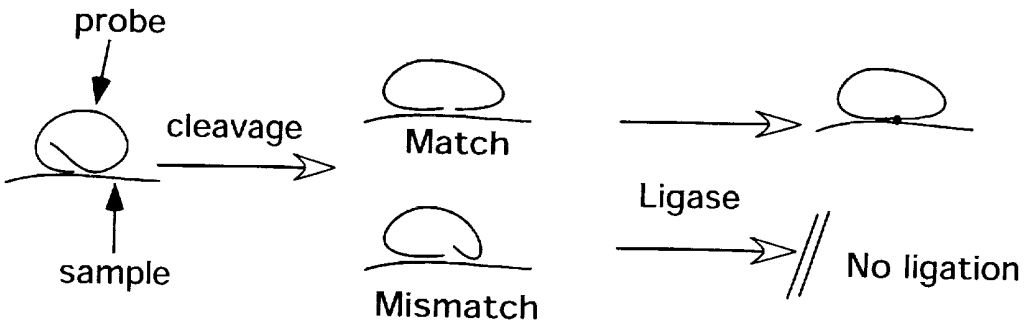
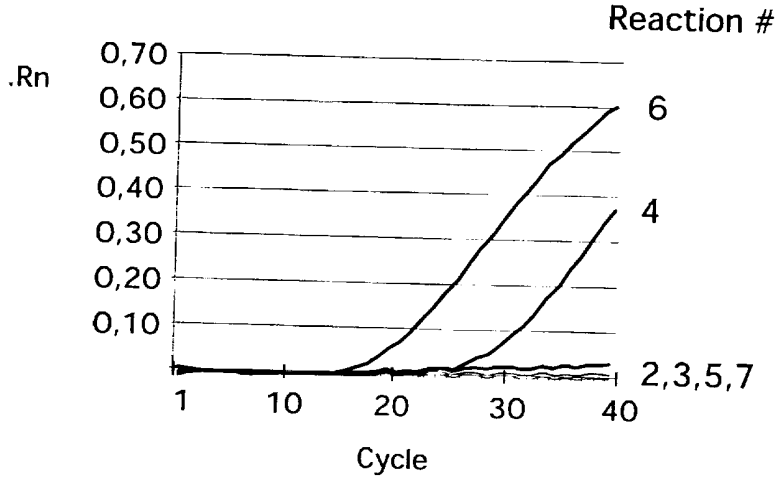


Figure 6

A


Reaction #
.Rn
0,70
0,60
0,50
0,40
0,30
0,20
0,10
6
4
2,3,5,7
1
10
20
30
40
Cycle

B

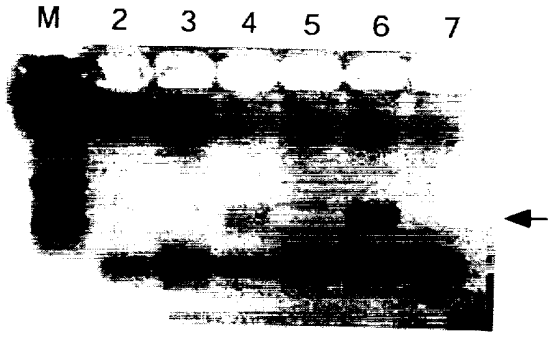
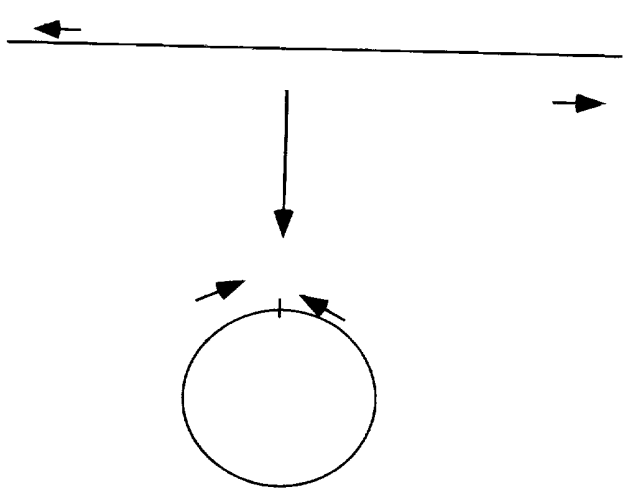

M
2
3
4
5
6
7

Figure 7

A



B

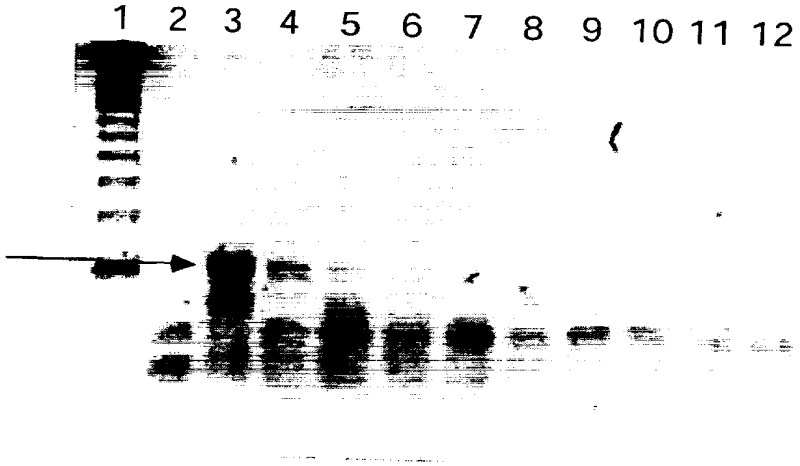
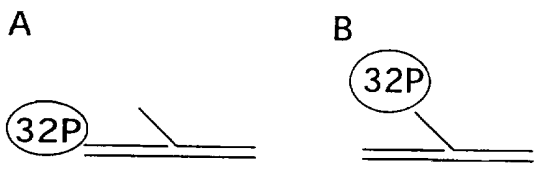


Figure 8



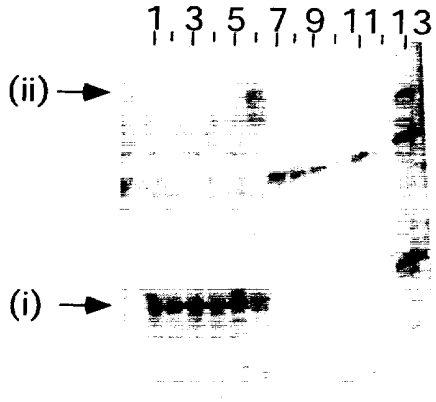
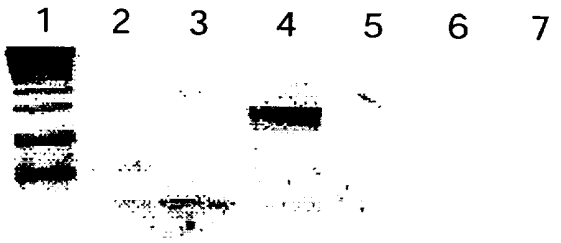


Figure 9



NUCLEIC ACID ENRICHMENT

Field of the Invention

This invention relates to methods of enriching nucleic acid sequences. More particularly, the present invention relates to sample preparation methods including sample modification, sample enrichment and amplification.

Background

Haplotype information can be vital in the analysis of disease by determining whether two or more sequence variants are located on the same nucleic acid fragment. This is of special interest in tumour research and diagnosis where it is important to know if two or more inactivating mutations occur on the same or different chromosomes. Similarly better information about which genotypes are located on the same nucleic acid segment can greatly increase the information derived from genotyping data and in the statistical analysis of genetic linkage or linkage disequilibrium of inherited traits and markers such as single nucleotide polymorphisms, SNP:s (e.g. [1], [2], [3]).

To date there has been no method developed that satisfactorily solves the problem of how to obtain haplotype information *in vitro*, that is to determine which gene variants are located, over some distance, on the same nucleic acid molecule.

Current analysis of heritable diseases is hampered by the fact that given the genotypes of the parents, it is still often impossible to confirm if a particular allele is obtained from the mother or from the father. In theory, with the ability to distinguish between haplotypes, more meioses would be informative and thereby facilitate genetic linkage analysis. Another area where haplotyping has proven to be of interest is in the study of genetic effects on a subject's response to different drugs. Recent publications have shown that haplotype information is important to be able to relate genetic factors to a patient's response to various drugs, [4].

Currently, there are only a few methods available for obtaining haplotype information. When lineage data and nucleic acid samples are available linkage analysis is applicable. It is also possible to use statistical methods to calculate possible allele-combinations from allele frequencies to gain information on a haplotype. However, this technique can only be used with a small number of alleles at the same time and on population-size data, and the analysis only provides statistical evidence for the presence of a given haplotype. Haplotype information can also be gained from hemizygous X- and Y-chromosomes, where haplotypes are immediately apparent from the genotype.

The possibility to study cells with only one autosome chromosome is utilised in some *in vivo* techniques. One approach is the creation of rodent-human hybrid cells, for example using the so called "Conversion technology" [2]. Some of the rodent-human hybrid cells will contain one of the two possible copies from a human chromosome. A second approach is to use hydatidiform moles, i.e. tissues that due to a fertilisation defect only contain genetic material from the sperm (complete hydatidiform mole) thereby containing only one copy of each chromosome.

There are also some *in vitro* molecular techniques that can be used to determine haplotypes. One technique is the sub-cloning of all nucleic acid sequences of interest, isolating individual clones and subsequently genotyping them. Allele-specific analysis through Fibre Fluorescent In Situ Hybridisation is another possible approach, however it has not yet been convincingly shown to be useful for SNP based haplotyping. A third approach is double PCR Allele Specific Amplification (double-PASA [5], a double allele-specific Polymerase Chain Reaction (PCR) which gives linkage information of two adjacent polymorphic sites. Pyrosequencing [6] and mass spectrometry may be used to analyse haplotypes over short distances, i.e. <100nt.

Summary of the Invention

Methods to analyse haplotypes, genotypes and enrichment of selected sequences are described herein. These methods are also of interest for population genetics, identification of lineage in plant and animal breeding and in analysis of micro-organisms.

In one aspect of the invention, a general technique is provided to obtain haplotypes through enrichment for one nucleic acid segment, selected to include a specific variant at a given position. Thereby any variant position in a sample could be used for selection, followed by analysing genetic variants elsewhere in the same nucleic acid fragment.

In another aspect of the invention the same principle can be used for genotyping or to generate probes that reveal the genotype at particular loci.

Accordingly, the present invention provides a method for sample preparation that optionally includes the steps of: (a) cleavage of a nucleic acid so that a fragment containing the sequence to be investigated is created with or without addition of oligonucleotide probes (b) selective modification of one variant of the nucleic acid sequences (c) enrichment of the selected variant, and (d) analysis of the nucleic acid.

The present invention also provides one or a set of probes for use in the described methods. A first set of probes/probe preferably directs site specific cleavage at predetermined sites of the sample upon hybridisation. Preferably, A second set of probes/probe is used to specifically modify the sample based upon the presence or absence of a given sequence variant. A third set of probes is used for amplification of the sample and a fourth set of probes is used for scoring the genotypes.

The present invention describes several ways to enrich a nucleic acid sequence or sequences from a multitude of sequences on the basis of the sequence or on the basis of a particular sequence variation at a given position.

Detailed Description of the Invention

The terms "nucleic acid", "nucleic acid sequence", "nucleic acid fragment", "nucleic acid segment ", "nucleic acid probe", "target nucleic acid sequence" or "target sequence" describe interchangeably and without preference, a plurality of nucleotides, covalently linked as such to form a linear molecules of DNA or RNA.

The term "variant" describes interchangeably and without preference a nucleic acid encoding a variant, which may for example be selected from the group including any one or more of the following; a single nucleotide sequence variant deletion sequence variant, insertion sequence variant, sequence length variants, and sequence variation among paralogous or orthologous nucleic acid sequence, or among edited sequences or splice variants

Examples of different approaches are as follows.

The first approach, described in part in Figure 1, is based on cleavage of DNA at any predetermined site through the use of so called nucleic acid adapters, hereafter called adapters, that are targets or part of targets for restriction enzymes preferably type II restriction enzymes [7],[8]. Adapters and sample are mixed, denatured and subsequently allowed to cool. The adapters hybridise to their complementary regions in the sample nucleic acid. One of the adapters is positioned so that the resulting cleaved sample DNA contains a variant position at the 5' position (A). Added restriction enzymes cleave the sample and, through addition of a ligation template that anneals to both the 5' and 3' end of the cleaved sample DNA, circular molecules are obtained by ligation of the ends that are brought next to each other (B). This circularisation is driven by the higher relative concentration of two ends belonging to the same molecule compared to those of two different copies of the same or similar molecules. If the added template is complementary to the sample DNA-ends, juxtaposing these, then ligation of the two ends can occur. If a mismatch between the sample DNA and the ligation template exists, at the variant position used for selection, or if there are no free ends at the site intended for ligation, then ligation will not occur. Circularised molecules can

then be enriched for through the use of exonucleases that degrade uncircularised DNA, and/or amplification of the circularised DNA, for example with rolling circle amplification (RCA) can be performed ([9][10]).

Alternatively, the adapter could be positioned upstream of the variant position used for selection. Optionally this adapter could be completely omitted. After cleavage, as described earlier, one or a plurality of oligonucleotides is added, (template), which hybridises to both the 3' end and to an upstream sequence around the variant position, as shown in Figure 2A. This provides a specificity step. The structure is then cleaved by chemical, enzyme or other means to generate a structure, as shown in Figure 2B. Where an enzyme is used, any enzyme capable of cleaving such a structure [11] may be used. The enzyme is preferably selected from, *FEN* nuclease, *Mja* nuclease, native or recombinant polymerase from *Thermus aquatiqus*, *Thermus thermophilus*, or *Thermus flavus*, or any enzyme selected according to the teachings of Lyamichev *et al* [11] or US 5846717, which are incorporated herein by reference. The variant position used for selection can either be removed by cleavage, or the cleavage can be performed so that the variant position is the 5'-most nucleotide of the sequence. Hence the major selective step is in the subsequent ligation reaction. The use of nucleic acid ligation for allele distinction is well described in the literature, for example [12][13]. To ensure that the cleaved substrate is eligible for ligation the 3' sample nucleotide must be complementary to the added template. This can be achieved directly from cleavage of the sample, in which case it is possible to ligate the DNA directly.

Another approach, which confers increased specificity, is to construct the added template so that it contains one extra nucleotide, giving a gap between the hybridised 3' and 5' sequences, similar to that observed for the SNP. By adding only the complementary nucleotide to the cleavage reaction a substrate for cleavage will only be generated from nucleic acid sequences that contain the complementary nucleic acid sequence.

Yet another approach is to construct the added template so that there will be a gap. This gap may be filled in by the addition of a complementary

oligonucleotide, as shown in Figure 2C. Optionally, this gap filling oligonucleotide can be labelled with an affinity tag, for example a specific sequence or specific molecule for subsequent affinity purification.

Cleavage of the sample DNA can also be achieved with restriction enzymes through the addition of oligonucleotides that hybridise to the selected sequence. The 5' cleavage site may or may not be influenced by the variable sequence.

Instead of circularising the DNA, the nucleic acid fragment ends can be protected via addition of protecting adapters to one or both ends based on selective addition at a variant position at at least one of the ends, as shown in Figure 3. Generation of the 3' or 5' sample ends could be achieved either through cleavage at the variable position or upstream at a generic site, as previously described. In the latter case cleavage will be performed via structure-specific cleavage as previously described. This protected linear substrate can now be enriched for, through degradation of unprotected sample using exonucleases. Selective amplification of the protected allele can be performed based on the presence of the added sequence/sequences.

It is not necessary to generate restriction sites in the sample or to denature double stranded DNA. Any restriction enzyme or pair of restriction enzymes having recognition sequences located on either side but not within the sequence of interest, can be used.

The enriched sample can be subjected to genotyping through any method and compared to results from genotyping of the total sample. Examples of methods which may be used are oligonucleotide ligation assays [13], primer extension assays [14], pyrosequencing [15], invader technology [16], mass-spectroscopy [19] or homogenous PCR methods e.g. Taqman [17] or molecular beacons [18]. However, other methods may be employed with equal utility. By using the enriched sample instead of a whole sample as the test sample it is also feasible to use any suitable method, to find new/unknown mutations or polymorphisms. Thereby all possible mutations in the enriched segment may be detected, also unknown ones, for example by Sanger sequencing or by hybridising the enriched sample to an array in order

to resequence the sample and in this respect also find new or unknown mutations. The methods could be, but are not limited to the use of, mismatch recognising enzymes for example T4 endo VII [20], DHPLC resequencing, Sanger or array, or pyrosequencing [15]. However, other methods may be employed with equal utility. The resulting genotypes will reveal the specific haplotype of the sample.

Accordingly, the present invention provides one or a set of probes. A first set of probes/probe directs site specific cleavage at predetermined sites of the sample upon hybridisation. A second set of probes/probe is used to specifically modify the sample based upon a sequence variant. A third set of probes is used for amplification of the sample and a fourth set of probes is used for scoring the genotypes.

Instead of investigating the genotypes all along the selected nucleic acid one can use the same principle for genotyping the variant position used for selection. Upon cleavage of sample DNA an oligonucleotide can be added that anneals to the 3' end of a generated fragment and to a stretch upstream, around the variant position to be scored, so that a probe with a hybridising region at its 5' end is formed, (as shown in Figure 4A), or a probe with a non-hybridising region at its 5' end is formed, (as shown in Figure 4B). If necessary this structure can then be cleaved as previously described. The use of ligase will complete the nucleic acid circle. The circle can then be enriched for, using exonuclease treatment and nucleic acid amplification, preferably rolling circle amplification. Preferentially the oligonucleotide added contains a sequence between the 3' and the 5' hybridising end that consist of a selected sequence used for later hybridisation that can be rendered double stranded through the addition of a second oligonucleotide, shown in Figure 4A as object 1. The added oligonucleotide could contain a recognition sequence for a type IIs restriction enzyme and preferably a sequence as dissimilar as possible compared to other oligonucleotides used for other loci, as described in co-pending application GB 0118959.6, the contents of which are incorporated herein by reference. Detection of the circularised nucleic acid or amplification products templated by the circularised nucleic acid is used to

score the genotype of the selected position.

Due to the intramolecular nature of the ligation reaction it is feasible to perform many reactions at the same time (from one to several tens of thousands). At any practical concentration the fragments will circularise intramolecularly in preference to intermolecular reactions.

Accordingly, the present invention further provides one or a set of probes. A first set of probes/probe directs site specific cleavage at predetermined sites of the sample upon hybridisation. A second set of probes/probe is used to specifically modify the sample based upon a sequence variant. A third set of probes is used for amplification of the enriched sample.

The variant position could be, but is not limited to a sequence variant polymorphism which may be selected from the group including any one or more; deletion variant, insertion variant, sequence length variant, single nucleotide polymorphism, substitution variant, paralogous or orthologous nucleic acid sequences, edited sequences or splice variants.

The present invention is also to be used as a method to isolate and enrich a specific sequence or sequences from a plurality of sequences, with the option of further manipulation of the enriched sequence/sequences. The method can be a single method or a combination of methods, for example a combination of any of the following: amplification, quantification, sequencing and variant scoring. Enriched sequence/sequences may be used as probes, or to compare different enriched samples.

Accordingly, the present invention further provides one or a set of probes. A first set of probes/probe directs site specific cleavage at predetermined sites of the sample upon hybridisation. A second set of probes/probe is used to specifically modify the sample based upon a nucleotide sequence. A third set of probes is used for amplification of the enriched sample.

In all of the above-mentioned methods where DNA samples are mentioned they could be exchanged with RNA or cDNA samples.

An added oligonucleotide probe can also be treated by the same principles and to be used for subsequent genotyping, as shown in Figure 5, if the added oligonucleotide anneals forming a non-hybridising region at the 5'

end. Cleavage of this structure will generate a molecule that can be circularised with a ligase. Ligation will depend on whether the 5' nucleotide is matched or not with the sample. This circularised probe can then be detected either directly or via the presence of amplification products (based on the presence of the circle or amplification products of the circle). The presence of such a product describes the nature of the variant position. The added oligonucleotide could preferentially contain a molecule or sequence in the 5' part that is used as an affinity tag for removal of unmodified circles before amplification of the circularised probes.

Accordingly, the present invention provides one or a set of probes. A first set of probes to be specifically modified based on the nature of a nucleotide in the target nucleic acid. A second set of probes could be used for purification of the sample. A third set of probes is used for amplification of the modified probes.

Embodiments of the invention will now be described in greater detail, by way of example only, with reference to the accompanying drawings, of which;

Figure 1 is a schematic representation of cleavage and circularisation of sample nucleic acid through the use of adapters;

Figure 2 is a schematic representation of structure specific cleavage for circularisation of sample nucleic acids;

Figure 3 is a schematic representation of addition of protecting ends to a linear nucleic acid sample;

Figure 4 is a schematic representation of the use of gap-oligonucleotides for circularisation of sample nucleic acids;

Figure 5 is a schematic representation of scoring SNPs through circularisation of nucleic acid probes;

Figure 6 shows (A) the result from a real-time PCR experiment and (B) the gel of the same amplification reactions from an experiment of cleaving, ligating and rolling circle amplification of BAC DNA;

Figure 7 is a schematic representation (A) of the experimental set-up for detection of circularisation of nucleic acids via inverse PCR and (B) a photo of an agarose gel showing the result of such an experiment where BAC DNA cut

with *Fok*I adapters, circularised with ligase, circular molecules enriched for via exonucleases and finally used for template in an inverse PCR reaction;

Figure 8 is showing an image of a poly acrylamide gel of radioactive labelled nucleic acids showing cleavage and ligation of structure specific cleaved nucleic acids with native DNA *Taq* polymerase and *Tth* ligase; and

Figure 9 is showing a photo of an ethidium bromide stained gel of amplification products obtained from an experiment with cleaved BAC DNA that had been circularised via cleavage by a structure specific enzyme and the two ends joined by a ligase.

Examples

Example 1

Circularisation of DNA after cleavage with restriction enzymes followed by enrichment through exonuclease treatment and rolling circle amplification. (See Figure 4)

A BAC clone (RP11-381L18, BacPac resources, Children's hospital, Oakland) with a genomic fragment containing the gene *ATP7B* was used. DNA was isolated by the rapid alkaline lysis miniprep method and the DNA concentration was determined measuring UV A_{260}.

*Hpa*II 5 U (New England Biolabs) was used to cleave a double stranded (ds) template in buffer (10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT) for 2 hours at 37°C before heat-inactivation of the enzyme. Two pmol of the ds template was cleaved with *Hpa*II. After the cleavage, the reaction was diluted to different concentrations (10^4-10^8 molecules/μl).

The template was ligated into a circle using 0.5 units of T4 DNA ligase, 1x T4 DNA ligase bf (66 mM Tris-HCl pH 7.6, 6.6 mM $MgCl_2$, 10 mM DTT, 66 μM ATP) and 10 nM ligation template, 5'Biotin- ttt ttt ttt ttt ttt gtc tgg aaa gca aac cgg tgc cca ccc atg a 3' SEQ ID N°1, in each reaction. After denaturation and subsequent addition of ligase to half of the reactions (see below), the samples were incubated at 37°C for 30 min and then the ligase was heat-inactivated at 65°C for 20 minutes.

After ligation, the samples were treated with exonucleases. Exonuclease V (5 units) was used for 30 min 37°C before heat-inactivation. The result was detected by performing a PCR with the following primers, 5 acg ccc acg gct gtc at 3' SEQ ID N°2 and 5' tgg acg tct gga aag caa a 3' SEQ ID N°3, (1 μM) located on both sides of the ligation junction. In 50 mM Tris HCl pH 8,3, 50 mM KCl, 200 μM dNTP, 0.125 u *Taq* GOLD polymerase (Perkin Elmer), 0.08xSYBR Green (Molecular Probes) as reporter molecule, and 1xROX (Molecular Probes) as standard, temperature cycles as follows 95°C 10 min activation of *Taq* polymerase followed by 40 cycles of 95°C 20 sec, 52°C 1 min, 72°C 20 sec. The experiments yield a cycle threshold value, Ct which is inversely proportional to the amount of starting material in the sample.

After the PCR amplification the reactions products were electrophoresed in a 3% agarose gel to ensure that a product of the correct length had been produced.

The results are shown in Figure 4, where A) Graph showing the fluorescence readings from a real-time PCR experiment read in an ABI 7700. The figures to the left corresponds to the numberings in B. Reactions were as follows; #2, 3 – No template control, #4 sample +ligase, #5 Sample –ligase, #6 sample + ligase + RCA, #7 Sample – ligase +RCA

B) A 3% agarose gel of the PCR reactions shown in A. Lane 1 in B is loaded with a 100 bp-ladder (lowest band around 50bp). Lane 2-7 corresponds to the same reactions. The arrow denotes the size for a correct length product.

Example 2

Enrichment of circular DNA over non-circular DNA through the use of different exonucleases.

BAC DNA as described in example 1 were cleaved and ligated as described in EXAMPLE 1. Half of the sample was ligated with T4 DNA ligase and half of the sample was not. The two reactions were further divided into five different reactions of each (+/- ligase) treated as follows.

1 5 u ExoV and 1 mM ATP,

2. 5 u ExoI, 50 u ExoIII and 25 u T7gene6
3. 5 u ExoI 50 u ExoIII and 2,45 u Lambda exo
4. 50 u ExoIII, 0,5 u ExoVII and 2,45 u Lambda Exo
5. 5 u ExoI, 0,5 u ExoVII in 1xTris buffer

All reactions were incubated at 37°C for 30 minutes before heat inactivation of the nucleases at 80°C for 20 minutes. The results were determined as described in example 1. After the PCR amplification, the reaction products were electrophoresed in a 3% agarose gel, and the nucleic acid visualised to ensure that a product of the correct length had been produced (not shown). The results are shown in table 1.

Table 1: Shows the result from an exonuclease treatment of cleave DNA that had been or had not been circularised with ligase.

Exo treatment:	+ligase: Ct value	–ligase: Ct value
*Exo*V(5U)+ATP(1mM)	24.49	35.82
*Exo*I (5U) +*Exo*III (50U) +T7Gen6 (25U)	21.33	33.52
*Exo*I (5U) +*Exo*III (50U) +λ*Exo* (2.45U)	21.52	35.07
*Exo*III (50U) +*Exo*VII (0.5U) +λ*Exo* (2.45U)	22.85	35.70
*Exo*I (5U)+ *Exo*VII (0.5U)+ 1x Tris bf	28.05	34.70

Example 3

Circularisation of DNA after denaturation of dsDNA, hybridisation of *Fok*I adapters, cleavage of the DNA at predetermined sites, specific circularisation of the cleaved fragment based on an SNP at the 5'prime end and enrichment of the circularised DNA. (See Figure 7)

BAC DNA was purified as described in example 1.

BAC DNA was diluted in a series and denatured by heat. After denaturation the samples were directly put on ice.

Different amounts (10^1-10^{10} molecules) of BAC DNA were cleaved with 2 units *Fok*I and 2 fmol FokI adapters (*Fok*I adapter 5'UTR 5' cgc atc cca cgt ggg atg cga aag caa aca ggg gt 3' SEQ ID N°4, *Fok*I adapter C2930T C-allele 5' gcc atc cgt gca cgg atg gct gca cag cac cgt gat 3' SEQ ID N°5, *Fok*I adapter C2930T T-allele 5' gcc atc cgt gca cgg atg gct gca cag cac cat gat 3' SEQ ID N°6) in 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 50 mM NaCl, 1xBSA1 for 2 hours 37°C before heat-inactivation of the enzyme.

The ends of the generated fragment nucleic acid were ligated into a circle using 8 fmol of the correct/incorrect ligation template (20+20 WDgDNA 5'UTR-Ex13 C-allele, 5' ctc ggc tct aaa gca aac agg tga tgg acg tct gga aag ctt t 3' SEQ ID N°7, 20+20 WDgDNA 5'UTR-Ex13 T-allele 5' ctc ggc tct aaa gca aac aga tga tgg acg tct gga aag ctt t 3' SEQ ID N°8). One unit T4 DNA ligase and 1x T4 DNA ligase buffer was used, and the reactions were incubated for approximately 30 minutes at 37°C before heat-inactivation of the DNA ligase. The circles were exonuclease treated with 5 units *Exo*V and 1 mM ATP and the samples were incubated in 37°C for 30 min before heat-inactivation at 80°C for 20 minutes.

PCR amplification was performed with primers (Frw WDgDNA 5'UTR-Ex13 5' cag agg tga tca tcc ggt ttg 3' SEQ ID N°9, Rew WDgDNA 5'UTR-Ex13 5' gga gag gag gcg cag agt gt 3' SEQ ID N°10), 0.5 μM of each, located on both sides of the ligation junction. With a total volume of 50 μl, 200 μM dNTP, 1 unit *Taq* GOLD polymerase, 1x PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatine). 40 amplification cycles were run after activation of the polymerase: 95°C 15 sec, 58°C 1 min and 72°C 20 sec. The amplified nucleic acids were detected by electrophoresis in a 3% agarose gel and visualisation by staining with ethidium bromide.

The results are shown in Figure 6 B. The following samples were loaded into the different lanes; 1 – Marker, 2 No template control, 3-8 samples from a 10-fold dilution series (10^{10}-10^1) of BAC DNA with correct ligation template

and ligase, 9 sample with correct ligation template but minus ligase, 10-12 samples from a 10-fold dilution series (10e10 to 10e9) with a ligation template corresponding to the wrong allele, T instead of C). The arrow denotes the size of a correct length product.

Example 4

Selective ligation of oligonucleotides cleaved with a structure specific enzyme. (See Figure 8)

The reactions were performed in 1x *Tth* buffer (1 mM NAD, 10 mM DTT and 0,1% Triton X-100). 20 µl reactions containing 0.5 pmol of the upstream, downstream and target oligonucleotides respectively (primer22+1 5' gta ttt gct ggg cac tca ctg ca 3' SEQ ID N°11, ArmC 5' tcc aga cgt cca tca cgg tgc tgt gca ttg cct g 3' SEQ ID N°12 or ArmT 5' tcc aga cgt cca tca tgg tgc tgt gca ttg cct g 3' SEQ ID N°13, Template2930 5'cag gca atg cac agc acc gtg cag tga gtg ccc agc aaa tac3' SEQ ID N°14), 1 unit of *Tth* ligase and native *Taq* polymerase. The reactions were prepared on ice and initiated by transfer to a Thermal Cycle where the following program was run: 95°C 20 sec, 72°C 30 min for 2 cycles. The upstream or downstream oligonucleotide was radio labelled and the samples were analysed on a 10 % denaturing polyacrylamide gel. Ten pmol target DNA was end-labelled with 1.65 pmol γ-^{32}P dATP (NEN). 4.9 U T4 PNK enzyme and 1x T4 PNK buffer (0.05 M Tris-HCl pH 7.6, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol) was added to each labelling reaction and the tubes were incubated for 45 min in 37°C. EDTA (1 mM) was added and the samples were boiled for 5 min in a water bath. The unincorporated nucleotides were removed from the labelling reaction with a MicroSpin™ G-50 column (Amersham Pharmacia Biotech).

The experiments with the radio labelled oligonucleotides were detected on a 10% polyacrylamide gel containing 7 M UREA. The gel was run with 0.5x Tris Borat EDTA buffer at 30 W for approximately 30 min and was dried in a gel dryer for 2 hours 80°C. The dried gel was exposed to a phosphorimager screen overnight.

The results are shown in Figure 8. Oligonucleotides yielding structure A was used in experiments 1-6 and oligonucleotides yielding structure B was used in experiments 7-12. (i) denotes the size of un-reacted oligonucleotide in experiments 1-6, (ii) the size for ligated product in reactions 1-6, (iii) uncleaved oligonucleotide used in reactions 7-12 and (iv) cleaved oligonucleotide in reactions 7-12. 32P denotes a radioactive label on respective oligonucleotide.

Lanes 1-6 shows the results from experiments with oligonucleotide 1 labelled with 32P. Lane 1, T-allel (wrong) -*Taq* polymerase, lane 2 C-allel (correct) – *Taq* polymerase, lane 3 T-allele –*Tth* ligase, lane 4 C-allel –*Tth* ligase, lane 5 T-allel , lane 6 C-allele..

Lanes 7-12 show the results from experiments with oligonucleotide 2 radio labelled with P^{32}. Lanes 7, 9, 11 is with the T-allele (incorrect) and lane 8,10,12 is with the C-allele (correct). Lanes 7-8 minus *Taq* polymerase, lanes 9-10 minus *Tth* ligase.

Lane 13 shows size markers.

Example 5

Circularisation of BAC DNA after cleavage with restriction enzymes, intramolecular hybridisation and cleavage with a structure specific enzyme followed by ligation, as shown in Figure 9.

BAC DNA was purified as described in example 1.

Denatured, ss BAC DNA ($1x10^{10}$ molecules) was cleaved with 10 units *Dra*III in buffer (10 mM NaCl, 5 mM Tris-HCl, 1 mM $MgCl_2$, 0.1 mM DTT pH 7.9) was used. *Dra*III was allowed to cleave the DNA for 1 hour 37°C before heat-inactivation.

This experiment was also done with genomic DNA. 10^{10} molecules were cleaved by 1 pmol of each adapter (cleave *Dra*III up2930 5' act gga cac aac gtg acg aac ttg ggt 3' SEQ ID N°15 and cleave *Dra*III down2930 5' cag ggc tca cac gca gtg agt gcc c 3' SEQ ID N°16) designed to hybridise to sequences in exon 13. The subsequent concerted structure-specific cleavage and ligation reaction contained the same reagents as above and 2 pmol of ligation any of

two different templates (20+20 *Dra*III-C 5' taa acg acc cgt gag tga cgc aca ggt cac ggg ggg ac 3' SEQ ID N°17 or 20+20 *Dra*III-G 5' taa acg acc cgt gag tga cgg aca ggt cac ggg ggg ac 3' SEQ ID N°18). The samples were divided into two parts and on one half was subjected to a RCA. A real-time PCR was performed on the samples with primers located on both sides of the ligation junction. In a total volume of 50 µl the following reagents were included: 2.5 µl sample, 1x PCR bf, 100 µM dNTP, 1 unit *Taq GOLD* polymerase, 0.5 µM of each primer, 1x ROX and 0.08x SYBR. After activating the polymerase 95°C for10 min, 40 cycles of the following program was run in a Thermal Cycler: 95°C 20 sec, 58°C 1 min and 72°C 30 sec.

*Dra*III and specific adapters designed to hybridise to sequences in exon 13 of *ATP7B* cleaved BAC DNA at predetermined sites. The target DNA was denatured to become single-stranded and the adapters were designed to create recognition and cleavage sites for *Dra*III. *Dra*III cleavage created a substrate that was used in the structure-specific cleavage, which generated the 5' located SNP.

The results are shown in Figure 8. Shown is the samples run on a 3% agarose gel. Lane 1 is a size marker, lane 2 and 3 PCR no template controls, lane 4 sample, lane 5 without *Taq* polymerase, lane 6 without *Tth* ligase and lane 7 without BAC DNA.

References

1. Stephens, J.C., *et al.*, *Haplotype variation and linkage disequilibrium in 313 human genes.* Science, 2001. 293(5529): p. 489-93.
2. Douglas, J.A., *et al.*, *Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies.* Nat Genet, 2001. 28(4): p. 361-4.
3. Bonnen, P.E., *et al.*, *Haplotypes at ATM identify coding-sequence variation and indicate a region of extensive linkage disequilibrium.* Am J Hum Genet, 2000. 67(6): p. 1437-51.
4. Drysdale, C.M., *et al.*, *Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness.* Proc Natl Acad Sci U S A, 2000. 97(19): p. 10483-8.
5. Liu, Q., *et al.*, *Overlapping PCR for bidirectional PCR amplification of specific alleles: a rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes.* Genome Res, 1997. 7(4): p. 389-98.
6. Ahmadian, A., *et al.*, *Analysis of the p53 tumor suppressor gene by pyrosequencing.* Biotechniques, 2000. 28(1): p. 140-4, 146-7.
7. Podhajska, A.J. and W. Szybalski, *Conversion of the FokI endonuclease to a universal restriction enzyme: cleavage of phage M13mp7 DNA at predetermined sites.* Gene, 1985. 40(2-3): p. 175-82.
8. Kim, S.C., A.J. Podhajska, and W. Szybalski, *Cleaving DNA at any predetermined site with adapter-primers and class- IIS restriction enzymes.* Science, 1988. 240(4851): p. 504-6.
9. Kool, E.T., *Circular oligonucleotides: new concepts in oligonucleotide design.* Annu Rev Biophys Biomol Struct, 1996. 25: p. 1-28.

   710. Baner, J., *et al.*, *Signal amplification of padlock probes by rolling circle replication.* Nucleic Acids Res, 1998. 26(22): p. 5073-8.
11. Lyamichev, V., M.A. Brow, and J.E. Dahlberg, *Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA*

polymerases. Science, 1993. 260(5109): p. 778-83.

12. Landegren, U., *et al.*, *A ligase-mediated gene detection technique.* Science, 1988. 241 http://www.biomednet.com/db/medline/88321655(4869): p. 1077-80.
13. Nilsson, M., *et al.*, *Padlock probes: circularizing oligonucleotides for localized DNA detection.* Science, 1994. 265(5181): p. 2085-8.
14. Syvanen, A.C., *et al.*, *A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E.* Genomics, 1990. 8(4): p. 684-92.
15. Ronaghi, M., *et al.*, *Real-time DNA sequencing using detection of pyrophosphate release.* Anal Biochem, 1996. 242(1): p. 84-9.
16. Lyamichev, V., *et al.*, *Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes.* Nat Biotechnol, 1999. 17(3): p. 292-6.
17. Heid, C.A., *et al.*, *Real time quantitative PCR.* Genome Res, 1996. 6(10): p. 986-94.
18. Tyagi, S. and F.R. Kramer, *Molecular beacons: probes that fluoresce upon hybridization.* Nat Biotechnol, 1996. 14(3): p. 303-8.
19. Griffin, T.J., *et al.*, *Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry.* Proc Natl Acad Sci U S A, 1999. 96(11): p. 6301-6.
20. Mashal, R.D., J. Koontz, and J. Sklar, *Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases.* Nat Genet, 1995. 9(2): p. 177-83.

CLAIMS

1. A method of enriching a preselected nucleic acid segment from a mixture of nucleic acid sequences, the preselected nucleic acid sequence encompassing a specific variant at a given position, the method comprising the steps of:

   (a) providing a nucleic acid mixture of sequences which includes the preselected nucleic acid segment to be enriched;

   (b) cleaving the nucleic acid sequences in the mixture to provide a nucleic acid fragment comprising the preselected nucleic acid segment;

   (c) providing a template oligonucleotide, one end of which hybridises to a sequence of the segment at or close to the variant position, and the other end of which hybridises to the end of a protecting sequence;

   (d) hybridising the template to the nucleic acid segment and to the protecting sequence such that the variant position and the end of the protecting sequence are brought into proximity of each other;

   (e) joining the end of the protecting sequence to the nucleic acid segment to form a ligated product, which ligated product is protected from degradation; and

   (f) enriching for the ligated product.

2. A method according to claim 1, step (b), in which cleavage generates a variant position located either at or close to one of the ends.

3. A method according to claim 1, step (f), in which the un-ligated nucleic acid is degraded.

4. A method according to claim 1, step (f) in which the ligated product is amplified.

5. A method according to claim 1, step (f), in which un-ligated nucleic acid is degraded and ligated product is amplified.

6. A method according to claim 1, step (b), in which the variant position is located at one end of the nucleic acid sequence and a defined nucleic acid sequence is located at the opposite end.

7. A method according to claim 1, claim 2 or claim 6 in which the cleavage generates a variant position located at or downstream of the 5' end of the segment.

8. A method according to any one of claims 1 to 7, in which degradation is effected with an exonuclease enzyme.

9. A method as claimed in any one of claims 1 to 8, in which the hybridisation of the template generates a structure that is a substrate for cleavage by a structure specific enzyme.

10. A method according to claim 9, in which the structure specific enzyme is selected from the group including any one or more of;

native or recombinant *Fen* nuclease,
native or recombinant *Mja* nuclease
recombinant polymerase from *Thermus aquatiqus*,
native or recombinant polymerase from *Thermus thermophilus*, and
native or recombinant polymerase from *Thermus flavus*.

11. A method according to claim 9, in which the structure specific enzyme is native polymerase from *Thermus aquaticus*.

12. A method according to any preceding claim, in which the protecting sequence is the opposite end of the nucleic acid segment, and in which the joining step (1e) circularises the nucleic acid segment.

13. A method according to any preceding claim, in which the nucleic acid segment comprises DNA.

14. A method according to any preceding claim, in which the nucleic acid segment comprises RNA.

15. A method according to any preceding claim, in which the amplifying step comprises rolling circle amplification.

16. A method according to any preceding claim, in which the protecting sequence is a nucleic acid adapter with a protected end.

17. A method according to any preceding claim, in which, cleavage of the nucleic acid sample in step (1b) uses at least one adapter and at least one restriction enzyme.

18. A method according to claim 17 in which, the one or more adapters hybridises to the nucleic acid sample such that, upon cleavage, the variant position is located at one end of the segment.

19. A method according to any preceding claim, in which there is a gap between the hybridised 3' end and the hybridised upstream sequence of the nucleic acid segment.

20. A method according to claim 19 in which, one of the sequences is extended to fill the gap there between.

21. A method according to claim 19 or claim 20, in which one of the sequences is extended with a polymerase.

22. A method according to claim 19, claim 20 or claim 21 in which the sequence is extended with an enzyme selected from the group including any or more of;

native or recombinant polymerase from *Thermus aquatiqus*,
native or recombinant polymerase from *Thermus thermophilus*, and
native or recombinant polymerase from *Thermus flavus*.

23. A method according to any one of claims 19 to 21, in which one of the sequences is extended with Klenow.

24. A method according to claim 19 in which, the gap is closed in with an oligonucleotide.

25. A method according to claim 24, in which the oligonucleotide inserted into the gap is labelled with an affinity tag.

26. A method according to claim 24 or claim 25, in which the oligonucleotide is labelled with biotin.

27. A method as claimed in any preceding claim, which includes an initial step of denaturing the nucleic acid when the nucleic acid is at least partially double stranded.

28. A method of analysing a target nucleic acid sequence for the presence of a mutation at a given variant position comprising the steps of:

a) providing a probe for the target sequence, the probe having a 3' end which is complementary to part of the target sequence, and

having a sequence downstream of the 5' end which is complementary to an adjacent part of the target sequence when the target sequence includes the mutation;

b) hybridising the probe with the target sequence, forming a non-hybridising region at the 5' end of the probe;

c) cleavage of the non hybridising region of the probe by a structure specific enzyme;

d) circularisation of the probe;

e) removal of uncircularised probe;

f) optionally amplifying the circularised probe;

g) and detection of the amplified product.

29. A method according to claim 28, in which the structure specific enzyme is selected from the group including any one or more of;

native or recombinant *Fen* nuclease,
native or recombinant *Mja* nuclease
recombinant polymerase from *Thermus aquatiqus*,
native or recombinant polymerase from *Thermus thermophilus*, and
native or recombinant polymerase from *Thermus flavus*.

30. A method according to claim 28, in which the structure specific enzyme is native *Taq* polymerase.

31. A method according to any one of claims 28 to 30, in which the removal of uncircularised probes in step (28e) comprises exonuclease treatment.

32. A method according to any one of claims 28 to 30, in which the removal of uncircularised probes (28e) is by binding to a portion in the 5' end of the probe that is cleaved off if hybridised to the sample

33. A method according to any one of claims 28 to 32, in which the amplification step (28f) comprises rolling circle amplification.

34. A method according to any one of claims 28 to 33, in which the detection of the product, step (28g) is by detection of the circularised probe directly or by detection of amplification products thereof.

35. A method according to any one of claims 28 to 34, in which a gap is present between the hybridised 3' end and the hybridised upstream sequence.

36. A method according to any one of claims 28 to 35, in which one of the sequences is extended to fill the gap therebetween.

37. A method according to claim 35 or claim 36, in which one of the sequences is extended with a polymerase.

38. A method according to claim 36 or claim 37, in which the sequence is extended with an enzyme selected from the group including any one or more of;

recombinant polymerase from *Thermus aquatiqus*,
native or recombinant polymerase from *Thermus thermophilus*, and
native or recombinant polymerase from *Thermus flavus*.

39. A method according to claim 36 or claim 37 in which one of the sequences is extended with native polymerase from *Thermus aquaticus*.

40. A method according to claim 36 or claim 37, in which one of the sequences is extended with Klenow.

41. A method according to claim 35 in which, the gap is closed in with a oligonucleotide.

42. A method according to claim 35 or 41 in which the oligonucleotide inserted into the gap is labelled with an affinity tag.

43. A method according to any one of claims 28 to 42, in which the oligonucleotide is labelled with biotin.

44. A method according to any one of claims 28 to 43 in which, the probe is labelled with a detectable cleavable molecule

45. A method according to claim 44 in which, the detectable cleavable molecule is fluorescent.

46. A method according to claim 44 or 45, in which cleavage of the probe is detected by fluorescence.

47. A method according to any preceding claim, in which the specific variant is a polymorphism, wherein the process includes a further step of variant scoring.

48. A method according to any preceding claim, in which the specific variant is a single nucleotide variant, wherein the process includes a further step of single nucleotide variant scoring.

49. A method according to any preceding claim, in which the specific variant is a deletion variant, wherein the process includes a further step of deletion variant scoring.

50. A method according to any preceding claim, in which the specific variant is an insertion variant wherein the process includes a further step of insertion variant scoring.

51. A method according to any preceding claim, in which the specific variant is a sequence variation wherein the process includes a further step of sequence variation scoring.

52. A method according to any preceding claim, in which the specific variant is a sequence length variation, wherein the process includes a further step of sequence length variation scoring.

53. A method of enriching a preselected nucleic acid segment from a mixture of nucleic acid sequences, comprising the steps of:

(a) providing a nucleic acid mixture of sequences which includes the preselected nucleic acid segment to be enriched;

(b) cleaving the nucleic acid sequences in the mixture to provide a nucleic acid fragment comprising the preselected nucleic acid segment;

(c) providing a template oligonucleotide, one end of which hybridises to a sequence of the segment, and the other end of which hybridises to the end of a protecting sequence;

(d) hybridising the template to the nucleic acid segment and to the protecting sequence such that the nucleic acid segment and the protecting sequence are brought into proximity of each other;

(e) joining the end of the protecting sequence to the nucleic acid segment to form a ligated product, which ligated product is protected from degradation; and

(f) enriching for the ligated product

54. A method according to claim 53, step (f), in which the un-ligated nucleic acid is degraded.

55. A method according to claim 53, step (f) in which the ligated product is amplified.

56. A method according to claim 53, step (f), in which un-ligated nucleic acid is degraded and ligated product is amplified.

57. A method according to claim 53, step (b), in which a pre-selected nucleic acid sequence is located at least at one of the ends.

58. A method according to any one of claims 53 to 57, in which degradation of steps (e) and (f) is effected with an exonuclease enzyme.

59. A method as claimed in any one of claims 53 to 58, in which the hybridisation of the template generates a structure that is a substrate for cleavage by a structure specific enzyme.

60. A method according to claim 59, in which the structure specific enzyme is selected from the group including any one or more of;

native or recombinant *Fen* nuclease,
native or recombinant *Mja* nuclease
recombinant polymerase from *Thermus aquatiqus*,
native or recombinant polymerase from *Thermus thermophilus*, and
native or recombinant polymerase from *Thermus flavus*.

61. A method according to claim 59, in which the structure specific enzyme is native polymerase from Thermus aquaticus.

62. A method according to any preceding claim, in which the protecting sequence is the opposite end of the nucleic acid segment, and in which the joining step (1e) circularises the nucleic acid segment.

63. A method according to any preceding claim, in which the nucleic acid segment comprises DNA.

64. A method according to any preceding claim, in which the nucleic acid segment comprises RNA.

65. A method according to any preceding claim, in which the amplifying step comprises rolling circle amplification.

66. A method according to claim 55, claim 56 or claim 65 in which amplification is performed with native or recombinant polymerase of phage phi29 or related polymerases.

67. A method according to any preceding claim, in which the protecting sequence is a nucleic acid adapter with a protected end.

68. A method according to any preceding claim, in which, cleavage of the nucleic acid sample in step (53b) uses at least one adapter and at least one restriction enzyme.

69. A method according to claim 67 in which, the one or more adapters hybridises to the nucleic acid sample such that, upon cleavage, the variant position is located at one end of the segment.

70. A method according to any preceding claim, in which there is a gap between the hybridised 3' end and the hybridised upstream sequence of the nucleic acid segment.

71. A method according to claim 69 in which, one of the sequences is extended to fill the gap there between.

72. A method according to claim 69 or claim 70, in which one of the sequences is extended with a polymerase.

73. A method according to claim 70, claim 71 or claim 72 in which the sequence is extended with an enzyme selected from the group including any or more of;

native or recombinant polymerase from *Thermus aquatiqus*,
native or recombinant polymerase from *Thermus thermophilus*, and
native or recombinant polymerase from *Thermus flavus*.

74. A method according to any one of claims 70 to 72, in which one of the sequences is extended with Klenow.

75. A method according to claim 70 in which, the gap is closed in with an oligonucleotide.

76. A method according to claim 75, in which the oligonucleotide inserted into the gap is labelled with an affinity tag.

77. A method according to claim 75 or claim 76, in which the oligonucleotide is labelled with biotin.

78. A method as claimed in any preceding claim, which includes an initial step of denaturing the nucleic acid when the nucleic acid is at least partially double stranded.

79. A method according to claim 55, claim 56, claim 65 or claim 66 in which the amplification product is labelled during amplification.

80. A method as claimed in any one of claims 53 - 79, in which the selected segment is subjected to a further step of nucleotide sequencing.

81. A method as claimed in any one of claims 53 - 79, in which the specific segment is subjected to a further step of single nucleotide variant scoring.

82. A method as claimed in any one of claims 53 - 79, in which the specific segment is subjected to a further step of deletion variant scoring.

83. A method as claimed in any one of claims 53 - 79, in which the specific segment is subjected to a further step of insertion variant scoring.

84. A method as claimed in any one of claims 53 - 79, in which the specific segment is subjected to a further step of sequence variation scoring.

85. A method as claimed in any one of claims 53 - 79, in which the specific segment is subjected to a further step of sequence length variation scoring.

86. A method as claimed in any one of claims 53 - 85, in which the specific segment is subjected to a further step of use as hybridisation probe for *in situ* analysis.

87. A method as claimed in any one of claims 53 - 85, in which the specific segment is subjected to a further step of hybridisation to an array.

88. A method as claimed in any one of claims 53 - 87, in which the specific segment is subjected to a further step of quantification.

Application No:	GB 0127704.5	**Examiner:**	Dr Rowena Dinham
Claims searched:	1-27 & 53-61; and 47-52 & 62-88 (in part)	**Date of search:**	4 November 2002

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.T):

Int Cl (Ed.7): C12Q 1/68

Other: ONLINE: WPI, EPODOC, JAPIO, CAPLUS, MEDLINE, EMBASE, BIOSIS, SCISEARCH

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
A, E	WO 02/057486 A2	(AMERSHAM BIOSCIENCES UK LTD) See especially page 4 line 19-27, Figure 4 and Examples	
A	WO 99/49079 A1	(LANDEGREN) See especially page 3 line 19- 29, page 8 line 2- page 9 line 1, Figure 4 and Examples	
A	US 6221603 B1	(MAHTANI) See especially column 6 line 50- column 7 line 29 and Examples	
A	*Nucl Acids Res*; Vol 29, pp e116 (2001). Qi *et al.* "L-RCA: a general method for genotyping....".See entire document		
A	*Nucl Acids Res*; Vol 28, pp e58 (2000). Antson *et al.* "PCR-generated padlock probes...." See entire document		

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.