

METHOD AND KIT FOR DETECTING A WILD-TYPE AND/OR A MUTATED TARGET DNA SEQUENCE

The present invention relates to a method and a kit
5 for detecting a wt target DNA sequence and/or a mutated target DNA sequence, which differ in that a single or multiple nucleotide substitution or deletion or insertion generates/eliminates a restriction site for a restriction endonuclease.

10 State of the art

Whole Genome Amplification on single or few cells is used to amplify DNA in order to allow different types of genetic analyses, including sequencing and SNP detection.

Whole Genome Amplification by means of a ligation-
15 mediated PCR (LM-PCR) based on a deterministic restriction site (hereinafter referred to as DRS-WGA) is known from EP1109938.

DRS-WGA has been shown to be better for the amplification of single cells (see for example: Lee YS, et
20 al: Comparison of whole genome amplification methods for further quantitative analysis with microarray-based comparative genomic hybridization. Taiwan J Obstet Gynecol. 2008, 47(1):32-41) and also more tolerant to DNA degradation due to fixative treatment (see for example:
25 Stoecklein N.H. et al: SCOMP is Superior to Degenerated Oligonucleotide Primed-PCR for Global Amplification of Minute Amounts of DNA from Microdissected Archival Samples. American Journal of Pathology 2002, Vol. 161, No. 1; Arneson N. et al.: Comparison of Whole Genome Amplification
30 methods for analysis of DNA extracted from microdissected early breast lesions in formalin-fixed paraffin-embedded tissue. ISRN Oncol. 2012; 2012;710692).

DRS-WGA DNA libraries comprise DNA fragments with the general structure shown in Figure 1A. Figure 1B shows a

specific example of the structure of the DNA library fragments obtained by DRS-WGA using the restriction endonuclease MseI.

Mutation detection assays downstream of DRS-WGA are normally carried out by designing primers within the restriction endonuclease (RE) amplicon. Although DRS-WGA provides best results in terms of uniform and balanced amplification, designing assays to determine the presence of mutations may be challenging in circumstances in which the mutation at issue generates or eliminates a restriction site for the DRS-WGA restriction endonuclease within the RE amplicon, because the usual way to design primers within the RE amplicon does not allow distinguishing the wild-type and the mutated DNA.

By way of explanation, examples of mutations giving rise to the above mentioned problem are shown hereinafter for the restriction site TTAA of the MseI restriction endonuclease, however the same problems occur with any other restriction site. The following examples should not be intended as limiting of the present invention, as it may apply also to other methods for DRS-WGA, including methods using a restriction endonuclease yielding blunt end DNA fragments.

Case A. A mutation introduces a new restriction site (RS)

Substitution

A substitution is a DNA mutation where one (or more) nucleotide(s) is (are) wrongly replaced with a different nucleotide. This generates a change in the nucleotide sequence of the particular DNA site.

The substitution may therefore introduce a RS in the mutated (M) DNA sequence where no RS was present in the wild type (WT) DNA sequence.

As an example for a single base-substitution:

WT Sequence	M Sequence	Case
-------------	------------	------

VTAA **TTAA** (1)

TVAA **TTAA** (2)

TTBA **TTAA** (3)

TTAB **TTAA** (4)

5 where V is A or C or G (not T), and B is C or G or T (not A).

Deletion

10 A DNA mutation may remove one (or more) nucleotide(s) producing a RS in the mutated (M) DNA sequence where no RS was present in the wild-type (WT) DNA sequence.

E.g. for single or multiple (n) base deletions:

WT Sequence	M Sequence	Case
T[V] _n TAA	TTAA	(5)
TT[V] _n AA	TTAA	(6)
15 TT[B] _n AA	TTAA	(7)
TTA[B] _n A	TTAA	(8)

Insertion

20 A DNA mutation may insert one (or more) nucleotide(s) producing a RS in the mutated (M) DNA sequence where no RS was present in the wild type (WT) DNA sequence.

E.g. for a single base insertion:

WT Sequence	M Sequence	Case
VTAA	[insT]TAA	(9)
TTAB	TT[insA]AB	(10)

25 and the related indistinguishable cases of:

VTAA	T[insT]AA	(9')
TTAB	TTA[insA]B	(10')

30 All of the above mutations introduce a RS, resulting in the mutation not being detectable in the DNA library fragment e.g. by PCR and Sequencing, when using primer pairs amplifying a region comprising the mutation site, as only the wild-type allele (if present) will be correctly amplified and sequenced. This situation is outlined in Figure 2, Case A, left inset.

Case B. The mutation removes the restriction site from the wild-type (WT) sequence

Substitution

A substitution may remove the RS present in the WT DNA sequence.

	WT Sequence	M Sequence	case
5	TTAA	VTAA	(11)
	TTAA	TVAA	(12)
	TTAA	TTBA	(13)
10	TTAA	TTAB	(14)

The above correspond to cases (1)-(4) where the M DNA sequence and the WT DNA sequence are swapped.

Deletion

A DNA mutation may remove one (or more) nucleotide(s) removing a RS where a RS was present in the wild type (WT) DNA sequence.

E.g. for single base deletions:

	WT Sequence	M Sequence	Case
	VTTAA	V[delT]TAA	(15)
20	TTAAB	TT[delA]AB	(16)

and the related indistinguishable cases of:

	VTTAA	VT[delT]AA	(15')
	TTAAB	TTA[delA]B	(16')

Insertion

A DNA mutation may insert one (or more) nucleotide(s) removing a RS where a RS was present in the wild type (WT) DNA sequence.

	WT Sequence	M Sequence	Case
	TTAA	T[insV] _n TAA	(17)
30	TTAA	TT[insV] _n AA	(18)
	TTAA	TT[insB] _n AA	(19)
	TTAA	TTA[insB] _n A	(20)

Any (and many other) cases comprising deletion of one or more bases as in the example above will remove the RS

existing in a WT sequence, resulting in a non-digested sequence.

While the mutated sequence may readily be identified designing primer pairs amplifying the DNA sequence comprising the mutation site, the wild-type allele (if present) fails to be amplified, giving an incorrect assessment of the genotype. This situation is outlined in Figure 2, Case B, right inset.

Moreover, when there is no mutation, there would be no signal at all from the PCR, and it would be impossible to determine whether there was a drop-out of the wild-type allele during DRS-WGA or the genotype is simply wild-type.

EP1350853 discloses the amplified fragment length polymorphism (AFLP) technique that reveals polymorphisms at restriction sites. The method for detecting sequence polymorphisms between one or more genomes comprises producing a nucleic acid fingerprint from said genomes by (a) providing from a starting nucleic acid a plurality of adaptor ligatable nucleic acid fragments with ends that are compatible to at least one adaptor, and wherein said nucleic acid fragments are obtained by fragmentation with restriction endonucleases; (b) performing a ligation reaction between said ends of said nucleic acid fragments and said at least one adaptor such as to produce adaptor-ligated nucleic acid fragments; (c) amplifying said adaptor-ligated nucleic acid fragments by using at least one amplification primer essentially complementary to the nucleotide sequence of said at least one adaptor; and (d) generating from said amplified adaptor-ligated nucleic acid fragments a nucleic acid fingerprint; comparing the obtained nucleic acid fingerprints for the presence or absence of, or differences between, amplified nucleic acid fragments such as to determine the presence of sequence polymorphisms.

This method however does not allow the detection of one specific polymorphic site.

An object of the present invention is therefore to provide a method for detecting a wild-type target DNA sequence and/or a mutated target DNA sequence in a library of DNA fragments having a structure such as that obtained by DRS-WGA, where the wild-type target DNA sequence and the mutated target DNA sequence differ in the presence of a restriction site for the restriction endonuclease of the DRS-WGA, that solves the above cited problems in a simple and efficient manner.

This object is achieved by the present invention as it relates to a method as defined in claim 1.

It is a further object of the present invention to provide a kit as defined in claim 10.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although many methods and materials similar or equivalent to those described herein may be used in the practice or testing of the present invention, preferred methods and materials are described below. Unless mentioned otherwise, the techniques described herein for use with the invention are standard methodologies well known to persons of ordinary skill in the art.

By the term "restriction site" or "RS" there is intended the sequence of nucleotides (typically 4-8 base pairs in length) along a DNA molecule recognized by the restriction endonuclease (or "RE"). At the restriction site, the restriction endonuclease cleaves nucleotides by hydrolysing a phosphodiester bond between them.

By the term "mutation-dependent restriction site" (or

"MDRS"), there is intended the RS which is introduced or removed by effect of the mutation.

By the term "cleavage site" (or "CS"), there is intended the position in the sequence of the restriction site in which the phosphodiester bonds hydrolysed by the RE are located.

By the term "mutation-dependent cleavage site" (or "MDCS"), there is intended the CS which is introduced or removed by effect of the mutation.

By the term "amplicon" there is intended a region of DNA produced by PCR amplification.

By the terms "DRS-WGA amplicon" or "WGA amplicon", there is intended a DNA fragment amplified during DRS-WGA, comprising a DNA sequence between two RS flanked by the ligated WGA primers.

By the terms "WGA PCR primer" or "universal WGA primer" or "adaptor", there is intended the additional oligonucleotide ligated to each fragment generated by the action of the restriction enzyme in DRS-WGA.

By the term "original DNA", there is intended the genomic DNA (gDNA) prior to amplification with DRS-WGA.

By the term "target sequence", there is intended the region of interest on the original DNA.

By the term "target sequence sense strand" there is generally intended the segment of the DNA strand running from 5' - 3', which has the same sequence as the mRNA and is complementary to the antisense strand. The sense strand may also be referred to as "positive strand".

For the sake of simplicity, in the present description, the term "target sequence positive strand" (TSPS) will be used with the following meanings:

- 1) it identifies the genomic DNA sequence by increasing nucleotide number in the case in which the mutation occurs on the 3' side of the mutation-

dependent cleavage site on the sequence with increasing nucleotide number

- 2) it identifies the reverse complementary of the genomic DNA sequence by increasing nucleotide number in the case in which the mutation occurs on the 5' side of the mutation dependent cleavage site position on the sequence with increasing nucleotide number

Coherently, the term "target sequence antipositive strand" (TSAS) will be used in the present description with the following meanings:

- 3) it identifies the reverse complementary of the genomic DNA sequence by increasing nucleotide number in the case in which the mutation occurs on the 3' side of the mutation-dependent cleavage site on the sequence with increasing nucleotide number
- 4) it identifies the genomic DNA sequence by increasing nucleotide number in the case in which the mutation occurs on the 5' side of the mutation dependent cleavage site position on the sequence with increasing nucleotide number

The expression "increasing nucleotide number" refers to numbering relative to the chromosome position (as found in sequence databases such as UCSC Genome Browser).

By the expression "5' end region of a sequence segment" there is intended that the localisation of the sequence of nucleotides referred to is towards the 5' terminal end of the sequence segment.

By the expression "3' end region of a sequence segment" there is intended that the localisation of the sequence of nucleotides referred to is towards the 3' terminal end of the sequence segment.

Brief description of the drawings

Figure 1A shows a sketch of the general structure of a DNA library fragment obtained by a specific DRS-WGA, using the restriction endonuclease MseI, when a mutation dependent cleavage site (MDCS) is not cleaved. Acronyms are as follows: FTS = first target sequence; CS = cleavage site; RS = restriction site; FTSPS = first target sequence positive strand; FTSAS = first target sequence antipositive strand; NUCL# = nucleotide number (increasing according to the direction of the arrow); WT = wild-type; MDCS = mutation dependent cleavage site.

Figure 1B shows a sketch of the general structure of a DNA library fragment obtained by a specific DRS-WGA, using the restriction endonuclease MseI, when a mutation dependent cleavage site (MDCS) is cleaved. Additional acronyms are as follows: STS = second target sequence; MDRS = mutation dependent restriction site; M = mutated; STSPS = second target sequence positive strand; STSAS = second target sequence antipositive strand.

Figure 1C shows a sketch of the first target sequence positive and antipositive strands and location of related reverse and forward primers. Additional acronyms are as follows: R1 = first reverse primer; F2 = second forward primer.

Figure 1D shows a sketch of the second target sequence positive and antipositive strands and location of related reverse and forward primers. Additional acronyms are as follows: F3 = third forward primer; F31 = first portion of third forward primer; F32 = second portion of third forward primer.

Figure 1E shows a sketch of the first target sequence positive and antipositive strands and location of related reverse and forward primers, when the mutation is located on the 5' side of the MDCS on the sequence by increasing nucleotide number.

Figure 1F shows a sketch of the second target sequence positive and antipositive strands and location of related reverse and forward primers, when the mutation is located on the 5' side of the MDCS on the sequence by increasing
5 nucleotide number.

In Figures 1A-1F, reference is made to the situation where the non-cleaved sequence is the WT sequence, and the cleaved sequence is the Mutant sequence. The alternative situation where the mutated sequence is non-cleaved and the
10 wild-type sequence is cleaved can be readily obtained by simply swapping WT with M.

Figure 2 shows a simplified sketch of the two situations comprising the introduction (Case A - left inset) or removal (Case B) of a restriction site in the
15 mutated DNA sequence and the consequences with traditional mutation detection methods.

Figure 3 shows an image of a gel electrophoresis of the separated products of a PCR amplification performed with a bivalent primer pair for wt and mutated DNA of
20 Example 1. Cnt: blank of WGA reaction. C-: blank of PCR reaction.

Figure 4A and 4B shows simplified sketches of the working principle of the method according to the invention. Figure 4A shows the situation in which the mutation
25 introduces a restriction site in the sequence. Figure 4B shows the situation in which the mutation removes the restriction site in the sequence.

Figure 5 shows an image of a gel electrophoresis of the separated products of a PCR amplification performed with a mutated specific 5'primer including the restriction
30 site, homologous by 86% of its length to the universal WGA primer.

Figure 6 shows an image of a gel electrophoresis of the separated products of a PCR amplification performed with the wild-type specific 5'primer of Example 3.

5 Figure 7 shows an image of a gel electrophoresis of the separated products of a PCR amplification performed with the mutated specific 5'primer of Example 3.

Figure 8 shows an example of sequencing of a wild-type allele of Example 3.

10 Figure 9 shows an example of sequencing of a mutated allele of Example 3.

Figure 10 shows a table summarising the results of Example 4.

Figure 11 shows an image of a gel electrophoresis of the separated products of a PCR amplification of M and WT individual cells performed with the mutated primer pair of Example 5.

20 Figure 12 shows an image of a gel electrophoresis of the separated products of a PCR amplification of M and WT individual cells performed with the wild-type primer pair of Example 5.

Figure 13 shows an example of reverse strand sequence of a wild-type single cell of example 6.

Figure 14 shows an example of reverse strand sequence of a mutated single cell of example 6.

25 Figure 15 shows a table summarising the results of Example 6.

Figure 16 shows an example of reverse strand sequence of a wild-type cell (a leukocyte) positive for mutant PCR product (i.e. a false positive for PCR product alone), which was instead confirmed to be wild-type by the assay as in Example 6.

Detailed description of the invention

The method according to the present invention for detecting at least one of at least one first target DNA

sequence and at least one second target DNA sequence from a library of DNA sequences comprises steps (a) to (c). The first target DNA sequence differs from the second target DNA sequence in that a single or multiple nucleotide substitution or deletion or insertion in the second sequence generates a restriction site for a restriction endonuclease. With reference to Figure 2, case A, left inset, and Figure 4A, the first target DNA sequence corresponds to the wild-type DNA sequence and the second target DNA sequence corresponds to the mutated DNA sequence, whereas with reference to Figure 2, case B, right inset, and Figure 4B, the first target DNA sequence corresponds to the mutated DNA sequence and the second target DNA sequence corresponds to the wild-type DNA sequence.

In step (a), the library of DNA sequences is provided. Each of the DNA sequences of the library comprises, respectively from the 5' end to the 3' end, a first sequence segment having a length from 15 to 50 nucleotides, a second sequence segment of genomic DNA as cleaved by the restriction endonuclease, and a third sequence segment reverse complementary to the union of the first sequence segment and, if any, the 5' overhang generated by the RE. With reference to Figure 1A, numeral 1 shows the first sequence segment, numeral 2 shows the second sequence segment, and numeral 3 shows the third sequence segment. In a preferred embodiment the first sequence segment corresponds to the WGA PCR Primer.

The restriction endonuclease is preferably MseI.

In step (b), the library of DNA sequences is amplified by PCR using:

- at least one first reverse primer which hybridises to the 3' end region of the second sequence segment of the at least one first or second target sequence positive

strand;

- at least one second forward primer which hybridises to the 3' end region of the second sequence segment of the at least one first target sequence antipositive strand;

5 - at least one third forward primer comprising a first portion hybridising to the 5' end region of the third sequence segment of the at least second target sequence antipositive strand and a second portion hybridising to the 3' end region of the second sequence segment of the at least one second target sequence antipositive strand, wherein the first portion of the at least one third forward primer has a length from 20% to 80% with respect to the total length of the at least one third forward primer.

15 The third forward primer is hereinafter sometimes referred to in short as "hybrid primer".

Preferably, at least one fourth reverse primer which hybridises to the 3' end region of the second sequence segment of the at least one second target sequence positive strand is used in step (b).

20 Preferably, the first portion of the at least one third forward primer has a length from 40 to 60%, with respect to the total length of the at least one third forward primer.

25 Preferably, the second portion of the at least one third forward primer has a length in bases comprised between a minimum corresponding to the consensus sequence of the restriction endonuclease minus, if any, the 5' overhang generated by the restriction endonuclease, all divided by two, and a maximum of 30 bases.

30 With reference to Figure 4A, the first reverse primer corresponds to the wild-type reverse primer (WT_R), the second forward primer corresponds to the wild-type forward primer (WT_F), the third forward primer corresponds to the mutated forward primer (M_F). In one embodiment, the first

reverse primer serves to amplify not only the first, but also the second target sequence positive strand. In a preferred embodiment, however, a fourth reverse primer which differs from the first reverse primer, is used to
5 amplify the second target sequence positive strand. In Figure 4A, the fourth reverse primer corresponds to the mutated reverse primer (M_R).

The same principle applies in Figure 4B, where the first reverse primer corresponds to the mutated reverse
10 primer (M_R), the second forward primer corresponds to the mutated forward primer (M_F), the third forward primer corresponds to the wild-type forward primer (WT-F), and the fourth reverse primer corresponds to the wild-type reverse primer (WT_R).

15 In step (c), the DNA sequences amplified in step (b) are detected. Step (c) may be performed by several detection methods known in the art, for example gel electrophoresis, capillary electrophoresis, DNA sequencing. Preferably, step (c) is performed by a DNA sequencing
20 method. Even more preferably, the DNA sequencing method is Sanger sequencing, or sequencing by synthesis.

The method of the present invention may be used with any library of DNA sequences having the structure shown in Figure 1A. The method is preferably used with a library of
25 DNA sequences obtained by deterministic restriction site whole genome amplification.

According to the present invention there is also provided a kit comprising a first and/or a second and/or a third primer as defined above.

30 More specifically, the kit for detecting at least one of at least one first target DNA sequence and at least one second target DNA sequence from a library of DNA sequences, wherein the first target DNA sequence differs from the second target DNA sequence in that a single or multiple

nucleotide substitution or deletion or insertion in the second sequence generates a restriction site for a restriction endonuclease, and wherein each of the DNA sequences of the library comprises, respectively from the
5 5' end to the 3' end, a first sequence segment having a length from 15 to 50 nucleotides, a second sequence segment of genomic DNA as cleaved by the restriction endonuclease, and a third sequence segment reverse complementary to the union of the first sequence segment and, if any, the 5'
10 overhang generated by the restriction endonuclease, comprises:

- at least one first reverse primer which hybridises to the 3' end region of the second sequence segment of the at least one first or second target sequence positive
15 strand;

- at least one second forward primer which hybridises to the 3' end region of the second sequence segment of the at least one first target sequence antipositive strand;

- at least one third forward primer comprising a first
20 portion hybridising to the 5' end region of the third sequence segment of the at least second target sequence antipositive strand and a second portion hybridising to the 3' end region of the second sequence segment of the at least one second target sequence antipositive strand,
25 wherein the first portion of the at least one third forward primer has a length from 20% to 80% with respect to the total length of the at least one third forward primer.

The kit preferably further comprises at least one fourth reverse primer which hybridises to the 3' end region
30 of the second sequence segment of the at least one second target sequence positive strand.

The kit may be used to detect any kind of mutation generating or eliminating a restriction site for the restriction endonuclease of the ends of the second sequence

segment of the DNA fragments of the library. The kit is preferably used in the diagnosis of mutations in the (anaplastic lymphoma kinase) ALK or (epidermal growth factor receptor) EGFR or (phosphatidylinositol 3-kinase catalytic alpha polypeptide) PIK3CA gene.

Examples

Example 1 - Bivalent primer approach

Preliminary tests were carried out on SY5Y cell lines (SH-SY5Y ATCC Catalog No. CRL-2266™), which harbour a heterozygous C to A substitution at codon 1174 of the ALK gene, turning a Phenylalanine into a Leucine (F1174L); considering the flanking sequence, the heterozygous substitution introduces one new restriction site (RS) in the mutated allele, whereas the wild type allele does not have any RS.

Figure 2 is a simplified sketch of the sequences and transformations in the WGA DNA library produced by the mutation.

To detect mutations occurring on the RS, the following approach was tested. The universal primer of the whole genome amplification (DRS-WGA primer, SEQ ID NO:1 having sequence AGTGGGATTCCTGCTGTCAGT) was exploited to design a 5' primer in a new PCR primer pair where the 3' primer overlaps a region in 3' with respect to the RS.

The strategy consisted in designing a bivalent primer pair comprising:

- a 5' primer having 95% homology with the DRS-WGA primer; and

- a 3' PCR primer which should provide the specificity required to the PCR, to selectively amplify the target region instead of other DRS-WGA amplicons.

This bivalent primer pair should in theory serve for the amplification of wild-type (WT) sequence and mutant (M) sequence.

Experimental evidence shows that this approach results to be poor and improper, and cannot guarantee the detection of the mutation at the RS. As shown in Figure 3, the use of a bivalent primer provides unspecific amplification, which
 5 results in many amplification bands having different sizes, and no clearly distinguishable bands of the expected size (e.g. on single SY5Y cells, bearing a F1174L heterozygous mutation, isolated with DEPAarray™ and amplified with DRS-WGA. 132bp for the mutated sequence, 169bp for the WT
 10 sequence). The amplification failed to give a clear and specific band both in mutated (M), wild-type (WT) and PCR negative control (C-) samples. The negative control of WGA (Ctr-), shows just an unspecific band.

One factor contributing to this poor result is that
 15 the 5' bivalent primer which corresponds by 95% to the ligated WGA-primer, is present on all DNA fragments of the DRS-WGA library, and the 3' bivalent primer does not provide the PCR reaction with sufficient specificity.

As an example, the human genome reference (Homo
 20 Sapiens hg 19) comprises 3,095,693,981 bases. If the genome is digested with a restriction endonuclease with a four base restriction site (e.g. TTAA), the mean length of the DNA fragments generated is 4 (the possible bases) to the power of 4 (the digestion sequence length considered) =
 25 256. The generated DNA library would thus comprise approximately $3,095,693,981/256 \sim 12.1$ million different fragments, with a simplified assumption of a random sequence of the nucleotides in the DNA. All of them would comprise the same 5' primer (corresponding to the WGA-
 30 primer from the primary PCR).

The use of the bivalent primer pair therefore gives unspecific bands.

Example 2 - Hybrid primer homology range limit

Amplification tests were carried out on the same SY5Y

cell line as that used in Example 1, but using the method of the present invention.

To test the amplification of both wild-type (WT) and mutated (M) alleles in DRS-WGA products, individual SY5Y cells were isolated with DEPArray™, which provides pure single cells.

The amplification approach of using one 5' PCR primer matching the WGA universal primer by 86% its length provided a solution for the amplification neither of the WT nor of the M allele. As shown in Figure 5, the amplification failed to give a clear and specific band both in mutated (M), wild-type (WT) and PCR negative control (C-) samples. The negative control of WGA (Ctr-), shows just an unspecific band.

Primers having different percentages of homology with the WGA universal primer were tested. The results are summarised in following Table 1.

Table 1

Primer	Homology to the WGA-primer		Homology to Original DNA		F32 [# of basis]	TEST
Universal	21/22	95%	1/22	5%	0	KO
Mutant 1	19/22	86%	3/22	14%	1	KO
Mutant 2	10/20	50%	10/22	50%	8	OK
Mutant 3	14/22	64%	8/22	36%	6	OK

In Table 1, column F32 reports the length in number of bases of the "second portion of third forward primer" i.e. the primer portion which has the same sequence as the original DNA, excluding the restriction endonuclease overhang.

It is clear from the results of Table 1 that a balanced compromise needs to be identified to meet the method requirements. Several tests have shown that the

ideal percentage of identity of the hybrid primer with the WGA universal primer is from 20% to 80%, with an even better efficiency in the range from 40% to 60%.

Example 3 Introduction of a new RS in the mutant allele (ALK gene) - assay design

The method according to the invention guarantees amplification (and sequencing) even in the case of incomplete digestion by the restriction endonuclease. In fact, the activity of the restriction endonuclease is not guaranteed for all the RS in the target DNA, and statistically a small percentage of undigested RS is present in the DRS-WGA, which nevertheless are successfully amplified by DRS-WGA, albeit with the WGA-primer (primary - PCR) being in another RS.

In case of an undigested site the use for the mutation assay of just one primer pair designed for the mutant sequence would not allow the amplification and the sequencing of the target.

Amplification tests were again carried out on the SY5Y cell line, which - as previously disclosed - harbours a heterozygous C to A substitution at codon 1174, turning a Phenylalanine into a Leucine (F1174L). The heterozygous substitution thus introduces a new RS in the mutated allele, whereas the wild-type allele does not have any RS.

The PCR primer sequences used for the amplification of WT and M alleles are shown in Table 2. For mutant allele forward primer, the first portion of the primer sequence homologous to the WGA primer is shown in bold and underlined, while the second portion of the primer which has the same sequence as the original DNA, excluding the restriction endonuclease overhang, (F32 = 8 basis) is shown boxed.

Table 2

Primer

Sequence

Name

ALK_WT_F	SEQ ID NO:2	5'	CCTCTCTGCTCTGCAGCAAAT	3'
ALK_WT_R	SEQ ID NO:3	5'	TCTCTCGGAGGAAGGACTTGAG	3'
ALK_M1_F	SEQ ID NO:4	5'	<u>TGCTGTCAGT</u> TA <u>AACCACCA</u>	3'
ALK_M1_R	SEQ ID NO:5	5'	GGTCTCTCGGAGGAAGGACT	3'

To test the amplification of both WT and M alleles in DRS-WGA products, individual SY5Y cells were isolated with DEPArray™, which provides pure single cells.

5 As negative control for the mutation detection, individual lymphocytes were also isolated with DEPArray™ and amplified with DRS-WGA.

The PCR amplification of the WT allele on both WT (lymphocytes) and heterozygous M (SY5Y) was achieved
10 perfectly by the use of the specifically designed WT 5' primer, which allows the exclusive amplification of the WT allele.

As can be observed in Figure 6, there are no aspecific amplification products. Instead, the expected PCR band (132
15 bp) is clearly distinguishable.

The M-specific 5' primer was tested for the same lymphocytes and SY5Y cells to detect the specificity of the amplification provided by the primer designed straddling the target sequence and the universal DRS-WGA primer.

20 As may be seen in Figure 7, in this case, as expected, the specific amplification was obtained only in the SY5Y single cell DRS-WGA DNA. DRS-WGA DNA from lymphocytes, being WT for the target mutation, was negative for the expected amplification, and only unspecific PCR
25 amplifications were present.

To demonstrate that the amplification achieved was specific and allowed sequencing, all the amplification products were sequenced from their 3' end. The corresponding WT or M status was confirmed for all

amplification products showing the specificity achieved with the described method. An example of sequencing of a WT allele is shown in Figure 8, whereas an example of sequencing of a M allele is shown in Figure 9.

5 Results are summarised in Table 3.

Table 3

	Single Cells Replicates	Sequence obtained with M-Specific 5'primer	Sequence obtained with WT-Specific 5'primer
WBC	1	No PCR Product	WT
	2	No PCR Product	WT
	3	No PCR Product	WT
SY5Y	1	M	WT
	2	M	WT
	3	M	WT

In a preferred embodiment, the second portion (F32) of the third forward primer (F3) is shorter than 30 nucleotides so as not to mis-prime on the first target sequence antipositive strand (FTSAS) - i.e. the wild-type sequence in this example -, thus starting a PCR reaction which may result in a false-positive (as per its PCR product length and sequence). More preferably, the length of the second portion (F32) is shorter than 20 nucleotides. Even more preferably, the length of said second portion (F32) of said third forward primer (F3) is shorter or equal to 10 nucleotides.

The second portion (F32) of the third forward primer (F3) should not be too short as to not provide enough specificity, (see for example results in table I). In particular the length of said second portion of the third forward primer, should be greater than the restriction site consensus sequence length minus the length of the 5'

overhang of the digested DNA, all divided by two. In order to obtain a greater specificity, the second portion (F32) of the third forward primer (F3) should be at least 3 nucleotides, and even more preferably at least 6 nucleotides, longer than the restriction site consensus sequence length minus the length of the 5' overhang of the digested DNA, all divided by two.

Example 4 Introduction of a new RS in the mutant allele (ALK gene) - assay validation

The method described above has been further validated with 54 single cells:

- 10 single live, fresh SY5Y;
- 19 single SY5Y, previously fixed with 2% paraformaldehyde (PFA) 20 minutes at room temperature, and permeabilised with Inside Perm (Miltenyi Biotec);
- 19 single SY5Y, previously fixed with CytoChex™, and permeabilised with Inside Perm;
- 2 single fresh, live lymphocytes;
- 2 single lymphocytes, previously fixed with 2% PFA 20 minutes at room temperature, and permeabilised with Inside Perm (Miltenyi Biotec).

The method amplified the WT allele in 100% of SY5Y and lymphocytes cells, and the mutant allele was amplified in 9/10=90% of live SY5Y, 16/19=84% of SY5Y cells fixed & permeabilised with cyto-chex/inside-perm, 17/19=89% of SY5Y cells fixed & permeabilized with PFA 2% 20' @ room temperature/inside-perm, and 0/4=0% of lymphocytes.

Results are shown in Figure 10 and summarised in Table 4.

Table 4

ALK	
PCR of WT Allele	PCR of F1174L M Allele

SY5Y	Live	100%	90%
	CytoChex, Inside Perm	100%	84%
	PFA, Inside Perm	100%	89%
Lymphocytes	Live	100%	0%
	PFA, Inside Perm	100%	0%

These results show the efficacy and robustness of the method of the present invention on larger numbers of samples.

5 Example 5 Removal of a RS in the mutant allele (EGFR gene) - assay design

Amplification tests were carried out on the HCC-827 cell line, harbouring a deletion of 5 codons in the EGFR gene. The deletion removes a restriction site (RS),
10 allowing the detection of the M allele, but not of the WT allele which has the RS, when using a single PCR and primer pairs on the human genome.

Individual HCC-827 cells were isolated with DEPArray™, along with lymphocytes as a control of the WT condition.

15 Two different primer pairs targeted for the M allele (with the deleted RS) and WT allele (still maintaining the RS) were designed and led to the correct identification of both WT and M conditions.

The PCR primer sequences used for the amplification of
20 WT and M alleles are shown in Table 5. For wild-type allele forward primer, the first portion of the primer sequence homologous to the WGA primer is shown in bold and underlined, while the second portion of the primer which has the same sequence as the original DNA, excluding the
25 restriction endonuclease overhang, (F32 = 16b) is shown boxed.

Table 5

Primer	Sequence
Name	

Ex19_M_F	SEQ ID NO:6	5'TAAAATTCCCGTCGCTATCAA3'
Ex19_M_R	SEQ ID NO:7	5'TGTGGAGATGAGCAGGGTCTAG3'
Ex19_WT_F	SEQ ID NO:8	5'CTGTCAGT TAA GAGAAGCAACATCTCC 3'
Ex19_WT_R	SEQ ID NO:9	5'AGAGCAGCTGCCAGACATGAG3'

Figure 11 shows the results of PCR amplification of M and WT individual cells with M primer pairs, while Figure 12 shows the results of PCR amplification of M and WT individual cells with WT primer pairs.

Figure 13 shows a reverse strand sequence of a WT single cell, compared to the gDNA amplified with DRS-WGA, while Figure 14 shows a reverse strand sequence of a M single cell, compared to the gDNA amplified with DRS-WGA.

10 Example 6 Removal of a RS in the mutant allele (EGFR gene)-assay validation

The method described above has been further validated with 60 single cells:

- 31 single HCC-827, treated according Veridex CellSearch enrichment protocol;
- 11 single lymphocytes, treated according Veridex CellSearch enrichment protocol;
- 17 single fresh, live lymphocytes.

The method amplified the WT allele in 28/31=90% of the single HCC-827 and the M allele in 31/31=100% of the single HCC-827.

Considering the 11 Veridex-treated lymphocytes, 11/11=100% resulted in a positive PCR product for the WT PCR, 3/11=27% resulted in a positive PCR product for the M-PCR. These products were sequenced and confirmed to be WT. Hence, detecting the DNA by sequencing, the specificity on Veridex-treated lymphocytes is still 100%, whereas, just relying on the PCR positivity the specificity is (in this test) 8/11=73%. Detecting the DNA product length by gel electrophoresis would similarly allow to distinguish the

length and determine that actually it is WT; detecting the DNA product by real-time PCR would not distinguish between WT and M products. Considering the 17 fresh lymphocytes, 17/17=100% resulted in a positive PCR product for the WT PCR, 0/17=0% resulted in a positive PCR product for the M-PCR. These products were sequenced and confirmed to be WT.

As there are 2 WT alleles per lymphocyte, the difference in undigested RS between Veridex-treated (3/22=14%) and fresh lymphocytes (0/34=0%) is statistically significant.

This demonstrates the robustness of the above described method in case of incomplete RE digestion activity.

Results are shown in Figure 15 and summarised in Table 6.

Table 6
EGFR Exon19

Treatment		n	PCR of WT Allele	PCR of Del. E746_A750 M Allele
HCC-827	Veridex	31	90%	100%
WBC	Veridex	11	100%	27% (*)
WBC	Fresh	17	100%	0%

(*) All sequences WT

The above examples show that the method according to the present invention guarantees the amplification (and the sequencing) even in case of incomplete digestion activity of the restriction endonuclease. The activity of the RE cannot always guarantee the effective digestion of all the RS present in the target DNA, because of the treatment which the cells have been subjected to (as in the previous example), or for other reasons linked to the specific sequence around the restriction site.

Statistically a small percentage of undigested RS is

present in the DRS-WGA, which nevertheless are successfully Whole Genome Amplified, albeit with the universal (primary -PCR) primer being connected to another RS.

5 In case of an undigested site the use of just one PCR for the third target sequence (with the MDRS) would not allow the amplification and the sequencing of said target. In case of incomplete DNA digestion by the restriction enzyme, the method of the invention allows the detection of both WT and M allele when they are present in the DRS-WGA
10 library.

Figure 16 shows an example of the sequencing results of one of the three Veridex-treated lymphocytes positive for the M-PCR. This is the case of the second target sequence (with the MDRS, but undigested during WGA), being
15 amplified and sequenced correctly with the second forward primer.

Example 7 Introduction of a new RS in the mutant allele (PIK3CA gene).

As another example, mutation M1043I, of the exon 21 of
20 the PIK3CA gene stemming from the single nucleotide change ATG/TAAT, can be detected by the method according to the present invention.

From an analysis of the features of the method and kit of the present invention, the resulting advantages are
25 apparent.

In particular, in virtue of the particular design of the primers used to amplify by PCR the library of DNA sequences, the method allows to differentially detect the first target DNA sequence and the second target DNA
30 sequence (differing in the presence of a restriction site for the restriction endonuclease of the DRS-WGA) with great specificity and robustness.

Further, the use of a fourth reverse primer allows an even more specific and robust detection and an amplicon-

size based detection, which is fast, simple and cost-effective.

Further, the method of the present invention may be applied downstream of deterministic restriction site whole
5 genome amplification to detect mutations in a specific and robust manner. These mutations are impossible to otherwise detect with the traditional detection methods available.

Moreover, the use of a DNA sequencing method, in particular Sanger sequencing or pyrosequencing, guarantees
10 the correct detection of even the false positives which could occur in the case of incomplete digestion of the restriction endonuclease of the DNA library.

Furthermore, a percentage of identity from 20% to 80%, better from 40% to 60%, of the third forward primer with
15 the WGA primer allows to obtain an optimal result.

Finally, it is clear that modifications and variants to the method and kit disclosed and shown may be made without because of this departing from the scope of protection of the appended claims.

20 In particular, the method may be multiplexed by using further pairs of primers which do not interfere with the PCR amplification with the first, second, third and possibly fourth primer.

Additionally, one or more of said primers may further
25 include a 5' end sequence which does not hybridize to any of said first or second target sequence positive or antipositive strand. This feature can advantageously be used for one or more of the following purposes:

- barcoding the PCR products with a sample tag,
- 30 - introducing in the PCR product an adaptor for next-generation sequencing
- preventing spurious priming in multiplexed target PCR reaction.

Furthermore, as the WGA products from the PCR reaction

may display some background signal, it may be of advantage to use a different primer for sequencing. This adds an extra layer of specificity, improving the signal-to-noise and readability of the sequence plot.

1484SequenceListing
SEQUENCE LISTING

<110> Silicon Biosystems S.p.A.

<120> METHOD AND KIT FOR DETECTING A WILD-TYPE AND/OR A MUTATED TARGET DNA SEQUENCE

<160> 9

<170> BISSAP 1.0

<210> 1

<211> 21

<212> DNA

<213> Artificial

<220>

<221> source

<222> 1..21

<223> /mol_type="DNA"
/note="universal WGA primer"
/organism="Artificial"

<400> 1

agtgggattc ctgctgtcag t

21

<210> 2

<211> 21

<212> DNA

<213> Homo sapiens

<220>

<221> source

<222> 1..21

<223> /mol_type="DNA"
/note="ALK_WT_F"
/organism="Homo sapiens"

<400> 2

cctctctgct ctgcagcaaa t

21

<210> 3

<211> 22

<212> DNA

<213> Homo sapiens

<220>

<221> source

<222> 1..22

<223> /mol_type="DNA"
/note="ALK_WT_R"
/organism="Homo sapiens"

<400> 3

tctctcggag gaaggacttg ag

22

<210> 4

<211> 20

<212> DNA

<213> Homo sapiens

<220>

<221> source

<222> 1..20

<223> /mol_type="DNA"

1484SequenceListing

/note="ALK_M1_F"
/organism="Homo sapiens"

<400> 4
tgctgtcagt taaaccacca 20

<210> 5
<211> 20
<212> DNA
<213> Homo sapiens

<220>
<221> source
<222> 1..20
<223> /mol_type="DNA"
/note="ALK_M1_R"
/organism="Homo sapiens"

<400> 5
ggtctctcgg aggaaggact 20

<210> 6
<211> 21
<212> DNA
<213> Homo sapiens

<220>
<221> source
<222> 1..21
<223> /mol_type="DNA"
/note="Ex19_M_F"
/organism="Homo sapiens"

<400> 6
taaaattccc gtcgctatca a 21

<210> 7
<211> 22
<212> DNA
<213> Homo sapiens

<220>
<221> source
<222> 1..22
<223> /mol_type="DNA"
/note="Ex19_M_R"
/organism="Homo sapiens"

<400> 7
tgtggagatg agcagggtct ag 22

<210> 8
<211> 27
<212> DNA
<213> Homo sapiens

<220>
<221> source
<222> 1..27
<223> /mol_type="DNA"
/note="Ex19_WT_F"
/organism="Homo sapiens"

<400> 8

ctgtcagtta agagaagcaa catctcc	1484SequenceListing	27
<210> 9 <211> 21 <212> DNA <213> Homo sapiens <220> <221> source <222> 1..21 <223> /mol_type="DNA" /note="Ex19_WT_R" /organism="Homo sapiens" <400> 9 agagcagctg ccagacatga g		
		21

CLAIMS

1. A method for detecting at least one of at least one first target DNA sequence and at least one second target DNA sequence from a library of DNA sequences, wherein the first target DNA sequence differs from the second target DNA sequence in that a single or multiple nucleotide substitution or deletion or insertion in the second sequence generates a restriction site for a restriction endonuclease, comprising the steps of:

(a) providing the library of DNA sequences, each of the DNA sequences comprising, respectively from the 5' end to the 3' end, a first sequence segment having a length from 15 to 50 nucleotides, a second sequence segment of genomic DNA as cleaved by the restriction endonuclease, and a third sequence segment reverse complementary to the union of the first sequence segment and, if any, the 5' overhang generated by the restriction endonuclease;

(b) amplifying the library of DNA sequences by PCR using:

- at least one first reverse primer which hybridises to the 3' end region of the second sequence segment of the at least one first or second target sequence positive strand;

- at least one second forward primer which hybridises to the 3' end region of the second sequence segment of the at least one first target sequence antipositive strand;

- at least one third forward primer comprising a first portion hybridising to the 5' end region of the third sequence segment of the at least second target sequence antipositive strand and a second portion hybridising to the 3' end region of the second sequence segment of the at least one second target sequence antipositive strand, wherein the first portion of the at least one third forward primer has a length from 20% to 80% with respect to the

total length of the at least one third forward primer;

(c) detecting DNA sequences amplified in step (b).

2. The method according to claim 1, wherein step (b) further uses at least one fourth reverse primer which
5 hybridises to the 3' end region of the second sequence segment of the at least one second target sequence positive strand.

3. The method according to claim 1 or 2, wherein the library of DNA sequences is obtained by deterministic
10 restriction site whole genome amplification.

4. The method according to any of claims from 1 to 3, wherein step (c) is performed by a DNA sequencing method.

5. The method according to claim 4, wherein the DNA sequencing method is Sanger sequencing or sequencing by
15 synthesis.

6. The method according to any of the preceding claims, wherein the first portion of the at least one third forward primer has a length from 40% to 60% with respect to the total length of the at least one third forward primer.

20 7. The method according to any of the preceding claims, wherein said second portion of the at least one third forward primer has a length in bases comprised between a minimum corresponding to the consensus sequence of said restriction endonuclease minus, if any, the 5'
25 overhang generated by the restriction endonuclease, all divided by two, and a maximum of 30 bases.

8. The method according to any of the preceding claims, wherein at least one of said primers further comprises a 5' end region which does not hybridize to any
30 of said first or second target sequence, positive or antipositive strand.

9. The method according to any of the preceding claims, wherein the restriction endonuclease is MseI.

10. A kit comprising a first, a second and a third

primer according to any of the preceding claims.

11. A kit for detecting at least one of at least one first target DNA sequence and at least one second target DNA sequence from a library of DNA sequences, wherein the first target DNA sequence differs from the second target DNA sequence in that a single or multiple nucleotide substitution or deletion or insertion in the second sequence generates a restriction site for a restriction endonuclease, and wherein each of the DNA sequences of the library comprises, respectively from the 5' end to the 3' end, a first sequence segment having a length from 15 to 50 nucleotides, a second sequence segment of genomic DNA as cleaved by the restriction endonuclease, and a third sequence segment reverse complementary to the union of the first sequence segment and, if any, the 5' overhang generated by the restriction endonuclease, comprising:

- at least one first reverse primer which hybridises to the 3' end region of the second sequence segment of the at least one first or second target sequence positive strand;

- at least one second forward primer which hybridises to the 3' end region of the second sequence segment of the at least one first target sequence antipositive strand;

- at least one third forward primer comprising a first portion hybridising to the 5' end region of the third sequence segment of the at least second target sequence antipositive strand and a second portion hybridising to the 3' end region of the second sequence segment of the at least one second target sequence antipositive strand, wherein the first portion of the at least one third forward primer has a length from 20% to 80% with respect to the total length of the at least one third forward primer.

12. The kit according to claim 10 or 11 for use in the diagnosis of ALK or EGFR or PIK3CA mutations.