

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 December 2006 (28.12.2006)

PCT

(10) International Publication Number
WO 2006/136170 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/DK2006/000367

(22) International Filing Date: 22 June 2006 (22.06.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PA 2005 00918 22 June 2005 (22.06.2005) DK
PCT/DK2005/000529
17 August 2005 (17.08.2005) DK
PA 2006 00187 9 February 2006 (09.02.2006) DK

(71) Applicant (for all designated States except US): AARHUS
UNIVERSITET [DK/DK]; Nordre Ringgade 1, DK-8000
Aarhus C (DK).

(71) Applicant (for US only): ARBEJDSMLJØINSTITUT-
TET [DK/DK]; Lersø Parkallé 105, DK-2100 Copenhagen
Ø (DK).

(72) Inventors; and

(75) Inventors/Applicants (for US only): NEXØ, Bjørn

Andersen [DK/DK]; Degnebakken 44, DK-8239 Åbyhøj
(DK). VOGEL, Ulla Birgitte [DK/DK]; Islevhusvej
46, DK-2700 Brønshøj (DK). BØRGLUM, Anders
[DK/DK]; Thunøgade 1E, 1.sal, DK-8000 Århus C. (DK).

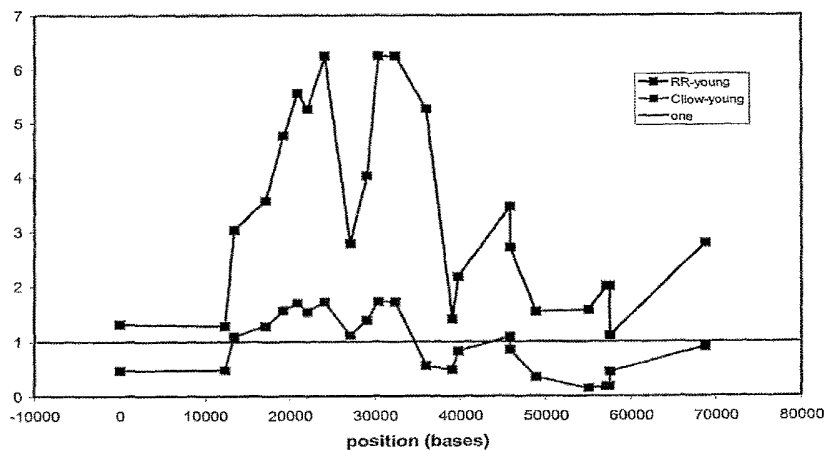
(74) Agent: HØIBERG A/S; St. Kongensgade 59 A, DK-1264
Copenhagen K (DK).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP,
KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT,
LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA,
NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC,
SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT,
RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: DISEASE RISK ESTIMATING METHOD USING SEQUENCE POLYMORPHISMS IN A SPECIFIC REGION OF
CHROMOSOME 19



Relative risk and lower confidence limit for different markers in relation to early breast cancer

(57) Abstract: The present invention provides methods and compositions for identifying human subjects with an increased risk of having or developing cancer. In particular, this invention relates to the identification and characterization of polymorphisms in the human chromosome 19q, the region r located approximately 19q 13.2-3 correlated with increased risk of developing cancer and the responsiveness of a subject to various treatments for cancer. An allele in the r region can be identified as correlated with an increased risk of developing cancer, the prognosis of developed cancer, and responsiveness to cancer treatment, on the basis of statistical analyses of the incidence of a particular allele in individuals diagnosed with cancer. The invention further relates to probes and kits comprising the probes useful in the diagnostic.



Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Disease risk estimating method using sequence polymorphisms in a specific region of chromosome 19

The present invention provides methods and compositions for identifying human subjects with an increased risk of having or developing disease. In particular, this invention relates to the identification and characterization of polymorphisms in the human chromosome 19q, the region r located approximately 19q13.2-3 correlated with increased risk of developing disease, in particular cancer and the responsiveness of a subject to various treatments for cancer.

Background

DNA polymorphisms provide an efficient way to study the association of genes and diseases by analysis of linkage and linkage disequilibrium. With the sequencing of the human genome a myriad of hitherto unknown genetic polymorphisms among people have been detected. Most common among these are the single nucleotide polymorphisms, also called SNPs, of which several millions are known. Other examples are variable number of tandem repeat polymorphisms, insertions, deletions and block modifications. Tandem repeats often have multiple different alleles (variants), whereas the other groups of polymorphisms usually just have two alleles.

Some of these genetic polymorphisms probably play a direct role in the biology of the individuals, including their risk of developing disease, but the virtue of the majority is that they can serve as markers for the surrounding DNA, and thus serve as leads during a search for a causative gene polymorphism, as substitutes in the evaluation of its role in health and disease, and as substitutes in the evaluation of the genetic constitution of individuals.

The association of an allele of one sequence polymorphism with particular alleles of other sequence polymorphisms in the surrounding DNA has two origins, known in the genetic field as linkage and linkage disequilibrium, respectively. Linkage arises because large parts of chromosomes are passed unchanged from parents to offspring, so that minor regions of a chromosome tend to flow unchanged from one generation to the next and also to be similar in different branches of the same family. Linkage is gradually eroded by recombination occurring in the cells of the germline, but typically operates over multiple generations and distances of a number of million bases in the DNA.

Linkage disequilibrium deals with whole populations and has its origin in the (distant) forefather in whose DNA a new sequence polymorphism arose. The immediate surroundings in the DNA of the forefather will tend to stay with the new allele for many generations. Recombination and changes in the composition of the population will again erode the association, but the new allele and the alleles of any other polymorphism nearby will often be partly associated among unrelated humans even today. A crude estimate suggests that alleles of sequence polymorphisms with distances less than 10000 bases in the DNA will have tended to stay together since modern man arose. Linkage disequilibrium in limited populations, for instance Europeans, often extends over longer distances. This can be the result of newer mutations, but can also be a consequence of one or more "bottlenecks" with small population sizes and considerable inbreeding in the history of the current population. Two obvious possibilities for "bottlenecks" in Europeans are the exodus from Africa and the repopulation of Europe after the last ice age.

Linkage disequilibrium is the results of many stochastic events and as such subject to statistical variation occasionally resulting in discontinuities, lack of a monotonic relationship between association and distance and differences between people of different ethnicity. Therefore, it is often advantageous to study more than one sequence polymorphism in a given region. This also allows for further definition of the genetic surroundings of the biologically relevant polymorphism by combining the associated alleles of the different markers into a so-called haplotype.

Humans in general carry two copies of each human chromosome in each cell. There are exceptions to this rule, not relevant to this application. We therefore speak about genotypes i.e. the combined analysis of both chromosomes at a given sequence polymorphism. The resulting genotypes of a person, analysed for instance on DNA from peripheral blood leukocytes, are inherently very stable over time. Therefore, this type of analysis can be performed any time in the life of a person and will be applicable to this person for his or her entire life. By the same token such genetic analyses are ideally suited to predict future risks of disease.

A variety of investigations suggest that many diseases in part are determined by the genetic constitution of the individual. One group of genes in particular has been as-

sociated with rare genetic predispositions to cancer. These are the genes involved in maintaining the integrity of a person's DNA, the so-called DNA repair genes. One set of such genes are the XP genes which participate in nucleotide excision repair, and, when mutated, give rise to a 1000 fold increased risk of getting skin cancer. For this reason we have previously investigated single nucleotide polymorphisms in one DNA repair gene XPD for association with risk of skin cancer in a cohort of Caucasian Americans, and found that one allele of the sequence polymorphism called XPDe6 was associated with a moderately increased risk of getting basal cell carcinoma, the most common form of skin cancer. Later other groups have studied the association between sequence polymorphisms in this and other DNA repair genes and various forms of cancer. Some have reported positive results.

Very little is known about the function of the gene RAI. It was cloned because its protein product binds to and inhibits RelA of the transcription regulator NF-kappaB. Other studies suggest that it may interact with the product p53 of a tumor suppressor gene.

Summary of the invention

The present invention relates in a first aspect to a group of nucleic acid sequences found to be associated with disease, in particular cancer. The invention further relates to transcriptional and translational products of said sequence. An allele in the region can be identified as correlated with an increased risk of developing disease, in particular cancer, the prognosis of developed disease, in particular cancer, and responsiveness to disease treatment, in particular cancer treatment on the basis of statistical analyses of the incidence of a particular allele in individuals diagnosed with disease, in particular cancer.

Thus, in a first aspect the invention relates to a method for estimating the disease risk of an individual comprising

A method for estimating the disease risk of an individual comprising

- in a sample from said individual assessing in the genetic material at least one sequence polymorphism
- in a region corresponding to SEQ ID NO: 132, or a part thereof, or
- in a region complementary to SEQ ID NO: 132, or a part thereof, or

- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 132, or a part thereof, or
- or translation product from a sequence in a region corresponding to SEQ ID NO: 132, or a part thereof, and optionally a further sequence polymorphism
- 5 - in a region corresponding to SEQ ID NO: 133, or a part thereof, or
- in a region complementary to SEQ ID NO: 133, or a part thereof, or
- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof, or
- 10 - or translation product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof,

- with the proviso that when using SEQ ID NO: 132 only the at least one sequence polymorphism is in a region corresponding to at least one region within SEQ ID NO: 132 selected from the regions flanked by and including the sequence polymorphism RAI3'd1 (SEQ ID NO: 19) and rs2377329 (SEQ ID NO: 22), or

In a region flanked by and including the sequence polymorphism positioned at 18149120 in Contig NT_011109 (SEQ ID NO: 3) and sequence polymorphism positioned at 18149154 in Contig NT_011109 (SEQ ID NO: 4), or

the sequence polymorphism positioned at 18150815 in Contig NT_011109 (SEQ ID NO: 5), or

the sequence polymorphism positioned at 18151158 in Contig NT_011109 (SEQ ID NO: 6)

- obtaining a sequence polymorphism response,
- 30 - estimating the disease risk of said individual based on the sequence polymorphism response.

The estimation of the disease risk of an individual can involve the comparison of the number and/or kind of polymorphic sequences identified with a predetermined disease risk profile. Such a profile can be based on statistical data obtained for a rele-

vant reference group of individuals. In particular the disease is a proliferative disease, such as cancer.

5 The sequence of the r region is set forth as SEQ ID NO: 1, originating from the cloning of human chromosome 19q published as part of the contig NT_011109 in the database of human sequences established by National Center for Biotechnology Information and located on the internet at <http://www.ncbi.nlm.nih.gov/genome/guide/human/> .

10 The presence of an allele is determined by determining the nucleic acid sequence of all or part of the region according to standard molecular biology protocols well known in the art as described for example in Sambrook et al. (1989) and as set forth in the Examples provided herein or products of the nucleic acid sequences.

15 In particular, the nucleic acid molecules of the present invention represent in a first aspect nucleic acid sequences forming part of the region r corresponding to SEQ ID NO: 132 and optionally a further region corresponding to SEQ ID NO: 133. The nucleic acid sequences of the present invention represent within the gene or controlling the gene referred to herein as RAI. As demonstrated in the Examples presented below, the RAI gene is in particular associated with human cancer diseases.

20

Furthermore, the invention relates to a method for estimating the disease prognosis of an individual comprising

A method for estimating the disease prognosis of an individual comprising

- 25
- in a sample from said individual assessing in the genetic material at least one sequence polymorphism
 - in a region corresponding to SEQ ID NO: 132, or a part thereof, or
 - in a region complementary to SEQ ID NO: 132, or a part thereof, or
 - in a transcription product from a sequence in a region corresponding to SEQ
 - 30 ID NO:132, or a part thereof, or
 - or translation product from a sequence in a region corresponding to SEQ ID NO: 132, or a part thereof, and optionally a further sequence polymorphism
 - in a region corresponding to SEQ ID NO: 133, or a part thereof, or
 - in a region complementary to SEQ ID NO: 133, or a part thereof, or

- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof, or
- or translation product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof,

5

- with the proviso that when using SEQ ID NO: 132 only the at least one sequence polymorphism is in a region corresponding to at least one region within SEQ ID NO: 132 selected from the regions flanked by and including the sequence polymorphism RAI3'd1 (SEQ ID NO: 19) and rs2377329 (SEQ ID NO: 22), or

10

in a region flanked by and including the sequence polymorphism positioned at 18149120 in Contig NT_011109 (SEQ ID NO: 3) and sequence polymorphism positioned at 18149154 in Contig NT_011109 (SEQ ID NO: 4), or

15

the sequence polymorphism positioned at 18150815 in Contig NT_011109 (SEQ ID NO: 5), or

the sequence polymorphism positioned at 18151158 in Contig NT_011109 (SEQ ID NO: 6)

20

- obtaining a sequence polymorphism response,

- estimating the disease prognosis of said individual based on the sequence polymorphism response.

25

The estimation of the disease prognosis of an individual can involve the comparison of the number and/or kind of polymorphic sequences identified with a predetermined disease prognosis profile. Such a profile can be based on statistical data obtained for a relevant reference group of individuals.

30

Additionally provided is a method of identifying a human subject as having an increased likelihood of responding to a treatment, comprising a) correlating the presence of an r region allele genotype with an increased likelihood of responding to treatment; and b) determining the r region allele genotype of the subject, whereby a subject having an r region allele genotype correlated with an increased likelihood of

35

responding to treatment is identified as having an increased likelihood of responding to treatment.

Thus, the present invention also relates to method for estimating a treatment re-

5 sponse of an individual suffering from disease to a disease treatment, comprising

A method for estimating a treatment response of an individual suffering from cancer to a disease treatment, comprising

in a sample from said individual assessing in the genetic material at least one sequence polymorphism

10 in a region corresponding to SEQ ID NO: 132, or a part thereof, or

- in a region complementary to SEQ ID NO: 132, or a part thereof, or
- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 132, or a part thereof, or
- or translation product from a sequence in a region corresponding to SEQ ID

15 NO:132, or a part thereof, and optionally a further sequence polymorphism in a region corresponding to SEQ ID NO: 133, or a part thereof, or

- in a region complementary to SEQ ID NO: 133, or a part thereof, or
- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof, or
- 20 - or translation product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof,

- with the proviso that when using SEQ ID NO: 132 only the at least one sequence polymorphism is in a region corresponding to at least one region
- 25 within SEQ ID NO: 132 selected from the regions flanked by and including the sequence polymorphism RAI3'd1 (SEQ ID NO:19) and rs2377329 (SEQ ID NO: 22), or

in a region flanked by and including the sequence polymorphism positioned

30 at 18149120 in Contig NT_011109 (SEQ ID NO: 3) and sequence polymorphism positioned at 18149154 in Contig NT_011109 (SEQ ID NO: 4), or

the sequence polymorphism positioned at 18150815 in Contig NT_011109 (SEQ ID NO: 5), or

the sequence polymorphism positioned at 18151158 in Contig NT_011109
(SEQ ID NO: 6)

- obtaining a sequence polymorphism response,
- estimating the individual's response to the disease treatment based on the sequence polymorphism response.

The estimation of the individual's response to disease treatment can involve the comparison of the number and/or kind of polymorphic sequences identified with a predetermined cancer treatment response profile. Such a profile can be based on statistical data obtained for a relevant reference group of individuals. In particular the disease is a proliferative disease, such as cancer.

The invention also comprises primers or probes for use in the invention, as well as kits including these. The primers and/or probes are preferably capable of hybridising to SEQ ID NO:132 and optionally to SEQ ID NO: 133, or a part thereof, in particular the regions relevant to this invention, or a part thereof, under stringent conditions, as well as to a sequence complementary thereto.

In particular, the present invention is based on the discovery of the correlation with single nucleotide polymorphisms (SNPs), deletion polymorphisms, insertion polymorphisms, dinucleotide polymorphisms and/or tandem repeats in the regions and disease. Thus, polymorphisms have been found in the r region as shown in table 1. However, the present invention is not limited to the polymorphisms shown in table 1, but does include any polymorphism in the region. Deletions and insertions have been found in the r region as shown herein. However, the present invention is not limited to the deletions shown in the tables, but does include any deletion and insertion in the region.

The term human includes both a human having or suspected of having a disease and a a-symptomatic human who may be tested for predisposition or susceptibility to disease. At each position the human may be homozygous for an allele or the human may be a heterozygote.

Drawings

Fig. 1 shows an overview of a subregion of chromosome 19q in which the exon-intron structure of RAI, XPD and ASE-1 is shown. Letters correspond to the specific marker position. The position in NT_011109 is shown. Marker abbreviations are as follows: a=XPDexon10, b=XPD exon6, c=XPD-4bp, d=XPD-81bp, e=XPD-5'2, f=XPD5'3, g=RAI-3'7, h=RAI-3'3, i=RAI-3'6, j=RAI-3'5, k=RAI-3'4, l=RAI-3'8, m=RAI exon6, n=RAI intron5-2, o=RAI intron3, p=RAI-intron1, q=RAI intron1-2, r=RAI intron1-3, s=RAI-5'2, t=RAI-5'3, u=RAI-5', v=ASE1-5'2, w= ASE1 exon1, x=ERCC1-3'

Fig. 2 shows a pair-wise linkage disequilibrium between all markers in breast cancer controls.

Fig. 3 shows an association of single polymorphisms with early breast cancer.

Fig. 4 shows an association of sets of neighboring SNPs with breast cancer.

Fig. 5 shows a maximal OR for breast cancer for haplotypes formed by two SNPs.

Fig. 6 shows a Lazzeroni estimation of the position of the causative variant using young breast cancer.

Fig. 7 shows the overall distribution of p-values for sets of markers plotted against the position on the chromosome for basal cell carcinoma and lung cancer. As position on the abscissa we used the median marker position in a given cluster of markers. The ordinate values are the negative logarithms to the overall p-values for a difference between cases and controls associated with a given set of markers. Each curve corresponds to a given size of marker sets, i.e. the length of the haplotypes.

Fig. 8 shows odds ratio for cancer versus control between the two homozygotes of each SNP in relation to location on chromosome 19.

Fig. 9 shows event free survival for patients who are wild-type carriers of ASE-1 (0) or homozygous or heterozygous carriers of the variant allele (1).

Fig. 10 shows overall survival subdivided by ASE-1 genotype. 0=homozygous carrier of the wild-type allele of ASE-1, 1= carrier of the variant allele of ASE-1.

Fig. 11 shows event free survival for women subdivided by ASE-1 genotype.

5 0=homozygous carrier of the wild-type allele of ASE-1, 1= carrier of the variant allele of ASE-1.

Fig. 12 shows overall survival for women subdivided by ASE-1 genotype.

10 0=homozygous carrier of the wild-type allele of ASE-1, 1= carrier of the variant allele of ASE-1

Fig. 13 shows event free survival for men subdivided by ASE-1 genotype.

15 0=homozygous carrier of the wild-type allele of ASE-1, 1= carrier of the variant allele of ASE-1.

Fig. 14 shows overall survival for men subdivided by ASE-1 genotype.

0=homozygous carrier of the wild-type allele of ASE-1, 1= carrier of the variant allele of ASE-1.

20 Fig. 15 shows Kaplan Meier plot of survival of lung cancer patients in relation to highrisk haplotype

Fig. 16 shows Kaplan Meier plot of survival in relation to XPD K751Q among those lung cancer patients homozygous for the high risk haplotype.

25 Fig. 17 shows Kaplan Meier plot of survival in relation to XPD K751Q among those lung cancer patients not homozygous for the highrisk haplotype.

Detailed description of the invention

30 The present invention relates to a characterization of a person's present and/or future risk of getting certain forms of disease, in particular a proliferative disease, such as cancer. The characterization is based on the analysis of sequence polymorphisms in a region of chromosome 19q in the person.

A number of polymorphisms in the chromosomal region 19q13.2-3 have been identified and characterised. Surprisingly, the sequence polymorphisms with strongest association to disease appeared to be located outside the gene XPD. More specifically, the sequences were located in a sub-region harboring the gene RAI and immediately downstream of the RAI gene. The nature of the association between haplotype and disease was examined together with the p-values associated with the individual haplotypes. The odds ratios for each haplotype of each set of three neighboring SNPs were also determined. The odds ratio for test of homozygotes of individual markers against cancer status was likewise determined. The most likely location of a single causative gene variant for individuals of the breast cancer cases was evaluated.

The region of chromosome 19q, more precisely the region located in 19q13.2-3, with which the present invention is concerned, is depicted in Fig. 1 as it is presently known together with the presently known or suspected genes. The arrows indicate the directions of transcription of the genes. The absolute chromosome positions shown are from the particular build of NCBI's map of chromosome 19, and will probably change with time. The position of markers used throughout the experiments is indicated. The intron-exon structure of the RAI gene is shown together with the position of the inter gene region between RAI and XPD genes.

The region *r* stretches from the beginning of, but not including the XPD gene, to approximately the end of ERCC1 and includes the genes RAI, LOC162978, and ASE-1. More specifically *r* is bounded by and includes the following two sequences:
AGAACCCCCG CCCCTCCACC TCGTCTCAA and TCCCTCCCCA GA-
GACTGCAC CAGCGCAGCC, and is defined by SEQ ID NO: 1.

In the present context the region *r* means SEQ ID NO: 1 and complementary sequence as well as transcriptional products and translational products thereof.

In one preferred embodiment, the gene RAI is defined in the claims as including transcribed sequences of the gene plus a 1500 base upstream promoter region. More specifically RAI is bounded by and includes the following sequences:
CATAACCACA ATGATGAGCA TGTATTGAGT and ATGTTGTCCA GGCTGGTCTT
GAACTCCTGA. In the present context this section of the region relates to SEQ ID

NO: 1 bases 7761-22885 and complementary sequence as well as transcriptional products and translational products thereof.

5 In another embodiment, one preferred section of the region stretches approximately from the the beginning of, but not including the XPD gene, to approximately the end of the RAI gene. In the present context the region means SEQ ID NO: 1 bases 1 to 25550 and complementary sequence as well as transcriptional products and translational products thereof.

10 In an even more preferred embodiment, one preferred section of the region stretches approximately from the the beginning of, but not including the XPD gene, to approximately within the RAI gene. In the present context the region means SEQ ID NO: 1 bases 1 to 15698 and complementary sequence as well as transcriptional products and translational products thereof.

15 In a most preferred embodiment, one preferred section of the region r stretches approximately from outside the XPD gene, to approximately within the RAI gene. In the present context the region means SEQ ID NO: 1 bases 4528 to 15698 and complementary sequence as well as transcriptional products and translational products thereof.

20 In a third embodiment, one preferred section of the region r stretches approximately from the beginning of, but not including the XPD gene, into the inter gene region between the RAI and XPD gene. In the present context the region means SEQ ID NO: 1 bases 1 to 1510 and complementary sequence as well as transcriptional products and translational products thereof.

25 In a further embodiment, one preferred section of the region r stretches approximately from the the beginning of, but not including the XPD gene, throughout the inter gene region between the RAI and XPD gene and into the 3' part of the RAI gene. In the present context the region means SEQ ID NO: 1 bases 1710 to 8685 and complementary sequence as well as transcriptional products and translational products thereof.

30

In yet a further embodiment, one preferred section of the region r stretches approximately over the central part of the RAI gene. In the present context the region means SEQ ID NO: 1 bases 8987 to 12090 and complementary sequence as well as transcriptional products and translational products thereof.

5

In yet another embodiment, one preferred section of the region r stretches approximately over the middle and 5' part of the RAI gene. In the present context the region means SEQ ID NO: 1 bases 15898 to 25550 and complementary sequence as well as transcriptional products and translational products thereof.

10

In a further embodiment, one preferred section of the region r stretches over the regions corresponding to SEQ ID NO: 132 and SEQ ID NO: 133 as described in detail elsewhere herein-

15

Modifications to the human genome map are known to occur from time to time. It is therefore possible that the defining sequences quoted above will change slightly in future maps.

20

Fragments or parts of the region r as used herein relates to any fragment of at least 5 nucleic acid residues in length, or multiples of 5 nucleic acid residues in length starting from SEQ ID NO: 1 position 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100. For example at least 21, such as at least 22, for example at least 23, such as at least 24, for example at least 26, such as at least 27, for example at least 28, such as at least 29, for example at least 31, such as at least 32, for example at least 33, such as at least 34, for example at least 36, such as at least 37, for example at least 38, such as at least 39, for example at least 41, such as at least 42, for example at least 43, such as at least 44, for example at least 46, such as at least 47, for example at least 48, or at least 100 nucleic acid residues in length, or multiples of 100 nucleic acid residues in length, starting from SEQ ID NO: 1 position 1, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, and so forth, each fragment starting position having an increment of 100 nucleic acid residues. Multiples are preferably multiples of e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26,

30

27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 and 50.

For fragments starting at position 1, the length of said fragments will thus be e.g.

5 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2600, 2700, 2800, 2900, 3000, and so forth, using suitable multipliers as listed herein above.

For fragments starting at position 100, the length of said fragments will thus be e.g.

10 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2600, 2700, 2800, 2900, 3000, and so forth, using suitable multipliers as listed herein above.

For fragments starting at position 7700, the length of said fragments will thus be e.g.

15 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2600, 2700, 2800, 2900, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 10500, 11000, 11500, 12000, 12500, 13000, 13500, 14000, 14500, 15000, and so forth, using suitable multipliers such as e.g. the
20 ones listed herein above.

The nucleic acid sequences according to the present invention make it possible to estimate cancer risk in an individual by using sequence polymorphisms originating from a specific region of chromosome 19.

25

Estimation of disease risks has a number of important applications, which in the following is exemplified with respect to cancer, but also apply to other disease, as described herein:

30 **(1)** Individuals with reasons to suspect that they are at risk for getting cancer would be able to clarify their situation and, if possible, take protective action. Alternatively, anti-cancer campaigns, companies, hospitals or other institutions could offer a service to help people clarify their situation. It would for instance be possible to test persons, when they got their first basal cell carcinoma, which is often recurrent and
35 also is a moderate predictor for other cancers. If the persons were in a high-risk

group, one could then advice them about, or they could of their own accord choose, risk-reducing behaviour, such as avoidance of excessive sun-exposure, abstaining from smoking etc. About 5 percent of the Danish population will at some point in their life get a basal cell carcinoma.

5

(2) Anti-cancer campaigns, companies, hospitals or other institutions would be able to define relevant target subpopulations and focus information on risk-reducing behaviour on these persons. They might perhaps also be in a position to inform the remainder of the population that they need not worry. Lung cancer affects approximately 10-15 percent of smokers and thus approximately 5 percent of the population, somewhat varying from country to country. Malignant melanoma, a sun-induced, often lethal form of skin cancer, affects approximately 700 persons a year in Denmark or about 1 percent of the Danish population.

10

(3) The drugs used in cancer treatment are often carcinogenic themselves and individual responses to them vary considerably, both with respect to tolerance to the treatment and with respect to efficacy of the treatment. It is an obvious possibility that the region of chromosome 19 here dealt with, which contains DNA repair genes known to modulate carcinogen responses, also modulates response to anti-cancer agents. Hence, analysis of the region may facilitate better choices of treatment for cancer, and/or help predict the future course of disease.

15

20

25

By sequence polymorphism is understood any single nucleotide, tandem repeat, insertion, deletion or block polymorphism, which varies among humans, whether it is of known biological importance or not.

Position of sequence polymorphism in the region r

30

In one embodiment of the methods of the invention, preferably the method for diagnosis as described herein, one or more single nucleotide polymorphism(s) at a predetermined position in the region (SEQ ID NO:1) are identified and used for e.g. cancer risk profiling and/or cancer treatment response profiling. Presently preferred polymorphism(s) are listed in Tables 1a, 1b and 1c, more preferably at least two polymorphism(s) are selected, most preferably at least three polymorphism(s) are selected. However, the present invention relates to any polymorphism in the region.

35

Table 1a

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative Position	Position in SEQ ID NO:1	SEQ ID NO:
XPD-81bp	81bp/-	rs3916787	NT_011109	18143027	19890	632	33
XPD-5'2	A/G	rs2097215	NT_011109	18144005	20868	1610	34
XPD-5'3	C/T	rs11878644	NT_011109	18145185	22048	2790	35
RAI-3'7	C/T	rs7252567	NT_011109	18146823	23686	4428	21
RAI-3'3	AA/-	rs3047560	NT_011109	18147192	24055	4797	23
RAI-3'6	A/C	10422489	NT_011109	18147886	24749	5491	24
RAI-3'5	C/T	10426701	NT_011109	18148193	25056	5798	25
RAI-3'4	A/G	rs4544343	NT_011109	18150199	27062	7804	26
RAI-3'8	A/G	rs8101662	NT_011109	18150911	27774	8516	27
RAI exon 6	A/T	rs6966	NT_011109	18151180	28043	8785	36
RAI-i5-2	A/T	rs8112723	NT_011109	18153497	30360	10204	37
RAI intron 3	A/G	rs2017104	NT_011109	18155483	32346	12190	38
RAI intron 1	A/G	rs1970764	NT_011109	18159091	35954	15798	28
RAI intron 1-2	A/G		AC092309	24351	39069	18913	45
RAI intron 1-3	A/C		AC092309	25115	39833	19677	46
RAI-5'2	C/T	rs4803814	NT_011109	18168944	45807	25650	39
RAI-5'3	C/T	rs4803815	NT_011109	18168948	45811	25564	40
RAI-5'	C/T	rs4572514	NT_011109	18171984	48847	28691	41
ASE1-5'2	G/T	rs2226949	NT_011109	18175327	52190	32035	42
ASE1 exon 1	A/G	rs967591	NT_011109	18178152	55015	34854	43
ERCC1-3'	C/T	rs762562	NT_011109	18180561	57424	37267	44

rs numbers were derived from the NCBI's database dbSNP.

Table 1b

dbSNP #		Position in SEQ ID NO: 1	SEQ ID NO:
rs3916787	gctgcagtg gctgt (-/ACACCTGTGGTCCCAGCTACTCTGG AAGCTGAGGTGGGAGGATCGCTTGAGCCCAAGAGGTG-GAGGCTGCAGTGAGCTGT) gactgtgcc ctgcactcca	632	33
rs2097215	TGACAGTAGA CATCCTGTCA T (A/G) ATAAGTCttt tttttt	1610	34
rs11878644	CATCCCCATA CCAAcccacc (c/t) tactgctctg atctcctct	2790	35
rs7252567	ttttagtagagacatggttcgcc (C/T) gttgccaggctggtctgaactc	4428	21
rs3047560	ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG-CATGGTGGTGGGGTGC	4797	23
rs10422489	AAAAAAC-TAAAGTGGGGTTTGCGGG(G/T)AGTGGGAGGGCCCTTC CTGCTAGG	5491	24
rs10426701	ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac	5798	25
rs4544343	TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTTTGAG	7804	26
rs8101662	CTCGGGAGGCTGAGGCAGGAGAATC (A/G)CTTGAACCTCAGGCAGAGGTTG	8516	27
rs6966	ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA	8785	36
rs8112723	ttgggagacc aaggcaggtg gatc (a/t) ttgaggtca gtagatcaaa	10204	37
rs2017104	gggaggctcg aggcgggc (A/G) gattgcatga gctcaggatt	12190	38
rs1970764	tgcagtgagc tgagatcgc (A/G) ccactgcact ccagcctggg	15798	28
AC092309	CTTGCTACAGAATTACAGGCA (T/C) GCGCCACCGCTCCGGGCTAA	18913	45
AC092309	CTAAAGACTACA (-/A) TTTCCAGCATCCCATTG	19677	46

rs4803814	CCCTCCCTGCTTGCTTGCTTTCTCT[C/T]TCTCTCTTTCTT TCTTTCTTTCTT	25650	39
rs4803815	CCCTGCTTGCTTGCTTTCTCTCTCT[C/T]TCTTTCTTTCTT TCTTTCTTTCTT	25564	40
rs4572514	AGAACCTGTTACAGGCTGGCGGCTCA[C/T]TTGGATGAAC AGGGAGTGTGTGAC	28691	41
rs2226949	CCCCCTTCTTAGGACG- CATGGGGGT[G/T]GAGAGAACGGGGAGATAGACAGAG	32035	42
rs967591	TGCGAGCAGCCCGGGCTA- CAGGGTT[A/G]CCTGAGGTGTGGGTCCCAGGATGG	34854	43
rs762562	GGCGCCTCAACAGCCAGAAG- GAGCG[A/G]AGCCTCAGGCCAGGCAGCTCTGG	37267	44

Table 1c

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative position	SEQ ID NO:
<i>XPD</i> exon 23	A/C	Rs1052559	NT_011109	18123137	0	
<i>XPD</i> exon 10	A/G	Rs1799793	NT_011109	18135477	12340	
<i>XPD</i> exon 6	A/C	Rs238406	NT_011109	18136527	13390	
<i>XPD</i> intron 3	A/G	Rs1799783	NT_011109	18140254	17117	
<i>XPD</i> -4bp	GACA/-	Rs3916791	NT_011109	18142345	19208	
<i>XPD</i> -81bp	81bp/-	Rs3916787	NT_011109	18143027	19890	33
<i>XPD</i> -5'2	A/G	Rs2097215	NT_011109	18144005	20868	34
<i>XPD</i> -5'3	C/T	Rs11878644	NT_011109	18145185	22048	35
<i>RAI</i> -3'7	C/T	Rs7252567	NT_011109	18146823	23686	21
<i>RAI</i> -3'3	AA/-	Rs3047560	NT_011109	18147192	24055	23
<i>RAI</i> -3'4	A/G	Rs4544343	NT_011109	18150199	27062	26
<i>RAI</i> exon 6	A/T	Rs6966	NT_011109	18151180	28043	36
<i>RAI</i> intron5-3	C/T	Rs10417235	NT_011109	18152171	29034	
<i>RAI</i> intron5-2	A/T	Rs8112723	NT_011109	18153497	30360	37
<i>RAI</i> intron 3	A/G	Rs2017104	NT_011109	18155483	32346	
<i>RAI</i> intron 1	A/G	Rs1970764	NT_011109	18159091	35954	28
<i>RAI</i> intron 1-2	C/T	Rs6509210	NT_011109	18162206	39069	45
<i>RAI</i> intron1-3	A/C		NT_011109	18162970	39833	46
<i>RAI</i> -5'2	C/T	Rs4803814	NT_011109	18168944	45807	39
<i>RAI</i> -5'3	C/T	Rs4803815	NT_011109	18168948	45811	40
<i>RAI</i> -5'	C/T	Rs4572514	NT_011109	18171984	48847	41
<i>ASE1</i> -5'2	G/T	Rs2226949	NT_011109	18175327	52190	42
<i>ASE1</i> exon 1	A/G	Rs967591	NT_011109	18178152	55015	43
<i>ERCC1</i> -3'2	A/G	Rs735482	NT_011109	18180220	57083	44
<i>ERCC1</i> -3'	C/T	Rs762562	NT_011109	18180561	57424	
<i>ERCC1</i> -3'3	A/G	Rs2336219	NT_011109	18180624	57487	
<i>ERCC1</i> exon 4	C/T	Rs3177700	NT_011109	18191871	68734	

In one embodiment of the methods of the invention, preferably the method for diagnosis as described herein, one or more single nucleotide polymorphism(s) at a predetermined position in the region (SEQ ID NO:1) are identified and used for e.g. cancer risk profiling and/or cancer treatment response profiling. Presently preferred polymorphism(s) are listed in Tables 2a and 2b, more preferably at least two polymorphism(s) are selected, most preferably at least three polymorphism(s) are selected. However, the present invention relates to any polymorphism in the region.

Table 2a

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative Position	Position in SEQ ID NO:1	SEQ ID NO:
RAI-3'4	A/G	rs4544343	NT_011109	18150199	27062	7804	26
RAI-3'8	A/G	rs8101662	NT_011109	18150911	27774	8516	27
RAI exon 6	A/T	rs6966	NT_011109	18151180	28043	8785	36
RAI-i5-2	A/T	rs8112723	NT_011109	18153497	30360	10204	37
RAI intron 3	A/G	rs2017104	NT_011109	18155483	32346	12190	38
RAI intron 1	A/G	rs1970764	NT_011109	18159091	35954	15798	28
RAI intron 1-2	A/G		AC092309	24351	39069	18913	45
RAI intron 1-3	A/C		AC092309	25115	39833	19677	46

rs numbers were derived from the NCBI's database dbSNP.

5 **Table 2b**

dbSNP #		Position in SEQ ID NO: 1	SEQ ID NO:
rs4544343	TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTGTTGAG	7804	26
rs8101662	CTCGGGAGGCTGAGGCAGGAGAATC (A/G)CTTGAATCAGGCAGAGGTTG	8516	27
rs6966	ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA	8785	36
rs8112723	ttgggagacc aaggcaggtg gatc (a/t) ttgaggtca gtagatcaaa	10204	37
rs2017104	gggaggctcg aggcgggc (A/G) gattgcatga gtcaggatt	12190	38
rs1970764	tgcagtgagc tgagatcgc (A/G) ccactgcact ccagcctggg	15798	28
AC092309	CTTGCTACAGAATTACAGGCA (T/C) GCGCCACCGCTCCGGGCTAA	18913	45
AC092309	CTAAAGACTACA (-/A) TTTCCCAGCATCCCATTG	19677	46

In one embodiment of the methods of the invention, preferably the method for diagnosis as described herein, one or more single nucleotide polymorphism(s) at a pre-determined position in the region (SEQ ID NO: 1) are identified and used for e.g. cancer risk profiling and/or cancer treatment response profiling. Presently preferred polymorphism(s) are listed in Tables 3a, 3b and 3c, more preferably at least two polymorphism(s) are selected, most preferably at least three polymorphism(s) are selected. However, the present invention relates to any polymorphism in the region.

15 **Table 3a**

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative Position	Position in SEQ ID NO:1	SEQ ID NO:
XPD-81bp	81bp/-	Rs3916787	NT_011109	18143027	19890	632	33
XPD-5'2	A/G	Rs2097215	NT_011109	18144005	20868	1610	34
XPD-5'3	C/T	Rs1187864	NT_011109	18145185	22048	2790	35

RAI-3'7	C/T	Rs7252567	NT_011109	18146823	23686	4428	21
RAI-3'3	AA/-	Rs3047560	NT_011109	18147192	24055	4797	23
RAI-3'6	A/C	10422489	NT_011109	18147886	24749	5491	24
RAI-3'5	C/T	10426701	NT_011109	18148193	25056	5798	25
RAI-3'4	A/G	Rs4544343	NT_011109	18150199	27062	7804	26
RAI-3'8	A/G	Rs8101662	NT_011109	18150911	27774	8516	27
RAI exon 6	A/T	Rs6966	NT_011109	18151180	28043	8785	36
RAI-i5-2	A/T	Rs8112723	NT_011109	18153497	30360	10204	37
RAI intron 3	A/G	Rs2017104	NT_011109	18155483	32346	12190	38
RAI intron 1	A/G	Rs1970764	NT_011109	18159091	35954	15798	28
RAI intron 1-2	A/G		AC092309	24351	39069	18913	45
RAI intron 1-3	A/C		AC092309	25115	39833	19677	46

rs numbers were derived from the NCBI's database dbSNP.

Table 3b

dbSNP #		Position in SEQ ID NO: 1	SEQ ID NO:
rs3916787	gctgcagtga gctgt (-/ACACCTGTGGTCCCAGCTACTCTGG AAGCTGAGGTGGGAGGATCGCTTGAGCCCAAGAGGTG- GAGGCTGCAGTGAGCTGT) gactgtgccactgactcca	632	33
rs2097215	TGACAGTAGA CATCCTGTCA T (A/G) ATAAGTCttt tttttt	1610	34
rs11878644	CATCCCCATA CCAAcccacc (c/t) tactgctctg atctctcct	2790	35
rs7252567	tttagtagagacatggttcgccca (C/T) gttgccaggtggtctgaactc	4428	21
rs3047560	ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG- CATGGTGGTGGGGTGC	4797	23
rs10422489	AAAAAAC- TAAAGTGGGTTTTCGGG(G/T)AGTGGGAGGGCCCTTC CTGCTAGG	5491	24
rs10426701	ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac	5798	25
rs4544343	TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTGTTGAG	7804	26
rs8101662	CTCGGGAGGCTGAGGCAGGAGAATC (A/G)CTTGAACCTCAGGCAGAGGTTG	8516	27
rs6966	ATTAAGTGCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA	8785	36
rs8112723	ttgggagacc aaggcaggtg gatc (a/t) ttgaggatca gtatgcaaa	10204	37
rs2017104	gggaggctcg aggcgggc (A/G) gattgcatga gtcaggatt	12190	38
rs1970764	tgcagtgagc tgagatgcg (A/G) ccactgcact ccagcctggg	15798	28
AC092309	CTTGCTACAGAATTACAGGCA (T/C) GCGCCACCGCTCCGGGCTAA	18913	45
AC092309	CTAAAGACTACA (-/A) TTTCCAGCATCCCATTG	19677	46

5

More preferably polymorphism(s) are listed below in tables 3c and 3d, more preferably at least two polymorphism(s) are selected, most preferably at least three polymorphism(s) are selected from SEQ ID NO:1 below

10

Table 3c

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative Position	Position in SEQ ID NO:1	SEQ ID NO:
--------------	------	---------	----------	----------------------------	----------------------	-------------------------------	---------------

XPD-81bp	81bp/-	Rs3916787	NT_011109	18143027	19890	632	33
XPD-5'2	A/G	Rs2097215	NT_011109	18144005	20868	1610	34
XPD-5'3	C/T	Rs1187864	NT_011109	18145185	22048	2790	35
		4					
RAI-3'7	C/T	rs7252567	NT_011109	18146823	23686	4428	21
RAI-3'3	AA/-	rs3047560	NT_011109	18147192	24055	4797	23
RAI-3'6	A/C	10422489	NT_011109	18147886	24749	5491	24
RAI-3'5	C/T	10426701	NT_011109	18148193	25056	5798	25
RAI-3'4	A/G	rs4544343	NT_011109	18150199	27062	7804	26
RAI-3'8	A/G	rs8101662	NT_011109	18150911	27774	8516	27
RAI exon 6	A/T	rs6966	NT_011109	18151180	28043	8785	36
RAI-5-2	A/T	rs8112723	NT_011109	18153497	30360	10204	37
RAI intron 3	A/G	rs2017104	NT_011109	18155483	32346	12190	38

rs numbers were derived from the NCBI's database dbSNP.

Table 3d

dbSNP #		Position in SEQ ID NO: 1	SEQ ID NO:
rs3916787	gctgcagtga gctgt (-/ACACCTGTGGTCCCAGCTACTCTGG AAGCTGAGGTGGGAGGATCGCTTGAGCCCAAGAGGTG- GAGGCTGCAGTGAGCTGT) gactgtgccca ctgcactcca	632	33
rs2097215	TGACAGTAGA CATCCTGTCA T (A/G) ATAAGTCttt tttttt	1610	34
rs11878644	CATCCCCATA CCAAcccacc (c/t) tactgctctg atctcctcct	2790	35
rs7252567	tttagtagagacatggtccgccca (C/T) gttgccaggtggtcttgaactc	4428	21
rs3047560	ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG- CATGGTGGTGGGGTGC	4797	23
rs10422489	AAAAAAC- TAAAGTGGGGTTTGC GGG(G/T)AGTGGGAGGGCCCTTC CTGCTAGG	5491	24
rs10426701	ggatcacaag gtcaggagt (c/t) gagaccagcc tggccaacac	5798	25
rs4544343	TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTGTTGAG	7804	26
rs8101662	CTCGGGAGGCTGAGGCAGGAGAATC (A/G)CTTGAACCTCAGGCAGAGGTTG	8516	27
rs6966	ATTAAGTGCCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA	8785	36
rs8112723	ttgggagacc aaggcagggtg gatc (a/t) tttgaggtca gtagatcaaa	10204	37
rs2017104	gggaggctcg aggcgggc (A/G) gattgcatga gtcaggatt	12190	38

5

Even more preferably polymorphism(s) are those listed below in tables 3e and 3f,
more preferably at least two polymorphism(s) are selected, most preferably at least
three polymorphism(s) are selected from SEQ ID NO: 1 below:

10

Table 3e

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative Position	Position in SEQ ID NO:1	SEQ ID NO:
RAI-3'3	AA/-	rs3047560	NT_011109	18147192	24055	4797	23
RAI-3'6	A/C	10422489	NT_011109	18147886	24749	5491	24
RAI-3'5	C/T	10426701	NT_011109	18148193	25056	5798	25

RAI-3'4	A/G	rs4544343	NT_011109	18150199	27062	7804	26
RAI-3'8	A/G	rs8101662	NT_011109	18150911	27774	8516	27
RAI exon 6	A/T	rs6966	NT_011109	18151180	28043	8785	36
RAI-i5-2	A/T	rs8112723	NT_011109	18153497	30360	10204	37
RAI intron 3	A/G	rs2017104	NT_011109	18155483	32346	12190	38

rs numbers were derived from the NCBI's database dbSNP.

Table 3f

dbSNP #		Position in SEQ ID NO: 1	SEQ ID NO:
rs3047560	ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG- CATGGTGGTGGGGTGC	4797	23
rs10422489	AAAAAAC- TAAAGTGGGGTTTGC GGG(G/T)AGTGGGAGGGCCCTTC CTGCTAGG	5491	24
rs10426701	ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac	5798	25
rs4544343	TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTGTTGAG	7804	26
rs8101662	CTCGGGAGGCTGAGGCAGGAGAATC (A/G)CTTGAACCTCAGGCAGAGGTTG	8516	27
rs6966	ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA	8785	36
rs8112723	ttgggagacc aaggcagggtg gatc (a/t) ttgagggtca gtagatcaaa	10204	37
rs2017104	gggaggctcg aggcgggc (A/G) gattgcatga gctcaggatt	12190	38

- 5 Most preferably polymorphism(s) are those listed in tables 4a and 4b below, more preferably at least two polymorphism(s) are selected, most preferably at least three polymorphism(s) are selected from the polymorphisms shown below:

Table 4a

Trivial name	Kind	dbSNP #	Sequence	Position in SEQ ID 1	Relative Position	SEQ ID NO:
RAI-3'3	AA/-	rs3047560	NT_011109		24055	23
RAI exon 6	A/T	rs6966	NT_011109		28043	36
RAI intron 3	A/G	rs2017104	NT_011109		32346	38

rs numbers were derived from the NCBI's database dbSNP.

Table 4b

dbSNP #		Position in SEQ ID NO: 1	SEQ ID NO:
rs3047560	ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG- CATGGTGGTGGGGTGC	4797	23
rs6966	ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA	8785	36
rs2017104	gggaggctcg aggcgggc (A/G) gattgcatga gctcaggatt	12190	38

In yet another embodiment according to the methods of the invention, preferably the method for diagnosis as described herein, one or more single nucleotide polymorphism(s) at a predetermined position in the region (SEQ ID NO:1) are identified and used for e.g. cancer risk profiling and/or cancer treatment response profiling. In this embodiment the preferred polymorphism(s) are listed in Tables 5a, more preferably at least two polymorphism(s) are selected, most preferably at least three polymorphism(s) are selected. However, the present invention relates to any polymorphism in the region.

Table 5a

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative Position	Position in SEQ ID NO:1	SEQ ID NO:
XPD-81bp	81bp/-	rs3916787	NT_011109	18143027	19890	632	33
XPD-5'3	C/T	rs11878644	NT_011109	18145185	22048	2790	35
RAI-3'7	C/T	rs7252567	NT_011109	18146823	23686	4428	21
RAI-3'3	AA/-	rs3047560	NT_011109	18147192	24055	4797	23
RAI-3'6	A/C	10422489	NT_011109	18147886	24749	5491	24
RAI-3'5	C/T	10426701	NT_011109	18148193	25056	5798	25
RAI-3'4	A/G	rs4544343	NT_011109	18150199	27062	7804	26
RAI-3'8	A/G	rs8101662	NT_011109	18150911	27774	8516	27
RAI-i5-2	A/T	rs8112723	NT_011109	18153497	30360	10204	37
RAI intron 1-2	A/G		AC092309	24351	39069	18913	45
RAI intron 1-3	A/C		AC092309	25115	39833	19677	46

rs numbers were derived from the NCBI's database dbSNP.

More preferably, polymorphism(s) are those listed in table 5b below, more preferably at least two polymorphism(s) are selected, most preferably at least three polymorphism(s) are selected from the polymorphisms shown below:

Table 5b

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative Position	Position in SEQ ID NO:1	SEQ ID NO:
XPD-5'3	C/T	rs11878644	NT_011109	18145185	22048	2790	35
RAI-3'7	C/T	rs7252567	NT_011109	18146823	23686	4428	21
RAI-3'3	AA/-	rs3047560	NT_011109	18147192	24055	4797	23
RAI-3'6	A/C	10422489	NT_011109	18147886	24749	5491	24
RAI-3'5	C/T	10426701	NT_011109	18148193	25056	5798	25
RAI-3'4	A/G	rs4544343	NT_011109	18150199	27062	7804	26
RAI-3'8	A/G	rs8101662	NT_011109	18150911	27774	8516	27
RAI-i5-2	A/T	rs8112723	NT_011109	18153497	30360	10204	37
RAI intron 1-2	A/G		AC092309	24351	39069	18913	45
RAI intron 1-3	A/C		AC092309	25115	39833	19677	46

rs numbers were derived from the NCBI's database dbSNP.

In an even more preferred embodiment polymorphism(s) are those listed in table 5c below, more preferably at least two polymorphism(s) are selected, most preferably at least three polymorphism(s) are selected from the polymorphisms shown below:

Table 5c

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative Position	Position in SEQ ID NO:1	SEQ ID NO:
XPD-5'3	C/T	rs11878644	NT_011109	18145185	22048	2790	35
RAI-3'7	C/T	rs7252567	NT_011109	18146823	23686	4428	21
RAI-3'3	AA/-	rs3047560	NT_011109	18147192	24055	4797	23
RAI-3'6	A/C	10422489	NT_011109	18147886	24749	5491	24
RAI-3'5	C/T	10426701	NT_011109	18148193	25056	5798	25
RAI-3'4	A/G	rs4544343	NT_011109	18150199	27062	7804	26
RAI-3'8	A/G	rs8101662	NT_011109	18150911	27774	8516	27
RAI-i5-2	A/T	rs8112723	NT_011109	18153497	30360	10204	37

rs numbers were derived from the NCBI's database dbSNP.

In a most preferred embodiment polymorphism(s) are those listed in table 5d below, more preferably at least two polymorphism(s) are selected, most preferably at least three polymorphism(s) are selected from the polymorphisms shown below:

Table 5d

Trivial name	Kind	dbSNP #	Sequence	Position in sequence	Relative position	Position in SEQ ID NO:1	SEQ ID NO:
XPD-5'3	C/T	rs11878644	NT_011109	18145185	22048	2790	35
RAI-3'7	C/T	rs7252567	NT_011109	18146823	23686	4428	21
RAI-3'3	AA/-	rs3047560	NT_011109	18147192	24055	4797	23
RAI-3'6	A/C	10422489	NT_011109	18147886	24749	5491	24
RAI-3'5	C/T	10426701	NT_011109	18148193	25056	5798	25
RAI-3'4	A/G	rs4544343	NT_011109	18150199	27062	7804	26
RAI-3'8	A/G	rs8101662	NT_011109	18150911	27774	8516	27

rs numbers were derived from the NCBI's database dbSNP.

In a preferred embodiment at least one of the following combinations of polymorphisms is included in the methods:

1	XPD 5'3	RAI- 3'7	RAI- 3'3
2	XPD 5'3	RAI- 3'5	RAI- 3'4
3	XPD 5'3	RAI- 3'8	
4	XPD -81bp	XPD-5'3	RAI-3'7
5	XPD -81 bp	RAI-3'3	RAI 3'6
6	XPD -81 bp	RAI-3'5	RAI-3'4

7	XPD -81 bp	RAI-3'8	
8	XPD -81 bp	RAI intron 5-2	RAI intron 1-2
9	XPD -81 bp	RAI intron 1-3	
10	XPD -81 bp	RAI exon 6	RAI intron 3
11	RAI exon 6	XPD 5'3	RAI intron 3
12	RAI exon 6	RAI- 3'7	RAI intron 3
13	RAI exon 6	RAI 3'6	RAI intron 3
14	RAI exon 6	RAI-3'5	RAI intron 3
15	RAI exon 6	RAI-3'4	RAI intron 3
16	RAI exon 6	RAI-3'8	RAI intron 3
17	XPD-81bp	XPD-5'2	XPD-5'3
18	XPD-5'2	XPD-5'3	RAI-3'7
19	XPD-5'3	RAI-3'7	RAI-3'3
20	RAI-3'7	RAI-3'3	RAI-3'6
21	RAI-3'3	RAI-3'6	RAI-3'5
22	RAI-3'6	RAI-3'5	RAI-3'4
23	RAI-3'5	RAI-3'4	RAI-3'8
24	RAI-3'4	RAI-3'8	RAI exon6
25	RAI-3'8	RAI exon6	RAI intron5-2
26	RAI exon6	RAI intron5-2	RAI intron 3
27	RAI intron5-2	RAI intron 3	RAI intron 1
28	RAI intron 3	RAI intron 1	RAI intron 1-2
29	RAI intron 1	RAI intron 1-2	RAI intron 1-3
30	RAI intron 1-2	RAI intron 1-3	RAI-5'2
31	RAI intron 1-3	RAI-5'2	RAI-5'3
32	RAI-5'2	RAI-5'3	RAI-5'
33	RAI-5'3	RAI-5'	ASE1-5'2
34	RAI-5'	ASE1-5'2	ASE1-exon1
35	ASE1-5'2	ASE1-exon1	ERCC1-3'
36	ASE1-exon1	ERCC1-3'	
37	RAI-3'4	RAI-3'8	
38	RAI-3'3	RAI e6	RAI intron 3
39	XPD-5'2	RAIi3	RAIi1
40	XPD-5'2	RAIi1	RAI-5'2
41	RAIe6	RAIi1	ASE1e1
42	RAIe6	RAIi1	RAI-5'
43	RAIe6	RAIi1	ERCC1-3'
44	RAIe6	ASE1e1	ERCC1-3'
45	RAIe6	RAIi1	
46	RAIi3	RAI-5'	ERCC1-3'
47	RAIi3	ASE1e1	ERCC1-3'
48	RAIi1	RAI-5'2	
49	RAIi1	RAI-5'2	
50	RAIi1	RAI-5'2	RAI-5'
51	RAIi1	ASE1-5'2	ASE1e1
52	RAIi1	ASE1-5'2	ERCC1-3'
53	RAIi1	RAI-5'	ERCC1-3'
54	RAIi1	ASE1e1	
55	RAIi1	ASE1e1	ERCC1-3'
56	RAIi1	RAI-5'2	
57	RAIi1	ASE1-5'2	
58	RAIi1	ASE1e1	
59	RAIi1		
60	RAIi1	ERCC1-3'	
61	RAIi1		
62	RAIi1		
63	ASE1-5'2	RAI-5'	
64	ASE1-5'2	ASE1e1	

In another embodiment of the invention preferably the method described herein is one in which the tandem repeat is at a position as described in Table 6:

Table 6**Identification in uniSTS²**

D19S908

STS-W67936

5 D19S543

D19S393

STS-R48186

GDB:181915

RH47033

10 GDB:190019

² UniSTS is a database of unique sequence tag sites established by National Center for Biotechnology Information and located on the internet at

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists>

15 In another embodiment of the invention, the method for diagnosis described herein is preferably one in which the sequence polymorphism is in region r. Testing for the presence of the RAI gene allele is especially preferred because, without wishing to be bound by theoretical considerations, of its association with increased risk of cancer (as explained herein).

20

In one embodiment of the methods of the invention, preferably the method for diagnosis as described herein, one or more polymorphism(s) at a predetermined position in the region r (SEQ ID NO:1) are identified and used for e.g. cancer risk profiling and/or cancer treatment response profiling. Presently preferred polymorphism(s) are

25 the dinucleotide polymorphism RAI-3'3 in which AA or a deletion is present, the single nucleotide polymorphism RAI exon 6 in which A or T is present, and RAI intron 3 in which A or G is present. However, the present invention relates to any polymorphism and SNP in the r region.

30

The sequence polymorphism of the invention comprises at least one base difference, such as at least two base differences, such as at least three base differences, such as at least four base differences, such as eighty one base pair differences. As described above the sequence polymorphism(s) comprises at least one polymorphism, such as at least two polymorphisms, such as at least three polymorphisms,

such as at least four polymorphisms. Also, the sequence polymorphism comprises at least one polymorphism, such as at least two tandem repeat polymorphisms.

Also, the sequence polymorphism may be a combination of single nucleotide polymorphism and dinucleotide polymorphism, such as one single nucleotide polymorphism and one dinucleotide polymorphism.

The status of the individual may be determined by reference to allelic variation at one, two, three, four or more of the above loci.

The primary effectors i.e. the mutations causing cancer according to the present invention may be found within the region of SEQ ID NO: 132 including and flanked by the marker RAI-3'7 and the polymorphism having the sequence AAGTTTCTCTATT[G/T]TGTTTATAACA corresponding to position 18151158 of contig nt_011109 (SEQ ID NO.: 6). The region and markers herein are shown to be important in relation to cancer according to the present invention. The sequence polymorphisms considered to be primary effector may be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
rai-3'7	Rs7252567	18146823	TTTAGTAGAGACATGGTTCCGCCA[C/T]-GTTGCCCAGGCTGGTCTTGAACTCC	21
*RAI3'd1		18147012	GGTTTAT[ATTTT]Ntgagatggatttt	19
	Rs2377329	18147126	ctggggaggctgaggcagga-gaatc[A/G]cttgaaaccgggaggcggaggtgt	22
Rai3'3	Rs3047560	18147192	actaaaaataaaaaataaaaaaa[-/AA]atagccgagcatggtggtgggtgcc	23
*		18146233	GATTGTCATGT[G/T]ACATCAGCCAATACT	2
	rs10422489	18147886	AAAAAACTAAAGTGGGGTTTGCGGG[G/T]-AGTGGGAGGGCCCTTCCTGCTAGGT	24
	Rs10426701	18148193	caggcggatcacaaggtcaggagtt[C/T]-gagaccagcctggccaacacagtga	25

*		18149120	CACAGTGAAAC[C/T]CCATCTCTACTAAA	3
*		18149154	AAAATTAGCCGG[A/G]CGCCATGGCGGGAG	4
RAI3'4	Rs4544343	18150199	CCCTATGTTGTCCAAGCTGGCA- GAG[A/G]TTTTGTTTGTGTTTGTGAGAGGG A	26
*		18150815	AGCCTGGCCAACATG[C/G]TGAAACCCCGT CTCT	5
	rs8101662	18150911	ctcgggaggctgaggcagga- gaatc[A/G]cttgaactcaggaggcagaggttgc	27
*		18151158	AAGTTTCTCTATT[G/T]TGTTTATAAACA	6

* novel polymorphisms identified in the present invention

In yet another embodiment of the present invention the primary effectors i.e. the mutations causing cancer according to the present invention may be selected from the group consisting of

5

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
* RAI3'd1		18147012	GGTTTAT[ATTTT]NtgagatggatTTT	19
Rai3'3	Rs3047560	18147192	actaaaaataaaaaataaaaaaa[- /AA]atagccgagcatggtggtgggtgcc	23
	rs10422489	18147886	AAAAAACTAAAGTGGGGTTTGC GGG[G/T]- AGTGGGAGGGCCCTTCCTGCTAGGT	24
*		18149154	AAAAT- TAGCCGG[A/G]CGCCATGGCGGGAG	4
RAI3'4	Rs4544343	18150199	CCCTATGTTGTCCAAGCTGGCAGAG- [A/G]TTTTGTTTGTGTTTGTGAGAGGGA	26

In a preferred embodiment the primary effectors i.e. the mutations causing cancer according to the present invention may be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
--------------	-------	------------------	----------	-------------

*		18147012	GGTTTAT[ATTTT]Ntgagatggatttt	19
Rai3'd1				
Rai3'3	Rs3047560	18147192	actaaaaataaaaaataaaaaaa[- /AA]atagccgagcatggtggtgggtgcc	23
	rs10422489	18147886	AAAAAACTAAAGTGGGGTTTGCGGG[G/T]- AGTGGGAGGGCCCTTCCTGCTAGGT	24
*		18149154	AAAATTAGCCGG[A/G]CGCCATGGCGGGAG	4
RAI3'4	Rs4544343	18150199	CCCTATGTTGTCCAAGCTGGCAGAG- [A/G]TTTTTGTGTTGTTTGTGAGAGGGA	26

In another embodiment the primary effectors i.e. the mutations causing cancer according to the present invention may further be selected from the group consisting of

5

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18147012	GGTTTAT[ATTTT]Ntgagatggatttt	19
RAI3'd1				
Rai3'3	Rs3047560	18147192	actaaaaataaaaaataaaaaaa[- /AA]atagccgagcatggtggtgggtgcc	23
	rs10422489	18147886	AAAAAACTAAAGTGGGGTTTGCGGG[G/T]- AGTGGGAGGGCCCTTCCTGCTAGGT	24

In yet another embodiment the primary effectors i.e. the mutations causing cancer according to the present invention may further be selected from the group consisting of

10

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18147012	GGTTTAT[ATTTT]Ntgagatggatttt	19
RAI3'd1				
Rai3'3	Rs3047560	18147192	actaaaaataaaaaataaaaaaa[- /AA]atagccgagcatggtggtgggtgcc	23

In another embodiment the primary effectors i.e. the mutations causing cancer according to the present invention may further be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
rai-3'7	Rs7252567	18146823	TTTGTAGTAGAGACATGGTTCCGCCA[C/T]-GTTGCCCAGGCTGGTCTTGAAGTCC	21
	Rs2377329	18147126	ctggggaggctgaggcagga-gaatc[A/G]cttgaaaccgggaggcggaggtgt	22
	Rs10426701	18148193	caggcggatcacaaggtcaggagt[C/T]-gagaccagcctggccaacacagtga	25
*		18150815	AGCCTGGCCAACATG[C/G]TGAAACCCCGTCTCT	5
*		18151158	AAGTTTCTCTATT[G/T]TGTTTATAAACA	6

5

In another preferred embodiment the primary effectors i.e. the mutations causing cancer according to the present invention may further be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
* RAI3'd1		18147012	GGTTTAT[ATTTT]Ntgagatggatttt	19
*		18146233	GATTGTCATGT[G/T]ACATCAGCCAATACT	2
*		18149120	CACAGTGAAAC[C/T]CCATCTCTACTAAA	3
*		18149154	AAAATTAGCCGG[A/G]CGCCATGGCGGGAG	4
*		18150815	AGCCTGGCCAACATG[C/G]TGAAACCCGTCTCT	5
*		18151158	AAGTTTCTCTATT[G/T]TGTTTATAAACA	6

10

In yet another preferred embodiment the primary effectors i.e. the mutations causing cancer according to the present invention may further be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18147012	GGTTTAT[ATTTT]Ntgagatggatttt	19
RAI3'd1				
*		18149154	AAAATTAGCCGG[A/G]CGCCATGGCGGGAG	4

In yet another preferred embodiment the primary effectors i.e. the mutations causing cancer according to the present invention may further be selected from the group consisting of

5

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18146233	GATTGTCATGT[G/T]ACATCAGCCAATACT	2
*		18149120	CACAGTGAAAC[C/T]CCATCTCTACTAAA	3
*		18150815	AGCCTGGCCAACATG[C/G]TGAAACCCCGTCTCT	5
*		18151158	AAGTTTCTCTATT[G/T]TGTTTATAAACA	6

However, in another embodiment of the present invention the primary effectors i.e. the mutations causing cancer according to the present invention may be selected from the group of polymorphisms consisting of

10

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*RAI3'd1		18147012	GGTTTAT[ATTTT]Ntgagatggatttt	19
	Rs2377329	18147126	ctggggaggctgaggcagga-gaatc[A/G]cttgaaaccgggaggcggaggtgt	22

In yet another embodiment of the present invention the primary effectors i.e. the mutations causing cancer according to the present invention may be selected from the group of polymorphisms consisting of

15

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18149120	CACAGTGAAAC[C/T]CCATCTCTACTAAA	3
*		18149154	AAAATTAGCCGG[A/G]CGCCATGGCGGGAG	4

In a further embodiment of the present invention the polymorphism according to the present invention is

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18150815	AGCCTGGCCAACATG[C/G]TGAAACCCCGTCTCT	5

5

In yet a further embodiment of the invention the polymorphshism present in the region SEQ ID NO: 132 is

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18151158	AAGTTTCTCTATT[G/T]TGTTTATAAACA	6

10

The primary effector may be the mutation corresponding to the sequence TTTTAG-TAGAGACATGGTTCCGCCA[C/T]GTTGCCCAGGCTGGTCTTGA ACTCC positioned at 18146823 of Contig nt_011109 (SEQ ID NO.: 21) . In another embodiment the primary effector may be the mutation corresponding to the sequence

15 ctggggaggctgaggcaggagaatc[A/G]cttgaaaccgggaggcggaggttgt positioned at 18147126 of Contig nt_011109 (SEQ ID NO.: 22). However, in another preferred embodiment the primary effector may be the mutation corresponding to the sequence GATTGTCATGT[G/T]ACATCAGCCAATACT positioned at 18146233 of Contig nt_011109 (SEQ ID NO.: 2). In yet another embodiment the primary effector

20 may be the mutation corresponding to the sequence caggcggatcacaaggtcag-gagtt[C/T]gagaccagcctggccaacacagtga positioned at 18148193 of Contig nt_011109 (SEQ ID NO.: 25). In a further embodiment the primary effector may be the mutation corresponding to the sequence CACAGTGAAAC[C/T]CCATCTCTACTAAA posi-

tioned at 18149120 of Contig nt_011109 (SEQ ID NO.: 3). In another preferred embodiment the primary effector may be the mutation corresponding to the sequence AGCCTGGCCAACATG[C/G]TGAAACCCCGTCTCT positioned at 18150815 of Contig nt_011109 (SEQ ID NO.: 5). In yet another embodiment the primary effector may be the mutation corresponding to the sequence ctcgggaggctgaggcagga-gaatc[A/G]cttgaactcaggaggcagaggtgc positioned at 18150911 of Contig nt_011109 (SEQ ID NO.: 27). In a further embodiment the primary effector may be the mutation corresponding to the sequence AAGTTTCTCTATT[G/T]TGTTTATAAACA positioned at 18151158 of Contig nt_011109 (SEQ ID NO.: 6). However, in another preferred embodiment the primary effector may be the mutation corresponding to the sequence CCCTATGTTGTCCAAGCTGGCAGAG[A/G]TTTTT-GTTTGTTTGTTTGAGAGGGA positioned at 18150199 of Contig nt_011109 (SEQ ID NO.: 26). Similarly in yet a further embodiment, the primary effector may be the mutation corresponding to the sequence actaaaaataaaaaataaaaaaa[-/AA]atagccgagcatggtgggtgggtgcc positioned at 18147192 of Contig nt_011109 (SEQ ID NO.: 23). In another embodiment the primary effector may be the mutation corresponding to the sequence AAAAACTAAAGTGGGGTTTGCGGG[G/T]-AGTGGGAGGGCCCTTCCTGCTAGGT positioned at 18147886 of Contig nt_011109 (SEQ ID NO.: 24). In a further embodiment the primary effector may be the mutation corresponding to the sequence AAAATTAGCCGG[A/G]CGCCATGGCGGGAG positioned at 18149154 of Contig nt_011109 (SEQ ID NO.: 4). In an especially preferred embodiment the primary effector may be the mutation RAI3'd1 corresponding to the sequence GGTAT[ATTTT]Ntgagatggatttt positioned at 18147012 of Contig nt_011109 (SEQ ID NO.: 19), wherein N denotes the number of repeats of the ATTTT sequence. It is appreciated that the sequences may be disclosed as the complementary sequence. N is for example at least 2 repeats, such as at least 3 repeats, for example at least 4 repeats, such as at least 5 repeats, for example at least 6 repeats, such as at least 7 repeats, for example at least 8 repeats, such as at least 9 repeats, for example at least 10 repeats, such as at least 11 repeats, for example at least 12 repeats, such as at least 13 repeats, for example at least 14 repeats, such as at least 15 repeats, for example at least 16 repeats, such as at least 17 repeats, for example at least 18 repeats, such as at least 19 repeats, for example at least 20 repeats, such as at least 21 repeats, for example at least 22 repeats, such as at least 23 repeats, for example at least 24 repeats, such as at least 25 repeats, for example at least 26

repeats, such as at least 27 repeats, for example at least 28 repeats, such as at least 29 repeats, for example at least 30 repeats. In one embodiment of the invention the number of repeats ranging from at least 2 to at least 7 is correlated to an increased risk of developing breast cancer, for example ranging from at least 2 to at least 6 repeats, such as ranging from at least 2 to at least 5 repeats, for example ranging from at least 2 to at least 4 repeats, such as ranging from at least 3 to at least 5 repeats, for example ranging from at least 3 to at least 7 repeats, such as ranging from at least 3 to at least 6 repeats, for example ranging from at least 3 to at least 5 repeats, such as ranging from at least 3 to at least 4 repeats, for example ranging from at least 4 to at least 7 repeats, such as ranging from at least 4 to at least 6 repeats, for example ranging from at least 4 to at least 5 repeats. In general, the lower the number of repeats the higher the risk of developing breast cancer.

Information on the number of repeats may be combined with information on the age of the woman concerned. This combined information correlates to the risk of developing breast cancer. Furthermore, the number of repeats may be combined with information on the homozygosity or heterozygosity in the woman concerned. This combined information correlates to the risk of developing breast cancer. Additionally, information on the number of repeats may be combined with information on the age of the woman concerned and the homozygosity or heterozygosity status of the woman. This combined information correlates to the risk of developing breast cancer.

In one embodiment of the present invention homozygosity with regard to the presence of N repeats of ATTTT (ie. the presence of N repeats of ATTTT on both alleles) is correlated to an increased risk of developing breast cancer. In another embodiment the heterozygosity with regard to the presence of N repeats of ATTTT (ie. the presence of N repeats of ATTTT on one allele) is correlated to an increased risk of developing breast cancer.

Thus, in one embodiment the presence of N repeats of ATTTT on both alleles is correlated to an increased risk of developing breast cancer. In another embodiment of the present invention the presence of N repeats of ATTTT on one allele is correlated to an increased risk of developing breast cancer. The increased risk of devel-

oping breast cancer is correlated to all age groups. However, the increased risk of developing breast cancer is correlated to the age group under 55 years of age. In a further embodiment the increased risk of developing breast cancer is correlated to the age group between 55-60 years of age. Furthermore, for women aged over 60 years the risk of developing breast cancer is increased.

In one embodiment the presence of RAI3'd1 on both alleles wherein N is at least 4 in women in all age groups combined is correlated with an increased risk of developing breast cancer. In another embodiment of the present invention the presence of RAI3'd1 on both alleles wherein N is at least 4 in women under 55 years of age is correlated with an increased risk of developing breast cancer.

In yet another embodiment the presence of RAI3'd1 on one allele wherein N is at least 4 in women in all age groups combined, and/or women aged under 55 years, and/or women aged 55-60 years, and/or women over 60 years is correlated with an increased risk of developing breast cancer.

The primary effectors may however be any combination of the polymorphisms mentioned herein.

Among the polymorphisms employed herein for providing a method and/or compositions for identifying human subjects with an increased risk of having or developing disease, a number of polymorphisms are novel and have been identified by the present inventors. For such polymorphisms the position according to Contig nt_011109 and the nucleotide sequence of the identified polymorphism are provided, whereas the novel polymorphisms cannot be assigned trivial names or identification numbers according to the dbSNP database.

The primary effectors i.e. the mutations causing cancer according to the present invention may be found within the region around position 39000 which is shown to be important in relation to cancer according to the present invention. The region includes and is flanked by the marker Rai intron 1 and the polymorphism having the sequence TCCAGCCTGGGCAAGAA[C/G]AGTGAAACTCCAGCTT corresponding to position 18165052 of contig nt_011109. The primary effector may be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
RAli1	Rs1970764	18159091	ggagcttgagctgagctga-gatgcg[A/G]ccactgcactccagcctggcgaca	28
*		18159263	TTCTCCTGACCTC[A/G]TGATCCGCCCACCTCGG	7
*		18160363	GGGATTACAGGCATGC-[A/G]CCACCAGGCCAGCTAATTTTTGT	8
*		18160936	TCCAATGGTGA-CA[A/C]CAGTAAGAGCAGTTAACAG	9
*		18160937	TCCAATGGTGA-CAA[C/G]AGTAAGAGCAGTTAACAG	10
	rs10419090	18161433	tacaggcgcccgccaccacccc-cag[A/C]taattttgtatttttagtagagac	29
*		18161694	TTGCCTCAGCCTCCT-GA[G/T]TAGCTGGGATTGGAATGAGA	11
*		18161841	TACGATAAATAGCTA-GA[C/GACCTTGGCGCCACCATCTT]	20
*		18161896	AAAATAATAATAATAATAT-TAA[C/T]CCTGACCTTGGCGCCACCATCT	12
RAI intron 1-2 (own1)	rs6509210	18162206	tcgtctgtctacagaattacagg-ca[C/T]gcgccaccgctccgggctaatttt	30
*		18162309	CCTCATGAGCCACC-CAC[C/T]TCGGCCTCCCAAAGTGCT	13
*		18162356	TGAGC-CACCGCGCCC[A/G]GCCGAGACTCACTATTT	14
	rs12986272	18162599	Taaagcgggaggat-ggctgaacct[A/G]ggaggcgagggtgagtgagccga	31
*		18162903	GGAGAGAAGGAGCAGA-GAAC[A/C]TCTCTATGTGGCCA	15

RAI intron 1-3 (own2)		18162970	ATCCTAAAGACTAC[A/C]TTTCCCAGCATCCCA	32
*		18162986	TTTCCCAGCATCC- CA[C/T]TGCAATGAGGCTCCTGGCC	16
*		18163200	TCCTGACTC- CAGTG[A/C]GGTGCCTACAGTCCTG	17
*		18165052	TCCAGCCTGGGCAA- GAA[C/G]AGTGAAACTCCAGCTT	18

In another embodiment of the invention the primary effectors i.e. the mutations causing cancer according to the present invention may be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18160363	GGGATTACAGGCATGC- [A/G]CCACCAGGCCAGCTAATTTTTGT	8
*		18160936	TCCAATGGTGA- CA[A/C]CAGTAAGAGCAGTTAACAG	9
*		18160937	TCCAATGGTGA- CAA[C/G]AGTAAGAGCAGTTAACAG	10
	rs10419090	18161433	tacaggcgccgccaccacccc- cag[A/C]taattttgtatttttagtagagac	29
	rs12986272	18162599	taaagcgggaggat- ggcttgaacct[A/G]ggaggcggaggtgcagtgagccga	31
*		18165052	TCCAGCCTGGGCAA- GAA[C/G]AGTGAAACTCCAGCTT	18

5

In yet another embodiment of the invention the primary effectors i.e. the mutations causing cancer according to the present invention may be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
-----------------	-------	---------------------	----------	-------------------

*		18160363	GGGATTACAGGCATGC-[A/G]CCACCAGGCCAGCTAATTTTTGT	8
*		18160936	TCCAATGGTGA-CA[A/C]CAGTAAGAGCAGTTAACAG	9
*		18160937	TCCAATGGTGA-CAA[C/G]AGTAAGAGCAGTTAACAG	10
	rs10419090	18161433	tacaggcgcccgccaccacccc-cag[A/C]taattttgtatttttagtagagac	29

In a further embodiment of the invention the primary effectors i.e. the mutations causing cancer according to the present invention may be selected from the group consisting of

5

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18160363	GGGATTACAGGCATGC-[A/G]CCACCAGGCCAGCTAATTTTTGT	8
*		18160936	TCCAATGGTGA-CA[A/C]CAGTAAGAGCAGTTAACAG	9
*		18160937	TCCAATGGTGA-CAA[C/G]AGTAAGAGCAGTTAACAG	10

In another embodiment of the invention the primary effectors i.e. the mutations causing cancer according to the present invention may be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18160363	GGGATTACAGGCATGC-[A/G]CCACCAGGCCAGCTAATTTTTGT	8
*		18160937	TCCAATGGTGA-CAA[C/G]AGTAAGAGCAGTTAACAG	10

10

In yet a further embodiment of the invention the primary effectors i.e. the mutations causing cancer according to the present invention may be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
	rs12986272	18162599	taaagcgggag-gatggcttgaacct[A/G]ggaggcggagggttcagtgagccga	31
*		18165052	TCCAGCCTGGGCAA-GAA[C/G]AGTGAAACTCCAGCTT	18

5

In a further embodiment the primary effector may be selected from the group consisting of

Trivial name	dbSNP	Contig nt_011109	Sequence	SEQ ID NO.:
*		18159263	TTCTCCTGACCTC[A/G]TGATCCGCCCACCTCGG	7
*		18160363	GGGATTACAGGCATGC-[A/G]CCACCAGGCCAGCTAATTTTGT	8
*		18160936	TCCAATGGTGA-CA[A/C]CAGTAAGAGCAGTTAACAG	9
*		18160937	TCCAATGGTGA-CAA[C/G]AGTAAGAGCAGTTAACAG	10
*		18161694	TTGCCTCAGCCTCCT-GA[G/T]TAGCTGGGATTGGAATGAGA	11
*		18161841	TACGATAAATAGCTA-GA[C/G]ACCTTGGCGCCACCATCTT]	20
*		18161896	AAAATAATAATAATAT-TAA[C/T]CCTGACCTTGGCGCCACCATCT	12
*		18162309	CCTCATGAGCCACC-CAC[C/T]TCGGCCTCCCAAAGTGCT	13

*		18162356	TGAGC- CACCGCGCCC[A/G]GCCGAGACTCACTATTT	14
*		18162903	GGAGAGAAGGAGCAGA- GAAC[A/C]TCTCTATGTGGCCA	15
*		18162986	TTTCCCAGCATCC- CA[C/T]TGCAATGAGGCTCCTGGCC	16
*		18163200	TCCTGACTC- CAGTG[A/C]GGTGCCTACAGTCCTG	17
*		18165052	TCCAGCCTGGGCAA- GAA[C/G]AGTGAACTCCAGCTT	18

In one embodiment of the present invention the primary effector may be the mutation corresponding to the sequence ggagcttgagtgagctga-
gatcgc[A/G]ccactgcactccagcctggcgaca positioned at 18159091 of Contig
5 nt_011109 (SEQ ID NO.: 28). In another embodiment the primary effector may be the mutation corresponding to the sequence
TTCTCCTGACCTC[A/G]TGATCCGCCCACCTCGG positioned at 18159263 of
Contig nt_011109 (SEQ ID NO.: 7). In yet another embodiment the primary effector
may be the mutation corresponding to the sequence GGGATTACAGG-
10 CATGC[A/G]CCACCAGGCCCCAGCTAATTTTTGT positioned at 18160363 of Contig
nt_011109 (SEQ ID NO.: 8). In a further embodiment the primary effector may be the mutation corresponding to the sequence
TCCAATGGTGACA[A/C]CAGTAAGAGCAGTTAACAG positioned at 18160936 of
Contig nt_011109 (SEQ ID NO.: 9). In a preferred embodiment the primary effector
15 may be the mutation corresponding to the sequence
TCCAATGGTGACAA[C/G]AGTAAGAGCAGTTAACAG positioned at 18160937 of
Contig nt_011109 (SEQ ID NO.: 10). In a further preferred embodiment the primary
effector may be the mutation corresponding to the sequence tacaggcgccccccac-
cacccccag[A/C]taattttgtatttttagtagagac positioned at 18161433 of Contig nt_011109
20 (SEQ ID NO.: 29). In a further preferred embodiment the primary effector may be the mutation corresponding to the sequence
TTGCCTCAGCCTCCTGA[G/T]TAGCTGGGATTGGAATGAGA positioned at
18161694 of Contig nt_011109 (SEQ ID NO.: 11). However, in another embodiment
the primary effector may be the mutation corresponding to the sequence TACGA-

TAAATAGCTAGA[C/GACCTTGGCGCCACCATCTT] positioned at 18161841 of Contig nt_011109 (SEQ ID NO.: 20). In an especially preferred embodiment the primary effector may be the mutation corresponding to the sequence AAAATAATAATAATAATATTAA[C/T]CCTGACCTTGGCGCCACCATCT positioned at 18161896 of Contig nt_011109 (SEQ ID NO.: 12). In further preferred embodiment the primary effector may be the mutation corresponding to the sequence tcgtcctgctacagaattacaggca[C/T]gcgccaccgctccgggctaatttt positioned at 18162206 of Contig nt_011109 (SEQ ID NO.: 30). In another embodiment the primary effector may be the mutation corresponding to the sequence CCTCAT-GAGCCACCCAC[C/T]TCGGCCTCCCAAAGTGCT positioned at 18162309 of Contig nt_011109 (SEQ ID NO.: 13). In yet another embodiment the primary effector may be the mutation corresponding to the sequence TGAGCCACCGCGCCC[A/G]GCCGAGACTCACTATTT positioned at 18162356 of Contig nt_011109 (SEQ ID NO.: 14). In yet a further embodiment the primary effector may be the mutation corresponding to the sequence taaagcgggag-gatggcttgaacct[A/G]ggaggcggagggtgcagtgagccga positioned at 18162599 of Contig nt_011109 (SEQ ID NO.: 31). In a further embodiment. In an additional embodiment the primary effector may be the mutation corresponding to the sequence GGAGA-GAAGGAGCAGAGAAC[A/C]TCTCTATGTGGCCA positioned at 18162903 of Contig nt_011109 (SEQ ID NO.: 15). However, in another embodiment the primary effector may be the mutation corresponding to the sequence ATCCTAAAGAC-TAC[A/C]TTTCCCAGCATCCCA positioned at 18162970 of Contig nt_011109 (SEQ ID NO.: 32). Furthermore, in a further embodiment the primary effector may be the mutation corresponding to the sequence TTTCCCAGCATC-CCA[C/T]TGCAATGAGGCTCCTGGCC positioned at 18162986 of Contig nt_011109 (SEQ ID NO.: 16). In yet a further embodiment the primary effector may be the mutation corresponding to the sequence TCCTGACTCCAGTG[A/C]GGTGCCTACAGTCCTG positioned at 18163200 of Contig nt_011109 (SEQ ID NO.: 17). In yet a preferred embodiment the primary effector may be the mutation corresponding to the sequence TTCCAGCCTGGGCAAGAA[C/G]AGTGAAACTCCAGCTT positioned at 18165052 of Contig nt_011109 (SEQ ID NO.: 18).

The primary effectors may also be found within the region corresponding to SEQ ID NO: 133, wherein the polymorphisms are shown to be important in relation to the

present invention. The region includes and is flanked by the polymorphisms RAI 5' (SEQ ID NO: 4) and ASE 1 exon 3-2 also named ERCC1 3'3 or ASE1e1b (SEQ ID NO: 142). The polymorphisms may be selected from the group consisting of

Trivial name	dbSNP	Position in Contig nt_011109	Sequence	SEQ ID NO.:
RAI5'	Rs4572514	18171984	AGAACCTGTTTCAGGCTGGCGGCTCA[C/T]TTGGATGAACAGGGAGTGTGTGAC	41
ASE 1-5'2	Rs2226949	18178152	CCCCCTTCTTAGGAC-GCATGGGGGT[G/T]GAGAGAACGGGGAGATAGACAGAG	42
ASE 1 exon 1	Rs967591	18178152	TGCGAGCAGCCCCGGGCTA-CAGGGTT[A/G]CCTGAGGTGTGGGTCCCAGGATGG	43
ERCC1-3'	Rs762562	18180561	GGCGCCTCAACAGCCAGAAG-GAGCG[A/G]AGCCTCAGGCCAGGCAGCTCTGG	44
ASE 1 exon 3-2 (also ERCC1-3'3)	Rs2336219	18180624	AGAAAGAAAAACAGCAA A/G ATGCCA-CAGTGGAGCCAGAG	SEQ ID NO: 142

5

In another embodiment of the present invention the polymorphisms are selected from the group consisting of

Trivial name	dbSNP	Position in Contig nt_011109	Sequence	SEQ ID NO.:
ASE 1-5'2	Rs2226949	18178152	CCCCCTTCTTAGGAC-GCATGGGGGT[G/T]GAGAGAACGGGGAGATAGACAGAG	42
ASE 1 exon 1	Rs967591	18178152	TGCGAGCAGCCCCGGGCTA-CAGGGTT[A/G]CCTGAGGTGTGGGTCCCAGGATGG	43
ERCC1-3'	Rs762562	18180561	GGCGCCTCAACAGCCAGAAG-GAGCG[A/G]AGCCTCAGGCCAGGCAGCTCTGG	44
ASE 1 exon 3-2 (also ERCC1-3'3)	Rs2336219	18180624	AGAAAGAAAAACAGCAA A/G ATGCCA-CAGTGGAGCCAGAG	SEQ ID NO: 142

In yet another embodiment of the present invention the polymorphisms are selected

10

from the group consisting of

Trivial name	dbSNP	Position in Contig nt_011109	Sequence	SEQ ID NO.:
RAI5'	Rs4572514	18171984	AGAACCTGTTTCAGGCTGGCGGCTCA[C/T]TTGGATGAACAGGGAGTGTGTGAC	41
ASE 1 exon 1	Rs967591	18178152	TGCGAGCAGCCCGGGCTA-CAGGGTT[A/G]CCTGAGGTGTGGGTCCCAGGATGG	43
ERCC1-3'	Rs762562	18180561	GGCGCCTCAACAGCCAGAAG-GAGCG[A/G]AGCCTCAGGCCAGGCAGCTCTGG	44
ASE 1 exon 3-2 (also ERCC1-3'3)	Rs2336219	18180624	AGAAAGAAAAACAGCAA A/G ATGCCA-CAGTGGAGCCAGAG	SEQ ID NO: 142

In a further embodiment of the present invention the polymorphisms are selected from the group consisting of

Trivial name	dbSNP	Position in Contig nt_011109	Sequence	SEQ ID NO.:
RAI5'	Rs4572514	18171984	AGAACCTGTTTCAGGCTGGCGGCTCA[C/T]TTGGATGAACAGGGAGTGTGTGAC	41
ASE 1-5'2	Rs2226949	18178152	CCCCCTTCTTAGGAC-GCATGGGGGT[G/T]GAGAGAACGGGGAGATAGACAGAG	42
ASE 1 exon 1	Rs967591	18178152	TGCGAGCAGCCCGGGCTA-CAGGGTT[A/G]CCTGAGGTGTGGGTCCCAGGATGG	43
ASE 1 exon 3-2 (also ERCC1-3'3)	Rs2336219	18180624	AGAAAGAAAAACAGCAA A/G ATGCCA-CAGTGGAGCCAGAG	SEQ ID NO: 142

5

In yet a further embodiment of the present invention the polymorphisms are selected from the group consisting of

Trivial name	dbSNP	Position in Contig nt_011109	Sequence	SEQ ID NO.:
--------------	-------	------------------------------	----------	-------------

ASE 1 exon 1	Rs967591	18178152	TGCGAGCAGCCCGGGCTA- CAGGGTT[A/G]CCTGAGGTGTGGGTCCCAGG ATGG	43
ASE 1 exon 3-2 (also ERCC1-3'3)	Rs2336219	18180624	AGAAAGAAAAACAGCAA A/G ATGCCA- CAGTGGAGCCAGAG	SEQ ID NO: 142

Thus in one embodiment of the present invention the at least one polymorphism is RAI5' (SEQ ID NO: 41). However, in another embodiment the at least one polymorphism is ASE 1- 5'2 (SEQ ID NO:42). A further embodiment is ERCC1-3' (SEQ ID NO: 44). A preferred embodiment of the invention is the at least one polymorphism ASE 1 exon 3-2 (SEQ ID NO: 142). In yet a further preferred embodiment the at least one polymorphism is ASE 1 exon 1 (SEQ ID NO:43).

In a further preferred embodiment the primary effectors may however be any combination of the polymorphisms.

One aspect of the present invention relates to a method for estimating the disease risk of an individual comprising assessing at least one sequence polymorphism in a region corresponding to SEQ ID NO: 132 and optionally at least one further sequence polymorphism in a region corresponding to SEQ ID NO: 133 with the proviso that when using SEQ ID NO:132 only, the at least one polymorphism is in a region corresponding to at least one region within SEQ ID NO:132 selected from the regions flanked by and including the sequence polymorphism RAI3'd1 (SEQ ID NO: 19) and rs 2377329 (SEQ ID NO: 22). In one embodiment the at least one polymorphism is RAI3'd1 (SEQ ID NO: 19). In another embodiment the at least one polymorphism is rs 2377329 (SEQ ID NO: 22).

However, another embodiment of the present invention relates to a method for estimating the disease risk of an individual comprising assessing at least one sequence polymorphism in a region corresponding to SEQ ID NO: 132 and optionally at least one further sequence polymorphism in a region corresponding to SEQ ID NO: 133 with the proviso that when using SEQ ID NO: 132 only, the at least one polymorphism is in a region corresponding to at least one region within SEQ ID NO: 132 selected from the regions flanked by and including the sequence polymorphism positioned at 18149120 (SEQ ID NO: 3) and sequence polymorphism positioned at 1814154 (SEQ ID NO: 4). Thus, in one embodiment of the present invention the

sequence polymorphism is the sequence polymorphism positioned at 18149120 (SEQ ID NO: 3). In another embodiment the is the sequence polymorphism is the sequence polymorphism positioned at 18149154 (SEQ ID NO: 4).

- 5 In yet other embodiments under said proviso when at least one polymorphism is assessed in the region corresponding to SEQ ID NO: 132 the sequence polymorphism is positioned at 18450815 (SEQ ID NO: 22), or for example at position 18151158 (SEQ ID NO: 6)
- 10 The at least one sequence polymorphisms residing in the region corresponding to SEQ ID NO: 132 may optionally further be combined with at least one polymorphism sequence polymorphism in the region corresponding to SEQ ID NO: 133 in order to estimate the disease risk of an individual. For example the at least one sequence polymorphism is assessed in a region of SEQ ID NO: 133 flanked by and including
- 15 the polymorphisms RAI 5' (SEQ ID NO.:4 and ASE 1 exon 3-2 (SEQ ID NO: 142). For example the at least one sequence polymorphism is assessed in a region of SEQ ID NO: 132 flanked by and including the regions flanked by and including the sequence polymorphism RAI3'd1 (SEQ ID NO: 19) and rs2377329 (SEQ ID NO: 22), and at least one sequence polymorphism is assessed in a region of SEQ ID
- 20 NO: 133 flanked by and including the polymorphisms RAI 5' (SEQ ID NO: 4 and ASE 1 exon 3-2 (SEQ ID NO: 142)

- In another embodiment at least one sequence polymorphism is assessed in a region of SEQ ID NO: 132 flanked by and including the regions flanked by and including
- 25 the sequence polymorphism positioned at 18149120 in Contig NT_011109 (SEQ ID NO: 3) and sequence polymorphism positioned at 18149154 in Contig NT_011109 (SEQ ID NO: 4), and at least one sequence polymorphism is assessed in a region of SEQ ID NO: 133 flanked by and including the polymorphisms RAI 5' (SEQ ID NO: 4 and ASE 1 exon 3-2 (SEQ ID NO: 142)

30

In yet another embodiment according to the present invention the at least one sequence polymorphism is assessed in SEQ ID NO: 133 said polymorphism is positioned at 18150815 in Contig NT_011109 (SEQ ID NO: 5), and at least one sequence polymorphism is assessed in a region of SEQ ID NO: 133 flanked by and

including the polymorphisms RAI 5' (SEQ ID NO: 4 and ASE 1 exon 3-2 (SEQ ID NO: 142).

Further embodiments include assessing at least one sequence polymorphism in SEQ ID NO: 132 said polymorphism is positioned at 18151158 in Contig NT_011109 (SEQ ID NO: 6), and at least one sequence polymorphism is assessed in a region of SEQ ID NO: 133 flanked by and including the polymorphisms RAI 5' (SEQ ID NO: 4 and ASE 1 exon 3-2 (SEQ ID NO: 142).

Yet other embodiments include a method for estimating the disease risk by assessing for example the sequence polymorphism RAI3' d1 (SEQ ID NO: 19) in SEQ ID NO: 132, and the sequence polymorphism assessed in SEQ ID NO: 133 is ASE 1 exon 3-2 (SEQ ID NO.: 142), such as the sequence polymorphism RAI3' d1 (SEQ ID NO: 19) in SEQ ID NO: 132, and the sequence polymorphism assessed in SEQ ID NO: 133 being ASE 1 exon 3-2 (SEQ ID NO: 142).

Cell sample

The cell sample used in the present invention may be any suitable cell sample capable of providing the genetic material for use in the method. In a preferred embodiment, the cell sample is a blood sample, a tissue sample, a sample of secretion, semen, ovum, a washing of a body surface (e.g. a buccal swap), a clipping of a body surface (hairs, or nails), such as wherein the cell is selected from white blood cells and tumour tissue.

It will be appreciated that the test sample may equally be a nucleic acid sequence corresponding to the sequence in the test sample, that is to say that all or a part of the region in the sample nucleic acid may firstly be amplified using any convenient technique e.g. PCR, before use in the analysis of variation in the region.

Detection methods

Detection may be conducted on the sequence of SEQ ID NO: 1 or a complementary sequence as well as on transcriptional (mRNA and/or cDNA) and translational products (polypeptides, proteins) therefrom. Similarly detection may thus be conducted on the sequence of SEQ ID NO: 132 and optionally further on SEQ ID NO: 133 or complementary sequences as well as on transcriptional (mRNA, and/or cDNA) and translational products (polypeptides, proteins) therefrom.

It will be apparent to the person skilled in the art that there are a large number of analytical procedures which may be used to detect the presence or absence of variant nucleotides at one or more of positions mentioned herein in the r region. Mutations or polymorphisms within or flanking the r region can be detected by utilizing a number of techniques. Nucleic acid from any nucleated cell can be used as the starting point for such assay techniques, and may be isolated according to standard nucleic acid preparation procedures that are well known to those of skill in the art. In general, the detection of allelic variation requires a mutation discrimination technique, optionally an amplification reaction and a signal generation system. Table 7 lists a number of mutation detection techniques, some based on the PCR. These may be used in combination with a number of signal generation systems, a selection of which is listed in Table 8. Further amplification techniques are listed in Table 9. Many current methods for the detection of allelic variation are reviewed by Nollau et al., Clin. Chem. 43, 1114-1120, 1997; and in standard textbooks, for example "Laboratory Protocols for Mutation Detection", Ed. by U. Landegren, Oxford University Press, 1996 and "PCR", 2nd Edition by Newton & Graham, BIOS Scientific Publishers Limited, 1997.

20 **Table 7**

Abbreviations:

ALEX	Amplification refractory mutation system linear extension
APEX	Arrayed primer extension
ARMS	Amplification refractory mutation system
25 b-DNA	Branched DNA
CMC	Chemical mismatch cleavage
bp	base pair
COPS	Competitive oligonucleotide priming system
DGGE	Denaturing gradient gel electrophoresis
30 FRET	Fluorescence resonance energy transfer
LCR	Ligase chain reaction
MASDA	Multiple allele specific diagnostic assay
NASBA	Nucleic acid sequence based amplification
OLA	Oligonucleotide ligation assay
35 PCR	Polymerase chain reaction

	PTT	Protein truncation test
	RFLP	Restriction fragment length polymorphism
	SDA	Strand displacement amplification
	SNP	Single nucleotide polymorphism
5	SSCP	Single-strand conformation polymorphism analysis
	SSR	Self sustained replication
	TGGE	Temperature gradient gel electrophoresis

10 Table 8 illustrates various mutation detection techniques capable of being used for SNP detection.

Table 8

General techniques: DNA sequencing, Sequencing by hybridisation, SNAPshot.

15 Scanning techniques: PJT*, SSCP, DOGE, TGGE, Cleavase, Heteroduplex analysis, CMC, Enzymatic mismatch cleavage

Hybridisation Based techniques

20 Solid phase hybridisation: Dot blots, MASDA, Reverse dot blots, Oligonucleotide arrays (DNA Chips)

25 Solution phase hybridisation: Taqman --U.S. Pat. No. 5,210,015 & 5,487,972 (Hoffmann-La Roche), Molecular Beacons -- Tyagi et al (1996), Nature Biotechnology, 14, 303; WO 95/13399 (Public Health Inst., New York), Lightcycler, optionally in combination with FRET.

30 Extension Based: ARMS, ALEX -- European Patent No. EP 332435 B1 (Zeneca Limited), COPS -- Gibbs et al (1989), Nucleic Acids Research, 17, 2347.

Incorporation Based: Mini-sequencing, APEX

Restriction Enzyme Based: RFLP, Restriction site generating PCR

35 Ligation Based: OLA

Other: Invader assay

Various Signal Generation or Detection Systems is listed below:

5

Fluorescence: FRET, Fluorescence quenching, Fluorescence polarisation--United Kingdom Patent No. 2228998 (Zeneca Limited)

10

Other: Chemiluminescence, Electrochemiluminescence, Raman, Radioactivity, Colorimetric, Hybridisation protection assay, Mass spectrometry

Table 9 illustrates examples of further amplification techniques.

Table 9

15

SSR, NASBA, LCR, SDA, b-DNA

Preferred mutation detection techniques include ARMS, ALEX, COPS, Taqman, Molecular Beacons, RFLP, and restriction site based PCR and FRET techniques.

20

Particularly preferred methods include FRET; taqman, ARMS and RFLP based methods.

25

In a preferred embodiment, mutations or polymorphisms can be detected by using a microassay of nucleic acid sequences immobilized to a substrate or "gene chip" (see, e.g. Cronin, et al., 1996, Human Mutation 7:244-255).

30

Further, improved methods for analyzing DNA polymorphisms, which can be utilized for the identification of region r specific mutations, have been described that capitalize on the presence of variable numbers of short, tandemly repeated DNA sequences between the restriction enzyme sites. For example, Weber (U.S. Pat. No. 5,075,217) describes a DNA marker based on length polymorphisms in blocks of (dC-dA)_n-(dG-dT)_n short tandem repeats. The average separation of (dC-dA)_n-(dG-dT)_n blocks is estimated to be 30,000-60,000 bp. Markers that are so closely spaced exhibit a high frequency co-inheritance, and are extremely useful in the iden-

tification of genetic mutations, such as, for example, mutations within the RAI gene, and the diagnosis of diseases and disorders related to RAI mutations.

Also Caskey et al. (U.S. Pat. No. 5,364,759) describe a DNA profiling assay for detecting short tri and tetra nucleotide repeat sequences. The process includes extracting the DNA of interest, such as the RAI gene, amplifying the extracted DNA, and labelling the repeat sequences to form a genotypic map of the individual's DNA.

The level of RAI gene expression can also be assayed. For example, RNA from a cell type or tissue known, or suspected, to express the RAI gene may be isolated and tested utilizing hybridization or PCR techniques such as are described, above. The isolated cells can be derived from cell culture or from a patient. The analysis of cells taken from culture may be a necessary step in the assessment of cells to be used as part of a cell-based gene therapy technique or, alternatively, to test the effect of compounds on the expression of the RAI gene. Such analyses may reveal both quantitative and qualitative aspects of the expression pattern of the RAI gene, including activation or inactivation of RAI gene expression.

In one embodiment of such a detection scheme, a cDNA molecule is synthesized from an RNA molecule of interest (e.g., by reverse transcription of the RNA molecule into cDNA). A sequence within the cDNA is then used as the template for a nucleic acid amplification reaction, such as a PCR amplification reaction, or the like. The nucleic acid reagents used as synthesis initiation reagents (e.g., primers) in the reverse transcription and nucleic acid amplification steps of this method are chosen from among the RAI gene nucleic acid reagents described above. The preferred lengths of such nucleic acid reagents are at least 9-30 nucleotides. For detection of the amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that the product may be visualized by standard ethidium bromide staining or by utilizing any other suitable nucleic acid staining method.

Additionally, it is possible to perform such RAI gene expression assays "in situ", i.e., directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents such as those described above may be used as probes and/or prim-

ers for such in situ procedures (see, for example, Nuovo, G. J., 1992, "PCR In Situ Hybridization: Protocols And Applications", Raven Press, NY).

Alternatively, if a sufficient quantity of the appropriate cells can be obtained, standard Northern analysis can be performed to determine the level of mRNA expression of the RAI gene.

Activity of the gene

Another method for detecting sequence polymorphism is by analysing the activity of gene products resulting from the sequences. Accordingly, in one embodiment the detection uses the activity of the RAI gene product as compared to a reference in the method. In particular if the activity of the genes are decreased or increased by at least or about 50 %, such as at least or about 40%, for example at least or about 30%, such as at least or about 20%, for example at least or about 10%, such as at least or about 10%, for example at least or about 5%, such as at least or about 2%, it indicates a sequence polymorphism in the gene.

Mutations outside the region

The present invention may combine the result of sequence polymorphism within the region r with sequence polymorphism outside the region in order to increase the probability of the correlation. The sequence polymorphisms outside the region which may be indicative of diseases according to the method of the present invention are linked to the primary effectors.

Primers

The primer nucleotide sequences of the invention further include: (a) any nucleotide sequence that hybridizes to a nucleic acid molecule of the region r or its complementary sequence or RNA products under stringent conditions, e.g., hybridization to filter-bound DNA in 6x sodium chloride/sodium citrate (SSC) at about 45°C followed by one or more washes in 0.2x SSC/0.1% SDS at about 50-65°C, or (b) under highly stringent conditions, e.g., hybridization to filter-bound nucleic acid in 6x SSC at about 45°C followed by one or more washes in 0.1x SSC/0.2% SDS at about 68°C, or under other hybridization conditions which are apparent to those of skill in the art (see, for example, Ausubel F.M. et al., eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New

York, at pp. 6.3.1-6.3.6 and 2.10.3). Preferably the nucleic acid molecule that hybridizes to the nucleotide sequence of (a) and (b), above, is one that comprises the complement of a nucleic acid molecule of the region s or r or a complementary sequence or RNA product thereof. In a preferred embodiment, nucleic acid molecules comprising the nucleotide sequences of (a) and (b), comprises nucleic acid molecule of RAI or a complementary sequence or RNA product thereof.

Among the nucleic acid molecules of the invention are deoxyoligonucleotides ("oligos") which hybridize under highly stringent or stringent conditions to the nucleic acid molecules described above. In general, for probes between 14 and 70 nucleotides in length the melting temperature (T_m) is calculated using the formula:

$$T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log [\text{monovalent cations (molar)}]) + 0.41(\% \text{ G+C}) - (500/N)$$

where N is the length of the probe. If the hybridization is carried out in a solution containing formamide, the melting temperature is calculated using the equation $T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log [\text{monovalent cations (molar)}]) + 0.41(\% \text{ G+C}) - (0.61\% \text{ formamide}) - (500/N)$ where N is the length of the probe. In general, hybridization is carried out at about 20-25 degrees below T_m (for DNA-DNA hybrids) or 10-15 degrees below T_m (for RNA-DNA hybrids).

Exemplary highly stringent conditions may refer, e.g., to washing in 6x SSC/0.05% sodium pyrophosphate at 37°C (for about 14-base oligos), 48°C (for about 17-base oligos), 55°C (for about 20-base oligos), and 60°C (for about 23-base oligos).

Accordingly, the invention further provides nucleotide primers or probes which detect the r region polymorphisms of the invention. The assessment may be conducted by means of at least one nucleic acid primer or probe, such as a primer or probe of DNA, RNA or a nucleic acid analogue such as peptide nucleic acid (PNA) or locked nucleic acid (LNA). The nucleotide primer or probe is preferably capable of hybridizing to a subsequence of the region corresponding to SEQ ID NO: 1, or a part thereof, or a region complementary to SEQ ID NO: 1.

According to one aspect of the present invention there is provided an allele-specific oligonucleotide probe capable of detecting a r region polymorphism at one or more of positions in the r region as defined by the positions in SEQ ID NO: 1.

- 5 The allele-specific oligonucleotide probe is preferably 5-50 nucleotides, more preferably about 5-35 nucleotides, more preferably about 5-30 nucleotides, more preferably at least 9 nucleotides.

10 The design of such probes will be apparent to the molecular biologist of ordinary skill. Such probes are of any convenient length such as up to 50 bases, up to 40 bases, more conveniently up to 30 bases in length, such as for example 8-25 or 8-15 bases in length. In general such probes will comprise base sequences entirely complementary to the corresponding wild type or variant locus in the region. However, if required one or more mismatches may be introduced, provided that the discriminatory power of the oligonucleotide probe is not unduly affected. The probes of
15 the invention may carry one or more labels to facilitate detection.

In one embodiment, the primers and/or probes are capable of hybridizing to and/or amplifying a subsequence hybridizing to a single nucleotide polymorphism containing the sequence shown herein selected from the group of subsequences below or a
20 sequence complementary thereto, wherein the polymorphism is denoted as for example T/C:

1. gctgcagtga gctgt (-/ACACCTGTGGTCCCAGCTACTCTGG
AAGCTGAGGTGGGAGGATCGCTTGAGCCCAAGAGGTGGAGGCTG-
25 CAGTGAGCTGT) gactgtgccca ctgcactcca (SEQ ID NO: 33)
2. TGACAGTAGA CATCCTGTCA T (A/G) ATAAGTCttt tttttt (SEQ ID NO: 34)
3. CATCCCCATA CCAAcccacc (c/t) tactgctctg atctcctct (SEQ ID NO: 35)
4. ttttagtagagacatggtccgccca (C/T) gttgccaggtggtcttgaactc (SEQ ID NO: 21)
5. ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG-
30 CATGGTGGTGGGGTGC (SEQ ID NO: 23)
6. AAAAACTAAAGTGGGGTTTGCGGG (G/T)
AGTGGGAGGGCCCTTCCTGCTAGG (SEQ ID NO: 24)
7. ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac (SEQ ID NO: 25)
8. TGTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTTTGAG

9. CTCGGGAGGCTGAGGCAGGAGAATC (A/G) CTTGAACTCAGGCA-
GAGGTTG
10. ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA (SEQ ID
NO: 36)
- 5 11. ttgggagacc aaggcaggtg gatc (a/t) ttgaggtca gtagatcaaa (SEQ ID NO: 37)
12. gggaggctcg aggcgggc (A/G) gattgcatga gtcaggatt (SEQ ID NO: 38)
13. tgcagtgagc tgagatcgc (A/G) ccactgcact ccagcctggg (SEQ ID NO: 28)
14. CTTGCTACAGAATTACAGGCA (T/C) GCGCCACCGCTCCGGGCTAA
(SEQ ID NO: 45)
- 10 15. CTAAAGACTACA (-/A) TTTCCCAGCATCCCATTG (SEQ ID NO: 46)
16. CCCTCCCTGCTTGCTTGCTTTCTCT [C/T]
TCTCTCTTTCTTTCTTTCTTTCTT
17. CCCTGCTTGCTTGCTTTCTCTCTCT [C/T]
TCTTTCTTTCTTTCTTTCTTTCTT (SEQ ID NO: 40)
- 15 18. AGAACCTGTTTCAGGCTGGCGGCTCA [C/T] TTGGAT-
GAACAGGGAGTGTGTGAC (SEQ ID NO: 41)
19. CCCCTTCTTAGGACGCATGGGGGT [G/T] GAGAGAACGGGGAGATA-
GACAGAG (SEQ ID NO: 42)
- 20 20. TGCGAGCAGCCCGGGCTACAGGGTT [A/G]
CCTGAGGTGTGGGTCCCAGGATGG (SEQ ID NO: 43)
21. GGCGCCTCAACAGCCAGAAGGAGCG [A/G] AGCCTCAGGCCAGG-
CAGCTCTGG (SEQ ID NO: 44)

25 In another embodiment of the methods of the invention the primers and/or probes
are capable of hybridizing to and/or amplifying a subsequence hybridizing to a single
nucleotide polymorphism containing the sequence shown herein selected from the
group of subsequences below or a sequence complementary thereto, wherein the
polymorphism is denoted as for example T/C:

- 30 1. TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTTTGAG
2. CTCGGGAGGCTGAGGCAGGAGAATC (A/G) CTTGAACTCAGGCA-
GAGGTTG
3. ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA (SEQ ID
NO: 36)
- 35 4. ttgggagacc aaggcaggtg gatc (a/t) ttgaggtca gtagatcaaa (SEQ ID NO: 37)

5. gggaggctcg aggcgggc (A/G) gattgcatga gctcaggatt (SEQ ID NO: 38)
6. tgcagtgagc tgagatcgc (A/G) ccactgcact ccagcctggg (SEQ ID NO: 28)
7. CTTGCTACAGAATTACAGGCA (T/C) GCGCCACCGCTCCGGGCTAA
(SEQ ID NO: 45)
- 5 8. CTAAAGACTACA (-/A) TTTCCCAGCATCCCATTG (SEQ ID NO: 46)

In a preferred embodiment of the methods of the invention the the primers and/or probes are capable of hybridizing to and/or amplifying a subsequence hybridizing to a single nucleotide polymorphism containing the sequence shown herein selected from the group of subsequences below or a sequence complementary thereto, wherein the polymorphism is denoted as for example T/C:

1. gctgcagtga gctgt (-/ACACCTGTGGTCCCAGCTACTCTGG
AAGCTGAGGTGGGAGGATCGCTTGAGCCCAAGAGGTGGAGGCTG-
CAGTGAGCTGT) gactgtgccca ctgcactcca (SEQ ID NO: 33)
2. TGACAGTAGA CATCCTGTCA T (A/G) ATAAGTCttt tttttt (SEQ ID NO: 34)
3. CATCCCCATA CCAAcccacc (c/t) tactgtctg atctcctct (SEQ ID NO: 35)
4. ttttagtagagacatggtccgccca (C/T) gttgccaggtggtcttgaactc (SEQ ID NO: 21)
5. ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG-
CATGGTGGTGGGGTGC (SEQ ID NO: 23)
6. AAAAACTAAAGTGGGGTTTGCGGG (G/T)
AGTGGGAGGGCCCTTCCTGCTAGG (SEQ ID NO: 24)
7. ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac (SEQ ID NO: 25)
8. TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTTTGAG
9. CTCGGGAGGCTGAGGCAGGAGAATC (A/G) CTTGAACTCAGGCA-
GAGGTTG
10. ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA (SEQ ID
NO: 36)
11. ttgggagacc aaggcaggtg gatc (a/t) ttgaggtca gtagatcaaa (SEQ ID NO: 37)
12. gggaggctcg aggcgggc (A/G) gattgcatga gctcaggatt (SEQ ID NO: 38)
13. tgcagtgagc tgagatcgc (A/G) ccactgcact ccagcctggg (SEQ ID NO: 28)
14. CTTGCTACAGAATTACAGGCA (T/C) GCGCCACCGCTCCGGGCTAA
(SEQ ID NO: 45)
15. CTAAAGACTACA (-/A) TTTCCCAGCATCCCATTG (SEQ ID NO: 46)

In a preferred embodiment, the primers and/or probes are capable of hybridizing to a subsequence selected from the group of subsequences below:

1. gctgcagtga gctgt (-/ACACCTGTGGTCCCAGCTACTCTGG
5 AAGCTGAGGTGGGAGGATCGCTTGAGCCCAAGAGGTGGAGGCTG-
CAGTGAGCTGT) gactgtgccca ctgcactcca (SEQ ID NO: 33)
2. TGACAGTAGA CATCCTGTCA T (A/G) ATAAGTCttt tttttt (SEQ ID NO: 34)
3. CATCCCCATA CCAAcccacc (c/t) tactgtctgt atctcctcct (SEQ ID NO: 35)
4. ttttagtagagacatggttccgccca (C/T) gttgccaggtggtcttgaactc (SEQ ID NO: 21)
- 10 5. ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG-
CATGGTGGTGGGGTGC (SEQ ID NO: 23)
6. AAAAACTAAAGTGGGGTTTGCGGG (G/T)
AGTGGGAGGGCCCTTCCTGCTAGG (SEQ ID NO: 24)
7. ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac (SEQ ID NO: 25)
- 15 8. TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTTTGAG
9. CTCGGGAGGCTGAGGCAGGAGAATC (A/G) CTTGAACTCAGGCA-
GAGGTTG
10. ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA (SEQ ID
NO: 36)
- 20 11. ttgggagacc aaggcaggtg gatc (a/t) ttgaggtca gtagatcaaa (SEQ ID NO: 37)
12. gggaggctcg aggcgggc (A/G) gattgcatga gctcaggatt (SEQ ID NO: 38)

In an even more preferred embodiment, the primers and/or probes are capable of hybridizing to a subsequence selected from the group of subsequences below:

- 25 1. ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG-
CATGGTGGTGGGGTGC (SEQ ID NO: 23)
2. AAAAACTAAAGTGGGGTTTGCGGG (G/T)
AGTGGGAGGGCCCTTCCTGCTAGG (SEQ ID NO: 24)
- 30 3. ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac (SEQ ID NO: 25)
4. TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTTTGAG
5. CTCGGGAGGCTGAGGCAGGAGAATC (A/G) CTTGAACTCAGGCA-
GAGGTTG
6. ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA (SEQ ID
35 NO: 36)

7. ttgggagacc aaggcaggtg gatc (a/t) ttgaggtca gtagatcaaa (SEQ ID NO: 37)
8. gggaggctcg aggcgggc (A/G) gattgcatga gctcaggatt (SEQ ID NO: 38)

Most preferred are the primers and/or probes capable of hybridizing to a subsequence selected from the group of subsequences below:

1. ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG-CATGGTGGTGGGGTGC (SEQ ID NO: 23)
2. ATTAAGTGCCTTCACACAGC (A/T) CTGGTTTAAT GTTTATAA (SEQ ID NO: 36)
3. gggaggctcg aggcgggc (A/G) gattgcatga gctcaggatt (SEQ ID NO: 38)

In a second embodiment according to the methods of the invention are the primers and/or probes capable of hybridizing to a subsequence selected from the group of subsequences below:

1. gctgcagtga gctgt (-/ACACCTGTGGTCCCAGCTACTCTGG AAGCTGAGGTGGGAGGATCGCTTGAGCCCAAGAGGTGGAGGCTG-CAGTGAGCTGT) gactgtgcc ctgcactcca (SEQ ID NO: 33)
2. CATCCCCATA CCAAcccacc (c/t) tactgctctg atctcctct (SEQ ID NO: 35)
3. ttttagtagagacatggtccgcca (C/T) gttgccaggtggtttgaactc (SEQ ID NO: 21)
4. ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG-CATGGTGGTGGGGTGC (SEQ ID NO: 23)
5. AAAAACTAAAGTGGGGTTTGCGGG (G/T) AGTGGGAGGGCCCTTCCTGCTAGG (SEQ ID NO: 24)
6. ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac (SEQ ID NO: 25)
7. TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTTTGAG (SEQ ID NO: 26)
8. CTCGGGAGGCTGAGGCAGGAGAATC (A/G) CTTGAACTCAGGCA-GAGGTTG (SEQ ID NO: 22)
9. ttgggagacc aaggcaggtg gatc (a/t) ttgaggtca gtagatcaaa (SEQ ID NO: 37)
10. CTTGCTACAGAATTACAGGCA (T/C) GCGCCACCGCTCCGGGCTAA (SEQ ID NO: 45)
11. CTAAAGACTACA (-/A) TTTCCCAGCATCCCATTG (SEQ ID NO: 46)

In a preferred embodiment, the primers and/or probes are capable of hybridizing to a subsequence selected from the group of subsequences below:

1. CATCCCCATA CCAAcccacc (c/t) tactgctctg atctcctcct (SEQ ID NO: 35)
- 5 2. ttttagtagagacatggttccgcca (C/T) gttgccccaggctggtcttgaactc (SEQ ID NO: 21)
3. ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG-
CATGGTGGTGGGGTGC (SEQ ID NO: 23)
4. AAAAACTAAAGTGGGGTTTGCGGG (G/T)
AGTGGGAGGGCCCTTCCTGCTAGG (SEQ ID NO: 24)
- 10 5. ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac (SEQ ID NO: 25)
6. TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTTTGAG
7. CTCGGGAGGCTGAGGCAGGAGAATC (A/G) CTTGAACTCAGGCA-
GAGGTTG
8. ttgggagacc aaggcaggtg gatac (a/t) ttgaggtca gtagatcaaa (SEQ ID NO: 37)
- 15

In an even more preferred embodiment the primers and/or probes are capable of hybridizing to a subsequence selected from the group of subsequences below:

1. CATCCCCATA CCAAcccacc (c/t) tactgctctg atctcctcct (SEQ ID NO: 35)
- 20 2. ttttagtagagacatggttccgcca (C/T) gttgccccaggctggtcttgaactc (SEQ ID NO: 21)
3. ACTAAAAATAAAAAATAAAAAAA(-/AA) ATAGCCGAG-
CATGGTGGTGGGGTGC (SEQ ID NO: 23)
4. AAAAACTAAAGTGGGGTTTGCGGG (G/T)
AGTGGGAGGGCCCTTCCTGCTAGG (SEQ ID NO: 24)
- 25 5. ggatcacaag gtcaggagtt (c/t) gagaccagcc tggccaacac (SEQ ID NO: 25)
6. TGTTGTCCAA GCTGGCAGAG (A/G) TTTTGTGTTG TTTGTTTGAG (SEQ
ID NO: 26)
7. CTCGGGAGGCTGAGGCAGGAGAATC (A/G) CTTGAACTCAGGCA-
GAGGTTG (SEQ ID NO: 22)

30

Even more preferred are the primers and/or probes are capable of hybridizing to a subsequence selected from the group of subsequences below:

1. CTTGCTACAGAATTACAGGCA (T/C) GCGCCACCGCTCCGGGCTAA
35 (SEQ ID NO: 45)

2. CTAAAGACTACA (-/A) TTTCCCAGCATCCCATTG (SEQ ID NO: 46)
3. ttgggagacc aaggcagggtg gatc (a/t) ttgaggtca gtagatcaaa (SEQ ID NO: 37)

In one embodiment the primers or probes able to detect subsequences as described
5 above are selected from one or more of the following:

1. agt cac agc tca ctg cag cct c (SEQ ID NO: 47)
2. acc tct tgg gct caa gcg atc ctc (SEQ ID NO: 48)
3. aaa aaa aga ctt atc atg aca gga tgt ct (SEQ ID NO: 49)
- 10 4. gca aga ctc cgt ccc aga aaa aga aaa (SEQ ID NO: 50)
5. tcctctctctccccagctcattttg (SEQ ID NO: 51)
6. ACCCCACCCTACTGCTCTGATCTC (SEQ ID NO: 52)
7. agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)
8. ggt tcc gcc acg ttg cc (SEQ ID NO: 54)
- 15 9. tgc gct att ttt ttt ttt att ttt tta tt (SEQ ID NO: 55)
10. att aca ggc acc cac cac cat g (SEQ ID NO: 56)
11. acc cca ctt tag ttt ttt ttt cct cta gtg atc gcc (SEQ ID NO: 57)
12. gcc ctc cca cta ccc gca (SEQ ID NO: 58)
13. aga cgg ggt ttc act gtg ttg gc (SEQ ID NO: 59)
- 20 14. agg ctg gtc tca aac tcc tga c (SEQ ID NO: 60)
15. CAAACAAACAAAAACCTCTGCCA (SEQ ID NO: 61)
16. CTTGGACAACATAGGGAGACCCTGTGT (SEQ ID NO: 62)
17. tgc ctc agc ctc ccg agt agc t (SEQ ID NO: 63)
18. cct cct gag ttc aag cga ttc tc (SEQ ID NO: 64)
- 25 19. tgc ctt cac aca gct ctg gtt taa tg (SEQ ID NO: 65)
20. tgc ctt cac aca gca ctg gtt taa tg (SEQ ID NO: 66)
21. ACCATGTTGGCCAGGCTGGTTTT (SEQ ID NO: 67)
22. ATCTACTGACCTCAAATGATCCACCT (SEQ ID NO: 68)
23. tgc aat ccg ccc gcc (SEQ ID NO: 69)
- 30 24. cca ggc tgg ttt gga aat cct gag ctc (SEQ ID NO: 70)
25. ctg aga tgc cac cac tgc ac (SEQ ID NO: 71)
26. ggg agg cgg agc ttg cag tga (SEQ ID NO: 72)
27. gcg cat gcc tgt aat tct gta (SEQ ID NO: 73)
28. cag gac gag cca cag aca aaa ctc c (SEQ ID NO: 74)
- 35 29. tgc aat gag gct cct ggc c (SEQ ID NO: 75)

30. act aca att tcc cag cat ccc a (SEQ ID NO: 76)
 31. cct ccc tcc ctc cct gc (SEQ ID NO: 77)
 32. tgc ttg ctt tct tct tct (SEQ ID NO: 78)
 33. tcc ctg ctt gct tgc ttt ctc t (SEQ ID NO: 79)
 5 34. tct ctc ttt ctt tct ttc ttt c (SEQ ID NO: 80)
 35. tgt tca tcc aaa tga gcc gc (SEQ ID NO: 81)
 36. agc ctg aac agg ttc tgt tcc ttc gac tt (SEQ ID NO: 82)
 37. caa gct gct atc tgc acc gat ctt (SEQ ID NO: 83)
 38. ggg tga cca ccc tgc cag cc (SEQ ID NO: 84)
 10 39. cgg gct aca ggg tta cct gag (SEQ ID NO: 85)
 40. tct gca acc tgg tgc gag cag c (SEQ ID NO: 86)
 41. tga ggc tcc gct cct tct gg (SEQ ID NO: 87)
 42. agc tgc cag agc tgc ctg ggc (SEQ ID NO: 88)

- 15 One or more primers able to detect subsequences by hybridisation as described may in a particularly preferred embodiment of the method of the invention be selected from

1. tgc gct att ttt ttt ttt att ttt tta tt (SEQ ID NO: 55)
 20 2. att aca ggc acc cac cac cat g (SEQ ID NO: 56)
 3. tgc ctt cac aca gct ctg gtt taa tg (SEQ ID NO: 65)
 4. tgc ctt cac aca gca ctg gtt taa tg (SEQ ID NO: 66)
 5. tgc aat ccg ccc gcc (SEQ ID NO: 69)
 6. cca ggc tgg ttt gga aat cct gag ctc (SEQ ID NO: 70)

25

According to another aspect of the present invention there is provided a diagnostic nucleic acid primer capable of detecting a r region polymorphism at one or more of positions in the r region as defined by the in SEQ ID NO: 1.

30

The primer or probe may be a diagnostic nucleic acid primer defined as an allele specific primer, used, generally together with a constant primer, in an amplification reaction such as a PCR reaction, which provides the discrimination between alleles through selective amplification of one allele at a particular sequence position. The diagnostic primer is preferably 5-50 nucleotides, more preferably about 5-35 nucleo-

tides, more preferably about 5-30 nucleotides, more preferably at least 9 nucleotides.

5 In accordance with the present invention diagnostic primers are provided, comprising the sequences set out below as well as derivatives thereof wherein about 6-8 of the nucleotides at the 3' terminus are identical to the sequences given below and wherein up to 10, such as up to 8, 6, 4, 2, or 1 of the remaining nucleotides may be varied without significantly affecting the properties of the diagnostic primer. Conveniently, the sequence of the diagnostic primer is as written below.

10

Furthermore, as described above at least two sets of primer(s) and/or probe(s), such as at least three sets of primer(s) and/or probe(s), for example four sets of primer(s) and/or probes may be combined in the method thereby increasing the correlation probability. This second or other set of primer(s) and/or probe(s) may be a nucleotide or nucleotide analogues hybridising to a region within the region r or to a sequence different from the region r. Said sequence different from the region r is preferably a region in chromosome 19, preferably in chromosome 19q. In particular such second or other primer or probe may be selected from one or more of the sequences below, or the complementary strands:

20

1. tgc ctc acc cct gta atc c (SEQ ID NO: 90)
2. gct tgt aat ccc agc tac tcg (SEQ ID NO: 91 and SEQ ID NO: 89)
3. caa cac tca cac ccc aca g (SEQ ID NO: 92)
4. aga tca cgc cac tgc act c (SEQ ID NO: 93)
- 25 5. ttg aca att gag caa aga gc (SEQ ID NO: 94)
6. ttg gat tac aga cgt gag c (SEQ ID NO: 95)
7. agt gca gcc tca act tcc (SEQ ID NO: 96)
8. cca gtc caa aca ata tga tcc (SEQ ID NO: 97)
9. cat gat tca ctg cac cca acc (SEQ ID NO: 98)
- 30 10. ttt cac tct tgt tgc cca agc (SEQ ID NO: 99)
11. cac cat gcc tgg ctc caa tgt (SEQ ID NO: 100)
12. agc cca gga att caa gg (SEQ ID NO: 101)
13. aga aca ttg gag cca gg (SEQ ID NO: 102)
14. aga atc act gga atc cag g (SEQ ID NO: 103)
- 35 15. ttt tca cac agg tcc aat cc (SEQ ID NO: 104)

16. act gca acc tcc atc tcc (SEQ ID NO: 105)
17. caa tta agt gcc ttc aca cag ca (SEQ ID NO: 106)
18. ggt caa gag ttc aag acc agc (SEQ ID NO: 107)
19. ccc tgc ccc acc tct cc (SEQ ID NO: 108)
5 20. agt caa ttt ctg tgc aaa cta ctt tta ttt (SEQ ID NO: 109)
21. gag gca aca gga aca aac c (SEQ ID NO: 110)
22. cat tgg aat gag cag aaa cc (SEQ ID NO: 111)
23. taa cat aaa gaa tca gga gga ggc (SEQ ID NO: 112)
24. agt tgg ctc atc tgc ctc tt (SEQ ID NO: 113)
10 25. tgg cta aca cgg tga aac c (SEQ ID NO: 114)
26. gga atc caa aga ttc tat gat gg (SEQ ID NO: 115)
27. act cct gac ttc aaa tga tcc (SEQ ID NO: 116)
28. tag ccc cca gtc acg ttc c (SEQ ID NO: 117)
29. aga agt cca aga gtt tgc agc (SEQ ID NO: 118)
15 30. ttc tca gtc cca gaa tga acc (SEQ ID NO: 119)
31. cca ctt agg taa aca cct ctt (SEQ ID NO: 120)
32. ctg caa tga gcc gag ata gaa (SEQ ID NO: 121)
33. cca ctt agg taa aca cct ctt (SEQ ID NO: 122)
34. ctg caa tga gcc gag ata gaa (SEQ ID NO: 123)
20 35. atg ttg ggg aga ctg agg (SEQ ID NO: 124)
36. ccg cat cta act tat tct gg (SEQ ID NO: 125)
37. aac tac ctc tgc aaa ccc agc (SEQ ID NO: 126)
38. ttg gaa tgg agg gat tct acc (SEQ ID NO: 127)
39. ggt ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
25 40. cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
41. gga cag atg gca atg atg g (SEQ ID NO: 130)
42. tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)

- However, in another embodiment a primer or probe for use in the methods of the
30 present invention may be selected from the group of nucleotides consisting of
agt gca gcc tca act tcc (SEQ ID NO: 96)
cca gtc caa aca ata tga tcc (SEQ ID NO: 97)
AAA AAA ATA GCC GAG CAT GG (SEQ ID NO: 134)
TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)
35 cat gat tca ctg cac cca acc (SEQ ID NO: 98)

- ttt cac tct tgt tgc cca agc (SEQ ID NO: 99)
ttt tca cac agg tcc aat cc (SEQ ID NO: 104)
act gca acc tcc atc tcc (SEQ ID NO: 105)
atg ttg ggg aga ctg agg (SEQ ID NO: 124)
5 ccg cat cta act tat tct gg (SEQ ID NO: 125)
aac tac ctc tgc aaa ccc agc (SEQ ID NO: 126)
ttg gaa tgg agg gat tct acc (SEQ ID NO: 127)
ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
10 gga cag atg gca atg atg g (SEQ ID NO: 130)
tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)
acc atg gcg cct caa ca (SEQ ID NO: 136)
gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:137)
- 15 In yet another embodiment of the present invention the primer or probe is selected
from the group of nucleotides consisting of
agt gca gcc tca act tcc (SEQ ID NO: 96)
cca gtc caa aca ata tga tcc (SEQ ID NO: 97)
AAA AAA ATA GCC GAG CAT GG(SEQ ID NO: 134)
20 TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)
cat gat tca ctg cac cca acc (SEQ ID NO: 98)
ttt cac tct tgt tgc cca agc (SEQ ID NO: 99)
ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
25 gga cag atg gca atg atg g (SEQ ID NO: 130)
tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)
acc atg gcg cct caa ca (SEQ ID NO: 136)
gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:137)
- 30 In a further embodiment the primer or probe for use in a methods described herein
is selected from the group of nucleotides consisting of
agt gca gcc tca act tcc (SEQ ID NO: 96)
cca gtc caa aca ata tga tcc (SEQ ID NO: 97)
AAA AAA ATA GCC GAG CAT GG(SEQ ID NO: 134)
35 TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)

cat gat tca ctg cac cca acc (SEQ ID NO: 98)

ttt cac tct tgt tgc cca agc (SEQ ID NO: 99)

ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

5 gga cag atg gca atg atg g (SEQ ID NO: 130)

tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)

acc atg gcg cct caa ca (SEQ ID NO: 136)

gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:137)

In yet a further embodiment the primer or probe is selected from the group of nu-

10 cleotides consisting of

agt gca gcc tca act tcc (SEQ ID NO: 96)

cca gtc caa aca ata tga tcc (SEQ ID NO: 97)

AAA AAA ATA GCC GAG CAT GG(SEQ ID NO: 134)

TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)

15 ggt ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

acc atg gcg cct caa ca (SEQ ID NO: 136)

gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:137)

In a particular embodiment of the present invention the primer or probe is selected

20 from the group of nucleotides consisting of

AAA AAA ATA GCC GAG CAT GG(SEQ ID NO: 134)

TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)

ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

25

In yet another embodiment the primer or probe is selected from the group of nucleotides consisting of

AAA AAA ATA GCC GAG CAT GG(SEQ ID NO: 134)

TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)

30 acc atg gcg cct caa ca (SEQ ID NO: 136)

gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:137)

However, the primer or probe may be selected from the group of nucleotides consisting of

35 AAA AAA ATA GCC GAG CAT GG(SEQ ID NO: 134)

- TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)
ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
acc atg gcg cct caa ca (SEQ ID NO: 136)
5 gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:137)

The primer or probe in one embodiment of the present invention includes the use of a primer or probe being selected from the group consisting of

- agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)
10 ggt tcc gcc acg ttg cc (SEQ ID NO: 54)
tcg gct att ttt ttt att ttt tta tt (SEQ ID NO: 55)
att aca ggc acc cac cac cat g (SEQ ID NO: 56)

15 cct gga caa cat agg gag acc ctg tgt (SEQ ID NO: 138)
caa aca aac aaa aac ctc tgc ca (SEQ ID NO:139)
atg ttg ggg aga ctg agg (SEQ ID NO: 124)
ccg cat cta act tat tct gg (SEQ ID NO: 125)
aac tac ctc tgc aaa ccc agc (SEQ ID NO: 126)
20 ttg gaa tgg agg gat tct acc (SEQ ID NO: 127)
ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
gga cag atg gca atg atg g (SEQ ID NO: 130)
tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)
25 acc atg gcg cct caa ca (SEQ ID NO: 140)
gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:141)

The primer or probe may, however, be selected from the group consisting of

- agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)
30 ggt tcc gcc acg ttg cc (SEQ ID NO: 54)
tcg gct att ttt ttt att ttt tta tt (SEQ ID NO: 55)
att aca ggc acc cac cac cat g (SEQ ID NO: 56)
ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
35 acc atg gcg cct caa ca (SEQ ID NO: 140)

gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:141)

Or the primer or probe is selected from the group consisting

tgc gct att ttt ttt ttt att ttt tta tt (SEQ ID NO: 55)

5 att aca ggc acc cac cac cat g (SEQ ID NO: 56)

ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

gga cag atg gca atg atg g (SEQ ID NO: 130)

tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)

10 acc atg gcg cct caa ca (SEQ ID NO: 140)

gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:141)

In another embodiment the primer or probe is selected from the group consisting of

agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)

15 ggt tcc gcc acg ttg cc (SEQ ID NO: 54)

ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

acc atg gcg cct caa ca (SEQ ID NO: 140)

gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:141)

20

In another embodiment of the present invention the primer or probe used in the methods described herein is selected from the group consisting of

agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)

ggg tcc gcc acg ttg cc (SEQ ID NO: 54)

25 acc atg gcg cct caa ca (SEQ ID NO: 140)

gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:141)

In a further embodiment the primer or probe is selected from the group consisting of

agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)

30 ggt tcc gcc acg ttg cc (SEQ ID NO: 54)

ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

It is understood that the primers may be selected individually or for example in pairs. However, it is also within the scope of the present application to combine any one of the primers in any combination.

5 In one embodiment the polymorphism observed at position 18147012 in config NT_011109, RAI3'd1, is detected using one fluorescent primer and one non-fluorescent in. The primers are: RAI3'd1-f1: 5'-AAA AAA ATA GCC GAG CAT GG (SEQ ID NO: 134) and RAI3'd1-r1: 5'-6FAM-TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135).

10

The primers and probes may be manufactured using any convenient method of synthesis. Examples of such methods may be found in standard textbooks, for example "Protocols for Oligonucleotides and Analogues; Synthesis and Properties," Methods in Molecular Biology Series; Volume 20; Ed. Sudhir Agrawal, Humana ISBN: 0-15 89603-247-7; 1993; 1.sup.st Edition. If required the primer(s) and probe(s) may be labelled to facilitate detection.

Kit

20 According to another aspect of the present invention, there is provided a diagnostic kit comprising at least one diagnostic primer of the invention and/or at least one allele-specific oligonucleotide primer of the invention.

The diagnostic kits may comprise appropriate packaging and instructions for use in the methods of the invention. Such kits may further comprise appropriate buffer(s) and polymerase(s) such as thermostable polymerases, for example taq polymerase.

25

Preferred kits can comprise means for amplifying the relevant sequence such as primers, polymerase, deoxynucleotides, buffer, metal ions; and/or means for discriminating the polymorphism, such as one or a set of probes hybridising to the polymorphic site, a sequence reaction covering the polymorphic site, an enzyme or an antibody; and/or a secondary amplification system, such as enzyme-conjugated antibodies, or fluorescent antibodies. The kit-of-parts preferably also comprises a detection system, such as a fluorometer, a film, an enzyme reagent or another highly sensitive detection device.

35

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits. The invention therefore also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (i.e., a test sample). Such kits can be used, e.g., to determine if a subject is suffering from or is at increased risk of developing a disorder associated with a disorder-causing allele, or aberrant expression or activity of a polypeptide of the invention. For example, the kit can comprise a labeled compound or agent capable of detecting the polypeptide or mRNA or DNA or RAI gene sequences, e.g., encoding the polypeptide in a biological sample. The kit can further comprise a means for determining the amount of the polypeptide or mRNA in the sample (e.g., an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level, or if the DNA correlates with presence of an RAI allele that causes a disorder.

One embodiment of the present invention is a diagnostic kit comprising at least one diagnostic primer of the invention directed against the RAI3'd1 polymorphic marker and/or at least one allele-specific oligonucleotide primer of the invention. However, the kit may comprise for example two diagnostic primers, such as three, for example four, such as five, for example six, such as seven, for example eight diagnostic primers and/or probes as described herein. For example the kit may further comprise at least one diagnostic primer of the invention directed against the ASE 1 exon 1 polymorphism. However, the kit may also comprise at least one diagnostic primer of the invention directed against the ASE 1 exon 3-2. In one embodiment the at least one diagnostic primer and/ or at least one allele-specific oligonucleotide primer is the RAI3'd1 -f2 primer, such as the RAI3' d1-r2 primer or for example SEQ ID NO: 134, such as SEQ ID NO: 135. However, it is appreciated that the present invention is not limited to the four above named primers, but that any primer capable of hybridizing or capable of amplifying the RAI3'd1 polymorphic marker is within the scope of the present application. Similarly, primers or probes directed against the ASE 1 exon 1 and/or ASE 1 exon 3-2 as described in the present invention may be used. However, it is understood that any primer or probe capable of hybridizing or capable of ampli-

fyng ASE 1 exon 1 and/or ASE 1 exon 3-2 is within the scope of the present invention.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or to the first antibody and is conjugated to a detectable agent.

Identification of an allele as having implication for risk of cancer

An allele in the r region can be identified as correlated with an increased risk of developing cancer on the basis of statistical analyses of the incidence of a particular allele in two groups of individuals with and without cancer, respectively, according to the χ^2 test, which is well known in the art. Furthermore, an allele in the region can be identified as an allele correlated with prognosis of cancer on the basis of statistical analyses of the incidence of a particular allele in individuals demonstrating different prognostic characteristics.

Identification of humans having increased likelihood of responding to treatment

It is further contemplated that the present invention provides a method for identifying a human subject as having an increased likelihood of responding positively to a cancer treatment, comprising determining the presence in the subject of a s or r region allele genotype correlated with an increased likelihood of positive response to treatment, whereby the presence of the genotype identifies the subject as having an increased likelihood of responding to cancer treatment.

The treatment mentioned herein may be any cancer treatment, such as conventional cancer treatment, for example X-ray, chemotherapeutics, surgical excision or combinations thereof.

Protein Products of the Gene(s)

Gene products of the region r or peptide fragments thereof, can be prepared for a variety of uses. For example, such gene products, or peptide fragments thereof, can be used for the generation of antibodies, in diagnostic assays.

The gene products of the invention include, but are not limited to, human RAI gene products. In the following the invention is described in relation to RAI gene product.

Gene product, sometimes referred to herein as an "protein" or "polypeptide", includes those gene products encoded by the RAI gene sequences shown as position 7760-22885 in SEQ ID NO: 1. Among gene product variants are gene products comprising amino acid residues encoded by the polymorphisms. Such gene product variants also include a variant of the RAI gene product.

In addition, RAI gene products may include proteins that represent functionally equivalent gene products. In preferred embodiments, such functionally equivalent RAI gene products are naturally occurring gene products. Functionally equivalent RAI gene products also include gene products that retain at least one of the biological activities of the RAI gene products described above, and/or which are recognized by and bind to antibodies (polyclonal or monoclonal) directed against RAI gene products.

Antibodies to Gene Products

Described herein are methods for the production of antibodies capable of specifically recognizing one or more gene product epitopes or epitopes of conserved variants or peptide fragments of the gene products. Furthermore, antibodies that specifically recognize mutant forms are encompassed by the invention. The terms "specifically bind" and "specifically recognize" refer to antibodies that bind to RAI gene product epitopes at a higher affinity than they bind to non-RAI (e.g., random) epitopes.

Such antibodies may include, but are not limited to, polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')₂ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above, including the polyclonal and monoclonal antibodies described below. Such antibodies may be used, for example, in the detection of a gene product in a biological sample and may, therefore, be utilized as part of a diagnostic or prognostic technique whereby patients may be tested for abnormal levels of gene products, and/or for the presence of abnormal forms of such gene products. Such antibodies may

also be utilized in conjunction with, for example, compound screening schemes, as described, below, for the evaluation of the effect of test compounds on gene product levels and/or activity.

5 For the production of antibodies against a gene product, various host animals may be immunized by injection with a RAI gene product, or a portion thereof. Such host animals may include, but are not limited to rabbits, mice, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and in-
10 complete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and Corynebacterium parvum.

15 Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as a gene product, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals such as those described above, may be immunized by injection with gene product supplemented with adjuvants as also described above.

20 Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein (1975, Nature 256:495-
25 497; and U.S. Pat. No. 4,376,110), the human B-cell hybridoma technique (Kosbor et al., 1983, Immunology Today 4:72; Cole et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030), and the EBV-hybridoma technique (Cole et al., 1985, Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass
30 thereof. The hybridoma producing the mAb of this invention may be cultivated in vitro or in vivo. Production of high titers of mAbs in vivo makes this the presently preferred method of production.

35 In addition, techniques developed for the production of "chimeric antibodies" (Morrison, et al., 1984, Proc. Natl. Acad. Sci., 81:6851-6855; Neuberger, et al., 1984, Na-

ture 312:604-608; Takeda, et al., 1985, Nature, 314:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from
5 different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, e.g., Cabilly et al., U.S. Pat. No. 4,816,567; and Boss et al., U.S. Pat. No. 4,816,397, which are incorporated herein by reference in their entirety.)

10 In addition, techniques have been developed for the production of humanized antibodies. (See, e.g., Queen, U.S. Pat. No. 5,585,089, which is incorporated herein by reference in its entirety.) An immunoglobulin light or heavy chain variable region consists of a "framework" region interrupted by three hypervariable regions, referred to as complementarily determining regions (CDRs). The extent of the framework
15 region and CDRs have been precisely defined (see, "Sequences of Proteins of Immunological Interest", Kabat, E. et al., U.S. Department of Health and Human Services (1983)). Briefly, humanized antibodies are antibody molecules from non-human species having one or more CDRs from the non-human species and a framework region from a human immunoglobulin molecule.

20 Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778; Bird, 1988, Science 242:423-426; Huston, et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:5879-5883; and Ward, et al., 1989, Nature 334:544-546) can be adapted to produce single chain antibodies against gene products. Single chain antibodies are formed by linking the heavy and light chain fragments of the
25 Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments that recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')₂ fragments, which can be produced by pepsin digestion of the antibody molecule and the Fab fragments, which can be generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed (Huse, et al., 1989, Science 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Immunoassays for gene products, conserved variants, or peptide fragments thereof will typically comprise incubating a sample, such as a biological fluid, a tissue extract, freshly harvested cells, or lysates of cells in the presence of a detectably labeled antibody capable of identifying gene product, conserved variants or peptide fragments thereof, and detecting the bound antibody by any of a number of techniques well-known in the art.

The biological sample may be brought in contact with and immobilized onto a solid phase support or carrier, such as nitrocellulose, that is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled gene product specific antibody. The solid phase support may then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on the solid support may then be detected by conventional means.

By "solid phase support or carrier" is intended any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

One of the ways in which the RAI gene product-specific antibody can be detectably labeled is by linking the same to an enzyme, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, α -glycero-phosphate dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, β -galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods

that employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

5 Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, by labeling the antibody with a fluorescent compound. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

10

The antibody can also be detectably labeled using fluorescence emitting metals such as ^{152}Eu , or others of the lanthanide series or by coupling it to a chemiluminescent compound.

15 Diseases

Described herein are various applications of gene sequences, gene products, including peptide fragments and fusion proteins thereof, and of antibodies directed against gene products and peptide fragments thereof. Such applications include, for example, prognostic and diagnostic evaluation of a disease, such as cancer, and the
20 identification of subjects with a predisposition to such disorders, as described above.

The method according to the invention may be used in relation to any cancer form, such as, but not limited to, skin carcinoma including malignant melanoma, breast cancer, lung cancer, colon cancer and other cancers in the gastro-intestinal tract,
25 prostate cancer, lymphoma, leukemia, multiple myeloma, pancreas cancer, head and neck cancer, ovary cancer and other gynecological cancers. In particular the method is relevant for skin cancer, lung cancer, colon cancer, multiple myeloma, and breast cancer, such as skin cancer, breast cancer, multiple myeloma, and lung cancer, such as skin cancer, breast cancer and lung cancer, such as skin cancer
30 and breast cancer, preferably wherein the skin cancer is basal cell carcinoma, such as lung cancer. For example the cancer is multiple myeloma, In another embodiment the cancer is breast cancer. The cancer may also be skin cancer, preferably basal cell carcinoma, for example early age basal cell carcinoma.

In particular, the method is relevant for both early age cancer and later age cancer, such as early age breast cancer, and such as later age breast cancer.

5 The method is of particular relevance for lung cancer, such as in patients with XPD exon 23^{AA}. The method is also in particular relevant for early age skin cancer, such as early age basal cell carcinoma.

10 Gene nucleic acid sequences, described above, can be utilized for transferring recombinant nucleic acid sequences to cells and expressing said sequences in recipient cells. Such techniques can be used, for example, in marking cells or for the treatment of cancer. Such treatment can be in the form of gene replacement therapy. Specifically, one or more copies of a normal RAI gene or a portion of the RAI gene that directs the production of an RAI gene product exhibiting normal RAI gene function, may be inserted into the appropriate cells within a patient, using vectors
15 that include, but are not limited to, adenovirus, adeno-associated virus, and retrovirus vectors, in addition to other particles that introduce DNA into cells, such as liposomes.

20 In another embodiment, the invention may be used in relation to inflammatory diseases, such as, but not limited thereto, rheumatoid arthritis, colitis ulcerosa, Crohn's disease, thyroiditis, neural inflammation as in Alzheimer's disease, and Guillain-Barré syndrome.

40 **RAI function**

Presumably, the primary effector influences activity or expression of *RAI* or *XPD*. Possibly, one of the markers analyzed is the effector itself. However, there are several alternative possibilities. First, the effector could modify the sequence of RAI protein, through amino acid substitution, splicing or termination. Secondly, one likely
45 position of the effector is in the 3' portion of the gene and that the effector modifies mRNA expression and stability (22). Thirdly, one could imagine that the effector was a modification of an enhancer situated between *RAI* and *XPD*, again modulating the expression of *RAI*, and possibly modulating other nearby genes as well. Finally, the primary effector may be located in the promoter or first intron of RAI and influence
50 the transcriptional activity of the gene.

If a primary result of the effector is modified *RAI* action, then one result may be modified apoptosis. *RAI* protein is an inhibitor of *RelA*, a subunit of the transcription factor $\text{NF-}\kappa\text{B}$ (23, 24). $\text{NF-}\kappa\text{B}$ has long been implicated in both cell proliferation and apoptosis. Modulation of $\text{NF-}\kappa\text{B}$ may well be part of a “crunch-time scenario”, invoked when the cell has to muster its forces and make life-and-death decisions. A recent scientific paper suggests that the choice between the cell survival and death is regulated by the relative activity of the two subunits encoded by *RelA* and *c-rel*, respectively (25). By neutralising the *RelA* product, *RAI* protein would presumably shift this balance towards apoptosis. *RAI* protein may also influence p53 activity (26).

Predictor of *RAI* action

The present invention relates to method of estimating the disease risk of an individual comprising further a predictor of *RAI* action. In one embodiment the length of the *RAI* transcript may be used as a predictor. In another embodiment the quantity of *RAI* transcripts may be used as a predictor. The present invention comprises the combined *RAI* transcript characteristics described above as a predictor of *RAI* action and the estimation of the disease risk of an individual.

Examples

Materials and methods

Study groups. Diet, Cancer and Health (DCH) is a Danish prospective follow-up study. Individuals eligible for invitation were born in Denmark, were living in the Copenhagen or Aarhus areas, and were not at the time of invitation registered with a previous diagnosis of cancer (including non-melanoma skin cancer) in the Danish Cancer Registry. Invited to participate were 160,725 individuals aged 50-64 years of which 57,053 individuals were recruited (22). Among these 542 were later notified to the Danish Cancer Register with a cancer diagnosed before the date of enrolment and were therefore excluded. At enrolment (1993-1997), detailed information on diet, smoking habits, lifestyle, weight, height, reproduction, medical treatment, and other socio-economic characteristics and environmental exposures were collected. For women hormone replacement therapy (HRT) and menopausal status were also recorded. Moreover, blood, urine, fat tissue and other biological material was sampled and stored in a biobank at -150°C . Cohort members were identified by a unique identification number, which is allocated to every Danish citizen by the Central

Population Registry. Cohort members were linked to The Central Population Registry for information on vital status and immigration. Information on cancer occurrence among cohort members was obtained through record linkage to the Danish Cancer Registry, which collects information on all inhabitants in Denmark who develop cancer (7). Linkage was performed by use of the personal identification number.

Study groups for breast cancer. In "Diet Health and Cancer" 79,729 women aged 50-64 years were invited to participate in the study and 29,875 accepted the invitation. Of these a total of 326 women that later were reported to the Danish Cancer Registry with a cancer before the visit to the study clinic were excluded from the study. In addition, 8 women were excluded from the study because they did not fill in the lifestyle questionnaire. Because the present analysis aimed at the subgroup of women, who were postmenopausal at study entry, we further excluded 4,844 supposedly pre-menopausal women, including 4,798 women who had reported at least one menstruation no more than 12 months prior to entry and no use of hormone replacement therapy (HRT), nine women who gave a lifetime history of no menstruations and 37 women who did not answer the questions about current or previous use of HRT, leaving 24,697 postmenopausal women.

Each cohort member was followed-up for breast cancer occurrence from date of entry, i.e. date of visit to the study centre until the date of diagnosis of any cancer (except for non-melanoma skin cancer), date of death, date of emigration, or 31 December 2000, whichever came first. A total of 434 women were diagnosed with incident breast cancer during the follow up period (Vogel et al. In press), and a similar number was selected as controls by individual matching for age, HRT use and menopausal status.

Study groups for lung cancer. These were also recruited from "Diet, Health and Cancer". Among the 56,511 individuals with no previous cancer diagnosis, 2,291 with incomplete, inconsistent or missing information on smoking habits were excluded. Among the remainder 54,220 cohort members included in this study, 265 cases of lung cancer, diagnosed between 1994 and 2001, were identified in the files of the Danish Cancer Registry (23). Also from among the cohort members, a sub-cohort of 272 persons (including 4 of the cases) were selected at random and weighted in numbers similarly to cases within strata defined by sex, year of birth (5-year inter-

vals) and duration of smoking (10 years intervals). Blood samples were available in the biobank for 521 of the 533 selected individuals (Vogel et al. In press).

Study groups for basal cell carcinoma (basocellular carcinoma BCC). The groups of Caucasian Americans with and without BCC have been described previously (Athas et al, Cancer Res. 51:5786-5793, 1991; Wei et al, Proc. Natl. Acad. Sci USA, 90: 1614-8, 1994). Briefly, the study was a clinic based case control study at the Johns Hopkins Hospital, which serves multiple participating dermatologists in Maryland. Cases were histo-pathologically confirmed primary BCCs and were diagnosed between 1987 and 1990. The controls were patients from the same physician practices and had a diagnosis of mild skin disorders. All participants were Caucasians living near Baltimore and were between 20 and 60 years of age. The controls were frequency matched to the cases by age and sex. Cases and controls with any other forms of cancer were excluded. In the questionnaire, the study subjects were asked if they had any blood relatives with skin cancer, and were asked to specify the type of cancer. Study subjects with relatives with basal cell carcinoma and squamous cell carcinoma and 'skin cancer' were included in the group of subjects with a family of skin cancer. Subjects with relatives with melanoma were not included. At the clinic visit the subjects gave informed consent, were examined by dermatologists, completed a structured questionnaire and provided blood. Available frozen lymphocytes were genotyped. Initially, 71 cases and 118 controls were included in this study. However, the number of persons varied between analyses, as the supply of DNAs gradually was depleted. In case of the SNP RAI intron 1 only 133 persons could be genotyped reliably.

The examples relate to prediction from sequence polymorphisms in and around the region r to cancer, see Fig. 1 for an overview of a subregion of chromosome 19q. In the present study 27 markers were investigated within this 69 kb stretch for association with breast, lung, skin cancer, and multiple myeloma, using linkage disequilibrium mapping based on single markers, as well as, on haplotypes combining several neighboring markers. We report that haplotypes with maximal association to all three cancers are centered around the gene RAI.

Typing of SNPs

Table 10 lists the polymorphisms used in this study, their nature, the numbers in the NCBI database dbSNP, the sequence defining the SNPs, and their position therein, the currently estimated chromosome positions and their relative position within the region of interest (which can readily be calculated based on the information provided in the table). Table 11 lists the primers used for the PCR reactions. Table 12 lists the probes used for detection and typing. Table 13 lists the PCR regimen for the polymorphisms using Lightcycler, Taqman and ABI 2720.

The particular sequence polymorphisms analysed in these examples are listed in Table 10, together with their sources of information and their definition as sequences.

Table 10

Trivial name	Kind	dbSNP #	Sequence	Position in se- quence	Relative Position
<i>XPD</i> exon 23	A/C	Rs1052559	NT_011109	18123137	0
<i>XPD</i> exon 10	A/G	Rs1799793	NT_011109	18135477	12340
<i>XPD</i> exon 6	A/C	Rs238406	NT_011109	18136527	13390
<i>XPD</i> intron 3	A/G	Rs1799783	NT_011109	18140254	17117
<i>XPD</i> -4bp	GACA/-	Rs3916791	NT_011109	18142345	19208
<i>XPD</i> -81bp	81bp/-	Rs3916787	NT_011109	18143027	19890
<i>XPD</i> -5'2	A/G	Rs2097215	NT_011109	18144005	20868
<i>XPD</i> -5'3	C/T	Rs11878644	NT_011109	18145185	22048
<i>RAI</i> -3'7	C/T	Rs7252567	NT_011109	18146823	23686
<i>None</i> 1)	ATTTT/-			18147012	<u>23875</u>
<i>None</i> 2)	A/G	Rs2377329		18147126	23989
<i>RAI</i> -3'3	AA/-	Rs3047560	NT_011109	18147192	24055
<i>None</i> 3)	G/T			18146233	
<i>None</i> 4)	G/T	rs10422489		18147886	
<i>None</i> 5)	C/T	rs10426701		18148193	
<i>None</i> 6)	C/T			18149120	
<i>None</i> 7)	A/G			18149154	
<i>RAI</i> -3'4	A/G	Rs4544343	NT_011109	18150199	27062
<i>None</i> 8)	C/G			18150815	
<i>None</i> 9)	A/G	Rs8101662		18150911	
<i>None</i> 10)	G/T			18151158	

RAI exon 6	A/T	Rs6966	NT_011109	18151180	28043
RAI intron5-3	C/T	Rs10417235	NT_011109	18152171	29034
RAI intron5-2	A/T	Rs8112723	NT_011109	18153497	30360
RAI intron 3	A/G	Rs2017104	NT_011109	18155483	32346
RAI intron 1	A/G	Rs1970764	NT_011109	18159091	35954
None 11)	A/G			18159263	
None 12)	A/G			18160363	
None 13)	A/C			18160936	
None 14)	C/G			18160937	
None 15)	A/C	rs10419090		18161433	
None 16)	G/T			18161694	
None 17)	C/GACCTTGG CGCCAC- CATCTT			18161841	
None 18)	C/T			18161896	
RAI intron 1-2	C/T		AC092309	26729	39249
			NT_011109	18162206	
None 19)	C/T			18162309	
None 20)	A/G			18162356	
None 21)	A/G	rs12986272		18162599	
None 22)	A/C			18162903	
RAI intron1-3	A/C		AC092309	27493	39833
			NT_011109	18162970	
None 23)	C/T			18162986	
None 24)	A/C			18163200	
None 25)	C/G			18165052	
RAI-5'2	C/T	Rs4803814	NT_011109	18168944	45807
RAI-5'3	C/T	Rs4803815	NT_011109	18168948	45811
RAI-5'	C/T	Rs4572514	NT_011109	18171984	48847
ASE1-5'2	G/T	Rs2226949	NT_011109	18175327	52190

<i>ASE1</i> exon 1	A/G	Rs967591	NT_011109	18178152	55015
<i>ERCC1-3'2</i>	A/G	Rs735482	NT_011109	18180220	57083
<i>ERCC1-3'</i>	C/T	Rs762562	NT_011109	18180561	57424
<i>ERCC1-3'3</i>	A/G	Rs2336219	NT_011109	18180624	57487
<i>ERCC1</i> exon 4	C/T	Rs3177700	NT_011109	18191871	68734

rs numbers were derived from the NCBI's database dbSNP; Nucleotide sequence of polymorphisms identified in the present invention with no trivial name as yet:

- 1) GGTATT[ATTTT]Ntgagatggatttt (SEQ ID NO: 19)
- 5 2) ctggggaggtcagggcaggagaatc[A/G]cttgaaaccgggagggcggaggtgt (SEQ ID NO: 22)
- 3) GATTGTCATGT[G/T]ACATCAGCCAATACT (SEQ ID NO: 2)
- 4) AAAAACTAAAGTGGGGTTTGC GGG[G/T]AGTGGGAGGGCCCTTCCTGCTAGGT (SEQ ID NO: 24)
- 5) caggcggatcacaaggtcaggagtt[C/T]gagaccagcctggccaacacagtga (SEQ ID NO: 25)
- 10 6) CACAGTGAAAC[C/T]CCATCTCTACTAAA (SEQ ID NO: 3)
- 7) AAAATTAGCCGG[A/G]CGCCATGGCGGGAG (SEQ ID NO: 4)
- 8) AGCCTGGCCAACATG[C/G]TGAAACCCCGTCTCT (SEQ ID NO: 5)
- 9) ctggggaggtcagggcaggagaatc[A/G]cttgaaactcaggagggcagaggtgc (SEQ ID NO: 27)
- 10) AAGTTTCTCTATT[G/T]TGTTTATAAACA (SEQ ID NO: 6)
- 15 11) TTCTCCTGACCTC[A/G]TGATCCGCCCACCTCGG (SEQ ID NO: 7)
- 12) GGGATTACAGGCATGC[A/G]CCACCAGGCCAGCTAATTTTTGT (SEQ ID NO: 8)
- 13) TCCAATGGTGACA[A/C]CAGTAAGAGCAGTTAACAG (SEQ ID NO: 9)
- 14) TCCAATGGTGACAA[C/G]AGTAAGAGCAGTTAACAG (SEQ ID NO: 10)
- 15) tacaggcgcccgccaccacccccag[A/C]taattttgtatttttagtagagac (SEQ ID NO: 29)
- 20 16) TTGCCTCAGCCTCCTGA[G/T]TAGCTGGGATTGGAATGAGA (SEQ ID NO: 11)
- 17) TACGATAAATAGCTAGA[C/GACCTTGCGCCACCATCTT] (SEQ ID NO: 20)
- 18) AAAATAATAATAATAATATTAA[C/T]CCTGACCTTGCGCCACCATCT (SEQ ID NO: 12)
- 19) CCTCATGAGCCACCCAC[C/T]TCGGCCTCCCAAAGTGCT (SEQ ID NO: 13)
- 25 20) TGAGCCACCGCGCCC[A/G]GCCGAGACTCACTATTT (SEQ ID NO: 14)
- 21) taaagcgggaggtggttgaaact[A/G]ggaggcggaggtgcagtgagccga (SEQ ID NO: 31)
- 22) GGAGAGAAGGAGCAGAGAAC[A/C]TCTCTATGTGGCCA (SEQ ID NO: 15)
- 23) TTTCCCAGCATCCCA[C/T]TGCAATGAGGCTCCTGGCC (SEQ ID NO: 16)
- 24) TCCTGACTCCAGTG[A/C]GGTGCCTACAGTCCTG (SEQ ID NO: 17)
- 30 25) TTCCAGCCTGGGCAAGAA[C/G]AGTGAAACTCCAGCTT (SEQ ID NO: 18)

Statistics and haplotype assignment

Data recording and calculations and tests of allele frequencies were performed in SPSS and Excel. Calculation of the relative risk and confidence intervals for the single polymorphisms was performed in SAS (SAS Institute, Cary, NC, USA).

5 Simultaneous analysis of multiple SNPs employing haplotype trend regression (17) was performed with HelixTree (GoldenTree, Bozeman, MT, USA). Haplotype trend regression is basically a two-stage procedure. First, genotype results in combination with population assumptions such as Hardy-Weinberg equilibrium are used to construct all haplotype probabilities corresponding to a given set of markers for each
10 individual. Secondly, the disease state (1 for cases, 0 for controls) is regressed on the haplotype probabilities of all individuals, resulting in a p-value for the overall association of the set of markers with disease, and parameters for association for each specific haplotype with disease. The HelixTree program will use sets of markers of defined size, typically 2 to 4 neighboring markers at a time, to scan the entire region.
15 It can then be used to derive frequencies of the individual haplotypes, to calculate an overall p-value for the distribution of haplotypes covering a given set of markers among cases and controls, and to calculate p-values for the distribution of each haplotype derived from a given set of markers.

The program RASCAL for performing gene localization according to Lazzeroni (16)
20 was implemented in Delphi (B.A. Nexø, unpublished). We used a bootstrap set of 10000 sets of haplotypes selected from the original set with replacement, and to avoid a Q-form with negative values we used a relaxation value of 0.25. A 95% confidence interval for the location of the causative gene variant was derived from the Q-form. Places, where it took on a value less than 3.85 above the minimum, were
25 considered inside the confidence interval.

Three programs were used for assigning haplotypes to individuals on the bases of genotype data: HelixTree, Arlequin (18) and Phase (19). Arlequin (18) like HelixTree is a maximum likelihood algorithm, while Phase in addition includes a penalty for each new haplotype that is brought into play. Furthermore, the program Arlequin
30 includes missing values in its table of frequencies. To compensate for the latter, we normalised the values corresponding to fully defined haplotypes before including them in the analysis. All three were run under Windows 2000. The figure of the link-

age disequilibrium among the controls (Fig. 2) was also derived with the program HelixTree.

Primers and probes

The primers used for typing the polymorphisms are listed in Table 11. Table 12 lists the probes used for the polymorphisms.

Polymorphisms listed in Table 10 as having no trivial name have been deduced by sequencing in the present study.

A person skilled in the art will appreciate that primers and probes can be designed based on the information provided herein, in order to detect polymorphism by other means than sequencing, for example by PCR amplification methods as also used herein.

Primers were purchased from DNA Technology (Aarhus, Denmark), probes for the lightcycler were obtained from MolBiol (Tempelhofer Weg 11-12, Berlin) and the probes for Taqman were supplied by Applied Biosystem.

Table 11 Primers for the polymorphisms

Trivial name	Primers
<i>XPD</i> exon 23	5' atg cac cag gaa ccg ttt atg g 5' tct gtt ctc tgc agg agg atc
<i>XPD</i> exon 10	5' gat caa aga gac aga cga gc 5' gaa gcc cag gaa atg c
<i>XPD</i> exon 6	5' gta cca gca tga cac cag cct 5' tcc ctc cct gag ccc tg
<i>XPD</i> intron 3	5' aag gca gac aaa gga agg 5' gca agg aga agg aac agg
<i>XPD</i> -4bp	5' caa tca aaa aga aaa cat gg 5' tga gac gag gtg gag g
<i>XPD</i> -81bp	5' tgc ctc acc cct gta atc c (SEQ ID NO.: 47) 5' gct tgt aat ccc agc tac tcg
<i>XPD</i> -5'2	5' caa cac tca cac ccc aca g 5' aga tca cgc cac tgc act c
<i>XPD</i> -5'3	5' ttg aca att gag caa aga gc 5' tgg gat tac aga cgt gag c
<i>RAI</i> -3'7	5' cca gtc caa aca ata tga tcc 5' agt gca gcc tca act tcc
<i>RAI</i> -3'3	5' cat gat tca ctg cac cca acc 5' ttt cac tct tgt tgc cca agc
<i>RAI</i> -3'4	5' ttt tca cac aag tcc aat cc 5' act gca acc tcc atc tcc
<i>RAI</i> exon 6	5' ccc tgc ccc acc tct cc 5' agt caa ttt ctg tgc aaa cta ctt tta ttt
<i>RAI</i> intron 5-3	5' cat gac gag acc ctg tct cta cta aa 5' cac ctc ccg gat tca agt ga
<i>RAI</i> intron 5-2	5' gag gca aca gga aca aac c 5' cat tgg att gag cag aaa cc
<i>RAI</i> intron 3	5' taa cat aaa gaa tca gga gga ggc 5' agt tgg ctc atc tgc ctc tt

<i>RAI</i> intron 1	5' tgg cta aca cgg tga aac c 5' gga atc caa aga ttc tat gat gg
<i>RAI</i> intron1-2	5' act cct gac ttc aaa tga tcc 5' tag ccc cca gtc acg ttc c
<i>RAI</i> intron 1-3	5' aga agt cca aga gtt tgc agc 5' ttc tca gtc cca gaa tga acc
<i>RAI</i> -5'2	5' cca ctt agg taa aca cct ctt 5' ctg caa tga gcc gag ata gaa
<i>RAI</i> -5'3	5' cca ctt agg taa aca cct ctt 5' ctg caa tga gcc gag ata gaa
<i>RAI</i> -5'	5' atg ttg ggg aga ctg agg 5' ccg cat cta act tat tct gg
<i>ASE1</i> -5'2	5' aac tac ctc tgc aaa ccc agc 5' ttg gaa tgg agg gat tct acc
<i>ASE1</i> exon 1	5' ggt ttt ctg ctc tgc aca cg 5' cct ttc tcc ttc cac caa cg
<i>ERCC1</i> -3'2	5'-aca gag ccc aca gtg gag aca 5'-cta gag gct cag tgt taa tct gtt cct
<i>ERCC1</i> -3'	5' gga cag atg gca atg atg g 5' tct tct tct tgg tgg atg tgg
<i>ERCC1</i> -3'3	5'-acc atg gcg cct caa ca 5'-gaa ttg gct cag tca ctg tgt ga
<i>ERCC1</i> exon 4	5' ggc cct gtg gtt atc aag g 5' tct cat aga aca gtc cag aac act g
<i>RAI</i> -3'3 nested 1	5' cat gat tca ctg cac cca acc 5' ttt cac tct tgt tgc cca agc
<i>RAI</i> -3'3 nested 2	5' 6FAM-ctt gca cag tgg ctc atg c 5' tct tgt tgc cca agc tgg
Trivial name	Primers
<i>ASE1</i> exon 1	5' ggt ttt ctg ctc tgc aca cg 5' cct ttc tcc ttc cac caa cg
<i>ERCC1</i> -3'2	5'-aca gag ccc aca gtg gag aca 5'-cta gag gct cag tgt taa tct gtt cct
<i>ERCC1</i> -3'	5' gga cag atg gca atg atg g 5' tct tct tct tgg tgg atg tgg
<i>ERCC1</i> -3'3	5'-acc atg gcg cct caa ca 5'-gaa ttg gct cag tca ctg tgt ga
<i>ERCC1</i> exon 4	5' ggc cct gtg gtt atc aag g 5' tct cat aga aca gtc cag aac act g
<i>RAI</i> -3'3 nested 1	5' cat gat tca ctg cac cca acc 5' ttt cac tct tgt tgc cca agc
<i>RAI</i> -3'3 nested 2	5' 6FAM-ctt gca cag tgg ctc atg c 5' tct tgt tgc cca agc tgg

Table 12 Probes for the polymorphisms

Trivial name	Probes
<i>XPD</i> exon 23	5' FAM-ctc tat cct ctg cag cg-MGB 5' VIC-tat cct ctt gag cgt ct-MGB
<i>XPD</i> exon 10	5' LC Red640- cgt gct gcc caa cga agt g -p 5' gga cgc cca cct ggc caa cc -fluoresceine
<i>XPD</i> exon 6	5' FAM-ccc cac tgc cgc ttc tat gag gt-TAMRA 5' VIC-ccc cac tgc cga ttc tat gag gtt-TAMRA
<i>XPD</i> intron 3	5' LC Red640- ccc tgc ccc cca act ttg ga -p 5' gcc tcc aat gaa cac aag ctc -fluoresceine
<i>XPD</i> -4bp	5' LC Red640- cct ggg ttc gat caa tac tca gac a -p 5' ctc gct atc ttg ctc aag ctg atc tog aac -fluoresceine
<i>XPD</i> -81bp	5' agt cac agc tca ctg cag cct c -fluoresceine

<i>XPD</i> -5'2	5' LC Red640- acc tct tgg gct caa gcg atc ctc -p (SEQ ID NO.: 48) 5' LC Red640- aaa aaa aga ctt atc atg aca gga tgt ct -p (SEQ ID NO.: 49)
<i>XPD</i> -5'3	5' gca aga ctc cgt ccc aga aaa aga aaa -fluoresceine 5' LC Red640- tcc tct ctc tcc ccc agc tca ttt tg -p
<i>RAI</i> -3'7	5' aac cca ccc tac tgc tct gat ctc -fluoresceine 5' LC Red640- agg ctg gtc ttg aac tcc tgg gct taa g -p
<i>RAI</i> -3'3	5' ggt tcc gcc acg ttg cc -fluoresceine 5' LC Red640- tcg gct att ttt ttt ttt att ttt tta tt -p
<i>RAI</i> -3'4	5' att aca ggc acc cac cac cat g -fluoresceine 5' LC Red640- cct gga caa cat agg gag acc ctg tgt -p
<i>RAI</i> exon 6	5' caa aca aac aaa aac ctc tgc ca -fluoresceine 5' 6-FAM-tgc ctt cac aca gct ctg gtt taa tg - TAMRA
<i>RAI</i> intron 5-3	5' VIC - tgc ctt cac aca gca ctg gtt taa tg - TAMRA 5' FAM-tgg tgg tgc atg cct gta atc cc-BHQ
<i>RAI</i> intron 5-2	5' Yakima Yellow-tgg tgg tgc atg ccc gta atc-BHQ 5' acc atg ttg gcc agg ctg gtt tt -fluoresceine
<i>RAI</i> intron 3	5' LC Red640- atc tac tga cct caa atg atc cac ct -p 5' LC Red640- tgc aat ccg ccc gcc -p
<i>RAI</i> intron 1	5' cca ggc tgg ttt gga aat cct gag ctc -fluoresceine 5' LC Red640- ctg aga tcg cac cac tgc ac -p
<i>RAI</i> intron1-2	5' ggg agg cgg agc ttg cag tga -fluoresceine 5' ggc cat gcc tgt aat tct gta -fluoresceine
<i>RAI</i> intron 1-3	5' LC Red640- cag gac gag cca cag aca aaa ctc c -p 5' LC Red640- tgc aat gag gct cct ggc c -p
<i>RAI</i> - 5'2	5' act aca ttt ccc agc atc cca -fluoresceine 5' cct ccc tcc ctc cct gc -fluoresceine
<i>RAI</i> -5'3	5' LC Red640- tgc ttg ctt tct ctc tct -p 5' tcc ctg ctt gct tgc ttt ctc t -fluoresceine
<i>RAI</i> -5'	5' LCRed 640- tct ctc ttt ctt tct ttc ttt c -p 5' tgt tca tcc aaa tga gcc gc -fluoresceine
<i>ASE1</i> -5'2	5' LC Red640- agc ctg aac agg ttc tgt tcc ttc gac tt -p 5' LC Red640- caa gct gct atc tcg acc gat ctt -p
<i>ASE1</i> exon 1	5' ggg tga cca ccc tgc cag cc -fluoresceine 5' LC Red640- cgg gct aca ggg tta cct gag -p
<i>ERCC1</i> -3'2	5' tct gca acc tgg tgc gag cag c -fluoresceine 5' FAM-aaa ggg aaa gaa acc t-MGB
<i>ERCC1</i> -3'	5' VIC- cca aag gga cag aaa-MGB 5' LC Red640- tga ggc tcc gct cct tct gg -p
<i>ERCC1</i> -3'3	5' agc tgc cag agc tgc ctg ggc -fluoresceine 5' VIC-aca gca aaa tgc cac agt-MGB
<i>ERCC1</i> exon 4	5' FAM-aca gca aga tgc cac ag-MGB 5' cgc aac gtg ccc tgg gaa t -fluorescein 5' LC Red640- tgg cga cgt aat tcc cga cta tgt gct g -p

Table 13. PCR regimens used

Trivial name	Tech	Buffer	Buffer modification	Denaturation	Annealing	Elongation
<i>XPD</i> exon 23	T	M	-	15 sec 94 °C		60 sec 60 °C
<i>XPD</i> exon 10	L	R	5% DMSO	10 sec 95 °C	15 sec 53 °C	30 sec 72 °C
<i>XPD</i> exon 6	T	M	-	15 sec 94 °C		60 sec 63 °C
<i>XPD</i> intron 3	L	H	-	2 sec 95 °C	15 sec 55°C	30 sec 72 °C

<i>XPD</i> -4bp	L	H	2x Buffer	10 sec 95 °C	15 sec 60 °C	30 sec 72 °C
<i>XPD</i> -81bp	L	H	5% DMSO	10 sec 95 °C	15 sec 66 °C	30 sec 72 °C
<i>XPD</i> -5'2	L	H	2x dNTP + 2x Buffer	10 sec 95 °C	15 sec 64 °C	30 sec 72 °C
<i>XPD</i> -5'3	L	H	-	2 sec 95 °C	15 sec 71 °C	30 sec 72 °C
<i>RAI</i> -3'7	L	H	-	2 sec 95 °C	15 sec 67 °C	30 sec 72 °C
<i>RAI</i> -3'3	L	H	2x dNTP + 2x Buffer	2 sec 95 °C	15 sec 70 °C	30 sec 72 °C
<i>RAI</i> -3'4	L	H	-	2 sec 95 °C	15 sec 67 °C	30 sec 72 °C
<i>RAI</i> exon 6	T	M	-	15 sec 94 C		60 sec 60 °C
<i>RAI</i> intron 5-3	T	M	-	15 sec 94 C		60 sec 63 °C
<i>RAI</i> intron 5-2	L	H	-	2 sec 95 °C	15 sec 60 °C	40 sec 72 °C
<i>RAI</i> intron 3	L	H	2x dNTP	2 sec 95 °C	15 sec 63 °C	30 sec 72 °C
<i>RAI</i> intron 1	L	H	2x Buffer	0 sec 95 °C	10 sec 57 °C	15 sec 72 °C
<i>RAI</i> in-tron1-2	L	H	-	2 sec 95 °C	15 sec 66 °C	30 sec 72 °C
<i>RAI</i> intron 1-3	L	H	-	2 sec 95 °C	15 sec 65 °C	30 sec 72 °C
<i>RAI</i> -5'2	L	R	-	2 sec 95 °C	15 sec 61 °C	30 sec 72 °C
<i>RAI</i> -5'3	L	R	-	2 sec 95 °C	15 sec 61 °C	30 sec 72 °C
<i>RAI</i> -5'	L	H	1M Betain	2 sec 95 °C	15 sec 63 °C	30 sec 72 °C
<i>ASE1</i> -5'2	L	H	5% DMSO + 1½x Primer	2 sec 95 °C	15 sec 63 °C	30 sec 72 °C
<i>ASE1</i> exon 1	L	H	-	2 sec 95 °C	15 sec 60 °C	30 sec 72 °C
<i>ERCC1</i> -3'2	T	M		15 sec 94 °C		60 sec 60 °C
<i>ERCC1</i> -3'	L	H	-	2 sec 95 °C	15 sec 62 °C	30 sec 72 °C
<i>ERCC1</i> -3'3	T	M		15 sec 94 °C		60 sec 62 °C
<i>ERCC1</i> exon 4	L	H	5% DMSO	0 sec 95 °C	15 sec 57 °C	25 sec 72 °C
<i>RAI</i> -3'3 nested 1	A	H	-	30 sec 94 °C	30 sec 64 °C	30 sec 72 °C
<i>RAI</i> -3'3 nested 2	A	H	-	30 sec 94 °C	30 sec 62 °C	30 sec 72 °C

Where R is Roche buffer, H is homebrew buffer and M is mastermix, respectively, and where T denotes Tagman technique, L is Lightcycler and A is ABI2720 sequencing, respectively.

5

R is Roche buffer: 10 pmole primers

- 5
- 1 pmole probes
 - 1.3 nmole dNTP
 - 1 ul Hybridization Mix for Lightcycler (Roche)
 - 1 ul MgCl₂ for Lightcycler (Roche)
 - 1 ul DNA
 - water to 10 ul
- 10
- H is Homebrew:
- 10 pmole primers
 - 1 pmole probes
 - 1.3 nmole dNTP
 - 0.25 ul Titanium Taq (BD Biosciences, Palo Alto, CA)
 - 1 ul Titanium Taq Buffer (BD Biosciences, Palo Alo, CA)
 - 1 ul DNA
 - water to 10 ul

15

Determination of polymorphisms by real-time PCR using Taqman probes. Polymorphisms analyzed on Taqman were scored on the basis of the relative reaction of the 2 probes. Some polymorphisms were analysed using the ABI Prism 7700 sequence detection system (Applied Biosystems, Foster City, Ca, USA). PCR Primers and

20 Taqman probes were designed using Primer Express v 1.0 (Applied Biosystems). The reactions were performed in MicroAmp optical tubes sealed with MicroAmp optical caps (Applied Biosystems) containing a 10 µl reaction volume (Mastermix): 1x Taqman buffer A, 2.5mM MgCl₂, 200 µM each of dATP dCTP, dGTP, 400µM dUTP, 800nM each primer, 200nm each probe, 0,01U/µL AmpErase UNG, 0,025 U/µL

25 AmpliTaq Gold Polymerase. Tubes were incubated at 50°C for 2 min followed by 10 min at 95°C. The incubation was succeeded by 45 cycles with thermal cycler conditions as shown in Table 13.

Determination of polymorphisms by Lightcycler. Genotypes of the American persons and the Danish persons for polymorphisms in XPD exon 10, XPD-4bp, XPD-81bp, XPD-5'2, RAI-3'3, RAI-3'7, RAI intron 3, RAI intron 1, RAI intron 1-2, RAI-5', RAI-5'2, RAI-5'3 ASE1-5'2, ASE1 exon 1, ERCC1-3' and ERCC1 exon 4 were detected using LightCycler™ (Roche Molecular Biochemicals, Mannheim, Germany), scored on the bases on the temperature dependency ("melting profile") of the fluorescence. For XPD exon 10 PCR was performed by rapid-cycling in a reaction vol-

30

35

ume of 20 μ l with 0.5 μ M of each primer, 0.045 μ M of anchor and sensor probe, 3.5 mM $MgCl_2$, approximately 7 - 25 ng genomic DNA, and 2 μ l LightCycler DNA Master Hybridization probe buffer (Roche Molecular Biochemicals, Cat. No 2158 825). This buffer contains Taq DNA polymerase, dNTP mix, and 10 mM $MgCl_2$ and 5% DMSO.

5 The temperature cycling consisted of denaturation at 95°C for 2 sec, followed by 46 cycles consisting of 10 sec at 95°C, 15 sec at 53°C, and 30 sec at 72°C, see table 4a for PCR regimen for the polymorphisms using LightCycler™. For detection of the polymorphisms XPD-4bp, XPD-81bp, XPD-5'2, RAI-3'3, RAI intron 3, RAI intron 1, RAI intron 1-2, RAI-5', ASE1-5'2, ASE1-5'2, ASE1 exon 1, ERCC1-3' and ERCC1

10 exon 4 using LightCycler™ PCR was performed using Homebrew buffer, the composition of which is listed in table 13a. XPD exon 10, RAI-5'2 and RAI-5'3 were run in Roche buffer, the composition of which is also listed in table 13. Buffer modifications for the individual assay are shown in table 13 below. The temperature cycling parameters are shown for each of the reactions in table 13. The last annealing period at 72°C was extended to 120 sec followed by denaturation for 10 sec at 95°C.

15 The melting profile was determined by a temperature ramp from 50°C to 95°C with a rate of 0.1 degree/sec.

EXAMPLE 1

20 Pair-wise linkage disequilibrium of all markers of controls

To get an initial overview over the many markers the pair-wise linkage disequilibrium of all markers in the controls was calculated, see Fig. 2. It is clear that the region consists of 2 major haplotype blocks with a few interspersed markers in between.

25 One haplotype block spans roughly 16 markers and 26 kb from XPD exon6 to RAI intron1-3, while the other haplotype block spans 6 markers and 9 kb from RAI-5' to ERCC1-3'3.

EXAMPLE 2

30 Relative risk of breast cancer

Blood samples were collected and frozen from a large number of Danish postmenopausal women who had suffered from breast cancer. An age limit of 55 years was used to separate early from late cases. The cut-off was forced by a previous decision to use 50 years of age as an entrance criteria for inclusion in the cohort. The

35

salient features of this cohort are shown in Table 14. DNAs were purified from the blood samples of these persons and the polymorphisms as shown in Table 10 were typed.

5 **Table 14 Features of the investigated breast cancer cases and controls.**

	Breast cancer
Background population	Danes
Recruitment	Population based
Epidemiological design	Cohort
No. of cases	428
No. of controls	433
Age at inclusion	50-64 yrs of age
Cutoff for young cases	55 yrs of age
No. of young cases	62

10 The relative risk of breast cancer was calculated for the two homozygous forms of all the markers as well as the confidence intervals. This relatively simple analysis has the advantage that no a priori assumptions about the distribution of genotypes are needed. Fig. 9 depicts the relative risks and the lower confidence limits of the relative risks for postmenopausal breast cancer below age 55 as a function of the location of the markers. The values were initially calculated with the wild-type allele as reference, however, if the calculation gave a RR values less than 1 we used the reciprocal value of the RR and the reciprocal value of the high confidence limit instead. It is obvious that a significant association of markers with cancer exists in the region from 13 000 to 39 000 bases, corresponding to the first mentioned haplotype block and covering most of the gene *RAI* and the 5' region of the gene *XPD*. The corresponding curves for the other age-brackets did not show significant regions of association (results not shown).

20

EXAMPLE 3

Distribution of p-values for sets of markers in relation to breast cancer

25 As previous experience (Vogel et al. In press; Nexø et al. In press) had indicated to us that combined analyses involving multiple markers were superior to analysis of individual markers, we used the analytical technique known as haplotype trend regression (Zaykin et al. 2002), more specifically the program HelixTree, to associate sets of markers with the individual diseases. Haplotype trend regression is basically a two-stage procedure. First, genotype results in combination with population assumptions such as Hardy-Weinberg equilibrium are used to construct all haplotype

30

probabilities corresponding to a given set of markers for each individual. Secondly, the disease state (1 for cases, 0 for controls) is regressed on the haplotype probabilities of all individuals, resulting in a p-value for the overall association of the set of markers with disease, and parameters for association for each specific haplotype with disease. The HelixTree program will use sets of markers of defined size, typically 2 to 4 neighboring markers at a time, to scan the entire region. It can then be used to derive frequencies of the individual haplotypes, to calculate an overall p-value for the distribution of haplotypes covering a given set of markers among cases and controls, and to calculate p-values for the distribution of each haplotype derived from a given set of markers.

The entire sets of controls were used for the analyses. This broke the matching of the data sets, but matching on other criteria than ethnicity and age may well be irrelevant for genetic association studies (Hemminki & Forsti 2001) (21).

Fig. 4 shows the overall distribution of p-values for sets of markers plotted against the position on the chromosome for breast cancer. As position on the abscissa we used the median marker position in a given set of markers. The ordinate values are the negative logarithms to the overall p-values for a difference between cases and controls associated with a given set of markers. Each curve corresponds to a given size of marker sets, i.e. the number of markers in the haplotypes.

The results in Fig. 4 suggest that the association with breast cancer has two peaks. One was located at roughly 24 kb and one at 39 kb. When the cases were broken down into early and late breast cancer the peak at 24 kb was present in both groups, while the peak at 39 kb was only present in the older group. The peaks were clearly present in curves for haplotypes of 2, 3 and 4 neighboring markers. Larger sets were less informative, presumably due to increased degrees of freedom, but they essentially corroborated the results (results not shown).

Remarkably, this association at 24 kb was present for breast cancer both in young and in older female cases. This is the first time an association to *RAI* has been reported for cancers in older populations. The position of the optimal set of markers corresponds to the 3' part of the gene *RAI* and the inter-gene region between *RAI*

and *XPD* and was identical for young and older cancer cases. The peak at 39 kb which is only present in the older breast cancers correspond to the 5' part of *RAI*.

EXAMPLE 4

5 Frequency of individual haplotypes made up from 2 neighbouring markers in relation to breast cancer.

To understand better the nature of the association, we tabulated the frequencies of individual haplotypes made up of 2 neighboring markers centring on the position at 24 kb with maximal association, in various groups of cases and controls (RAI-3'3
10 RAI-3'4; Table 15). We also tabulated p-values associated with the individual haplotypes. It was the same haplotypes that differed when young and older breast cancers were compared to controls (RAI-3'3^s RAI-3'4^c down; RAI-3'3^l RAI-3'4^t up). A similarity between patterns in the different cancer forms was not obvious, however, the frequencies of the haplotypes in the control groups were similar.

15

Similarly, we calculated haplotypes for the set of 2 markers, located at 39 kb. Again the deviation of young and old breast cancer cases from the controls had similarity (RAI intron1-2^g RAI intron1-3^a up; RAI intron1-2^g RAI intron1-3^c down). The same
20 pattern was present in the lung cancer data. This is remarkable, as neither the young breast cancers nor the lung cancers showed a peak in this position, and suggests an underlying similarity of haplotypes in the different groups.

Table 15. Frequency of 2-haplotypes of cancer cases and controls and their association with cancer

25

¹⁾ p-values in parentheses relative to breast cancer controls, all ages

Haplotype	Agegroup	Breast cancer cases ¹⁾		Breast cancer controls	Lung cancer cases ²⁾	Lung cancer Controls	Lung cancer cases with XPDexon23 ^{AA}	Lung cancer controls with XPDexon23 ^{AA}
		Young	Old	All ages	Young	All ages	Young	All ages
RAI-3'3 ^s	RAI-3'4 ^c	0.280 (0.013)	0.352 (0.066)	0.399	0.424 (0.72)	0.414	0.222 (0.44)	0.302
RAI-3'3 ^s	RAI-3'4 ^t	0.000 (0.360)	0.007 (0.173)	0.000	0.000 (0.87)	0.014	0.055 (0.017)	0.008
RAI-3'3 ^l	RAI-3'4 ^c	0.009 (0.10)	0.006 (0.118)	0.001	0.000 (0.86)	0.002	0.000 (0.78)	0.004
RAI-3'3 ^l	RAI-3'4 ^t	0.7 (0.020)	0.634 (0.014)	0.600	0.576 (0.93)	0.570	0.722 (0.74)	0.686
RAI intron1-2 ^a	RAI intron1-3 ^a	0.259 (0.78)	0.209 (0.011)	0.271	0.301 (0.36)	0.263	0.000 ₃₎	0.047
RAI intron1-2 ^l	RAI intron1-3 ^c	0.000 (0.89)	0.007 (0.13)	0.000	0.017 (0.832)	0.000	0.000	0.000
RAI intron1-2 ^g	RAI intron1-3 ^a	0.69 (0.12)	0.701 (0.001)	0.615	0.654 (0.181)	0.623	0.944 ₃₎	0.803
RAI intron1-2 ^g	RAI intron1-3 ^c	0.051 (0.046)	0.084 (0.11)	0.113	0.028 (0.088)	0.114	0.056 ₃₎	0.150

²⁾ p-values in parentheses relative to skin cancer controls, all ages

³⁾ p-value could not be calculated

5

10

15

The assignment of frequencies to individual haplotypes also made it possible to derive a measure of association which was stochastically independent of group sizes. Under the assumption that a putative causative variant must be at least equally as associated with disease as any surrogate haplotype we calculated the odds ratios for each haplotype of each set of two neighbouring markers versus the three other haplotypes of the set and chose the maximum value. This value we then plotted against the median position of the set (Fig. 5). The resulting curves indicate a maximal association of the set RAI-3'3 RAI-3'4 at 25 kb with a value of approximately 4.4 for the young women, with rapid fall-off at either side. The curve for older women as expected had two maxima, one at the same place as the young women with a maximum of 7.3, and one associated with the set RAI intron1-2 RAI intron1-3 with a maximum of the same height.

20

To make sure that the derived haplotype frequencies did not reflect a peculiarity of the particular program used, we also calculated the frequencies of individual haplotypes in cases from breast cancer and basal cell carcinoma combined and in the corresponding controls, using 2 other computer programs, Arlequin and Phase. Ar-

lequin (18) like HelixTree is a maximum likelihood algorithm, while Phase in addition includes a penalty for each new haplotype that is brought into play. Furthermore, the program Arlequin includes missing values in its table of frequencies. To compensate for this, we normalised the values corresponding to fully defined haplotypes before including them in the analysis. It was evident that results from all 3 programs were similar. In particular, HTR and Arlequin produced essentially identical results (0 to 2 percent difference, except when the values were essentially zero). The results obtained with Phase occasionally diverged 5 to 10 percent of the value, presumably a reflection of the different optimization criterion, but the change appeared to affect both cases and controls. Therefore, we are confident that our biological conclusions are independent of which computer program was used (results not shown).

We also assigned long haplotypes to all individuals of the breast cancer cases and controls using the program Phase (Stephens et al. 2001). To avoid overloading this program and to remove SNPs of little importance we reduced the set to 22 SNPs, by eliminating those with lowest heterogeneity. The resulting haplotypes were used as input to a program implementing Lazzeroni's technique (Lazzeroni 1998) (16) for locating gene variation (RASCAL, BA Nexø, unpublished). We used a bootstrap set of 10000 sets of haplotypes selected from the original set with replacement, and to avoid a Q-form with negative values we used a relaxation value of 0.25. With these values we found the most likely location to be at 20 kb very close to the marker *XPD*_81bp for the early breast cancers, (Fig. 6). Changing the relaxation factor to 0.5 moved the most likely location to 23 kb bases. A 95% confidence interval for the location of the causative gene variant was derived from the quadratic form for the residual variance. Places, where it took on a value less than 3.85 above the minimum were considered inside the confidence interval. The curve suggested that the CI stretches from approximately 10 kb to approximately 57 kb. Using the same parameters, we found an almost identical position for the gene variant among the late breast cancers, however in this case no part of the region could be statistically excluded as a possible location for the variant.

EXAMPLE 5

Distribution of p-values for sets of markers in relation to early basal cell carcinoma

DNA was isolated from lymphocytes from humans from the American cohort of patients with basal cell carcinoma and controls, described in Materials and Methods,

was typed with respect to a number of sequence polymorphisms located in and around the claimed region *r*. The salient features of the American cohort are shown in Table 16.

5 **Table 16. Features of the investigated basal cell carcinoma cases and controls.**

	Basal cell carcinoma
Background population	Caucasian Americans
Recruitment	Hospital based
Epidemiological design	Case-Control
No. of cases	71
No. of controls	118
Age at inclusion	29 – 60
Cutoff for young cases	50 yrs of age
No. of young cases	45

Figure 7 shows the same kind of result for early basal cell carcinoma. Again we found a strong association with the region around *RAI*, with the curve for sets of 3 markers possibly shifted slightly to the right, and the curve for 4 markers showing suggestions of an extra association at higher positions.

EXAMPLE 6

Distribution of p-values for sets of markers in relation to lung cancer

15

DNA from humans from the Danish cohort of patients with lung cancer and controls was typed with respect to a number of sequence polymorphisms located in and around the claimed region *r*.

Evidence exists that a separate effector is associated with the marker XPD exon 23 which influences the lung cancer risk and thus may interfere with accurate mapping. Therefore the Danish men and women who had suffered from lung cancer were stratified according to the value of the marker XPD exon 23. The salient features of the Danish cohort are shown in Table 17.

20

Table 17. Features of the investigated lung cancer cases and controls.

	Lung cancer	Lung cancer w/ XPD exon23 ^{AA}
Background population	Danes	Danes
Recruitment	Population based	Population based
Epidemiological design	Cohort	Subcohort
No. of cases	249	83
No. of controls	259	115
Age at inclusion	50 – 64 yrs of age	50 – 64 yrs of age
Cutoff for young cases	56 yrs of age	56 yrs of age
No. of young cases	33	9

To minimize the interference of this second effector, we stratified the young population for *XPB* exon 23, and calculated overall p-values for each value of *XPB* exon23. Fig. 7 shows that also this disease is associated with sets of markers spanning the distal end of *RAI* and the inter-gene region in the population with the *XPB* exon23^{AA} genotype. This is the group of patients free of influence of the *XPB* gene. No association was evident in the two other groups (results not shown).

EXAMPLE 7

The results obtained in examples 1, and 3 to 6 rely on the use of the HelixTree program. In an attempt to control for this dependency the odds ratios for test of the two repective homozygotes of each individual marker against cancer status were plotted against the marker position on the chromosome 19, see Fig. 8. Where one of the four homozygotic groups was empty heterozygotes were included in the analysis. This relatively simple analysis has the advantage that no assumptions about the phase of different markers are necessary.

The data for early cancers show a sharp maximum in the marker *RAI* intron1, and breast cancer in addition a minor peak in *XPB*_4bp. Among the late cancers there was no striking association with disease, but mapping with single SNPs is also not expected to be particular sensitive (results not shown).

EXAMPLE 8

The ASE-1 genotype influences relapse-free survival and survival in Multiple Myeloma

391 patients diagnosed with multiple myeloma were found eligible for transplantation in Denmark in the period 1993-2004.

304 patients were genotyped for the polymorphisms ERCC1 exon4, RAI intron1, ASE-1 e1, XPD exon23, XPD exon10. Clinical parameters were available for 360 patients.

The multiple myeloma patients were treated with high dose alkylating chemotherapy followed by autologous bone marrow transplantation. The overall survival was significantly shorter for patients who experienced relapse than for patients who did not.

ASE-1 genotype was found to strongly influence the event-free survival (EFS) and overall survival in patients with multiple myeloma who were auto-transplanted. Thus, homozygous carriers of the wild-type allele had mean EFS of 1160 days whereas carriers of the variant allele had a mean EFS of 1718 days ($p=0.0018$). Thus variant allele carriers had 1.5 year longer event-free survival, see Fig. 9. The EFS was similar for men and women, although the difference was only statistically significant for women (Table 18). This means that the ASE-1 genotype can predict who will benefit most from the present treatment regimen.

The overall survival was also found to be influenced by the ASE-1 genotype. Thus, homozygous carriers of the wild-type allele had an overall survival of 2117 days whereas carriers of the variant allele lived 2727 days, which is 1.67 years longer ($p=0.029$), see Fig. 10. There was no gender effect but the difference was only statistically significant for women.

This means that the ASE-1 genotype can predict who will benefit the most from the present treatment regimen.

Table 18

Event-free survival and overall survival for patients with different ASE-1 genotypes

Group	Median event-free survival, EFS, (5-95% CI)	p	Median overall survival (5-95 % CI)	P
All				
GG	736 (634, 838)	0.0018	1902 (1613, 2191)	0.0293

AA+AG	1397 (901, 1893)		3015 (2093, 3937)	
Men				
GG	736 (616,856)	0.106	2030 (1741, 2404)	0.2297
AA+AG	963 (693,1233)		2747 (2377, 3077)	
Women				
GG	714 (426, 1002)	0.0075	1897 (1384, 2410)	0.0368
AA+AG	1479 (1019, 2939)		3015 (2261, 3764)	

As Fig. 11 illustrates, female carriers of the variant allele of ASE-1 polymorphism had a longer relapse-free period than homozygous carriers of the wild-type allele (median 1479 days vs. 714 days), while Figure 12 illustrates that they lived longer (median 3015 vs. 1897). Similar analyses for the men revealed the same tendencies but no statistically significant differences (median 963 vs 736 days for the relapse-free period; mean 2747 vs. 2030 days for the survival), Fig. 13 and Fig. 14, respectively.

10 EXAMPLE 9

Survival from time of diagnosis and genotypes of relevant genes were determined for 432 lung cancer patients who had been enrolled in the prospective study 'Diet, Cancer and Health'.

15 High-risk haplotype carriers (defined as *ERCC1* exon 4^{AA}, *ASE1* exon1^{GG}, *RAI* Intron 1^{AA}) had a mean survival of 359 days compared to a mean survival of 244 for non-carriers (Table 19 and Fig. 15). The genotype XPD K751Q was also determined.

20 Table 19.

Genotype	N	Survival (days)	P ^a
High Risk carrier.			
Yes	105	359.3 +/- 37.1	
No	319	244.5 +/- 17.5	0.002

Missing	8		
XPD K751Q			
AA	146	330.8 +/- 33.8	0.053
AC	218	248.9 +/- 20.7	
CC	64	255.8 +/- 31.6	
Missing	4		

a) p for cox regression

When the two genotypes were combined, there were significant differences in survival between carriers of different genotype combinations, see Table 20. Thus, homozygous carriers of both favorable genotypes had a mean survival of 415 days whereas carrier of an unfavorable had a lower mean survival time, see Fig.16 and Fig. 17, respectively.

Table 20.

High-risk carrier	XPD K751Q	N	Survival (days)	P
No	AA	101	296 +/- 37	0.001 ^a
No	AC	164	222 +/- 24	
No	CC	51	217 +/- 29	
Yes	AA	44	415 +/- 74	
Yes	AC	51	318 +/- 42	
Yes	CC	10	346 +/- 83	
Missing		11		

a) the p-value is for a cox regression. P for high-risk is 0.004 and for XPD K751Q p=0.035.

EXAMPLE 10

Odds ratios for alleles in young cases vs controls

In analogy with the methods employed and results obtained in examples 1-9 the alleles of the markers listed below in table 20 have been found to have the indicated odds ratios for breast cancer in young cases versus controls when determined by sequencing of the region in approximately 10 cases and 10 controls. Furthermore, to make sure that an odds ratio could be calculated in all instances (i.e. that

the divisor was never 0) we added a value of 0.5 to each of the four alleles occurrences.

Table 20

Marker sequence in 24,000 region	Odds ratio	SEQ ID NO.:
TTTATAGTAGA- GACATGGTTCGCCA[C/T]GTTGCCAGGCTGGTCTTGAAGTCC GGTTTAT[ATTTT]Ntgagatggat	13,00	19
ctggggaggctgaggcaggagaatc[A/G]cttgaaaccgggaggcgagggtgt	0,56	22
actaaaaataaaaaataaaaaaa[-/AA]atagccgagcatgggtgggtgcc		23
GATTGTCATGT[G/T]ACATCAGCCAATACT	0,08	2
AAAAAAC- TAAAGTGGGGTTTGCGGG[G/T]AGTGGGAGGGCCCTTCCTGCTAGGT caggcggaacacaggtcaggagtt[C/T]gagaccagcctggccaacacagtg	3,86	24
	0,8	25
CACAGTGAAAC[C/T]CCATCTCTACTAAA	0,43	3
AAAATTAGCCGG[A/G]CGCCATGGCGGGAG	2,33	4
CCCTATGTTGTCCAAGCTGGCA- GAG[A/G]TTTTTGTGTTTGTGTTTGAAGGGA AGCCTGGCCAACATG[C/G]TGAAACCCCGTCTCT	2,33	26
	0,48	5
AAGTTTCTCTATT[G/T]TGTTTATAAACA	0,88	6
Marker sequence in 39,000 region:	Odds ratio	SEQ ID NO.:
ggagcttgacgtgagctgagatgc[A/G]ccactgcactccagcctgggcgaca	0,17	28
TTCTCCTGACCTC[A/G]TGATCCGCCACCTCGG	0,27	7
GGGATTACAGGCATGC[A/G]CCACCAGGCCAGCTAATTTTGT	10,33	8
TCCAATGGTGACA[A/C]CAGTAAGAGCAGTTAACAG	7,58	9
TCCAATGGTGACAA[C/G]AGTAAGAGCAGTTAACAG	7,86	10
tacaggcgcccgccaccacccag[A/C]taattttgtatttttagtagagac	3,29	29
TTGCCTCAGCCTCCTGA[G/T]TAGCTGGGATTGGAATGAGA	0,49	11
TACGATAAATAGCTAGA[C/GACCTTGGCGCCACCATCTT]	0,15	20
AAAATAATAATAATAATATTAA[C/T]CCTGACCTTGGCGCCACCATCT	0,34	12
tcgtcctgtacagaattacaggca[C/T]gcgccaccgctccgggctaatttt	0,60	30
CCTCATGAGCCACCCAC[C/T]TCGGCCTCCCAAAGTGCT	0,49	13
TGAGCCACCGCGCCC[A/G]GCCGAGACTCACTATTT	0,19	14
taaagcgaggaggtggctgaacct[A/G]ggaggcgagggtgacgtgagccga	2,51	31
GGAGAGAAGGAGCAGAGAAC[A/C]TCTCTATGTGGCCA	0,36	15
ATCCTAAAGACTAC[A/C]TTTCCAGCATCCCA	0,36	32
TTTCCAGCATCCCA[C/T]TGCAATGAGGCTCCTGGCC	0,41	16
TCCTGACTCCAGTG[A/C]GGTGCTACAGTCCTG	0,49	17

TTCCAGCCTGGGCAAGAA[C/G]AGTGAAACTCCAGCTT	1,5	18
---	-----	----

EXAMPLE 11.

Superior association of the marker RAI3'd1 with breast cancer.

- 5 To investigate the association of the length-polymorphism observed at position 18147012 in contig NT_011109, here called RAI3'd1, with breast cancer we developed a PCR reaction using one fluorescent primer and one non-fluorescent. The primers were:

RAI3'd1-f1: 5'-AAA AAA ATA GCC GAG CAT GG

- 10 RAI3'd1-r1: 5'-6FAM-TT TGG ACT GGG TAA GAA TTT CC

The PCR reaction contained

	DNA (ca 20 ng/ul)	2.5 µl
	Primers (100 uM)	0.075 µl each
15	dNTP (25 mM)	0.1 µl
	Titanium taq polymerase *)	0.065 µl
	Titanium taq buffer *)	1.25 µl
	Water	8.45 µl

- 20 * (BD Biosciences, Brøndby, Denmark)

The PCR regimen consisted of

- 25 Denaturation at 94°C for 5 min, 30 cycles consisting of a denaturation step(at 94°C for 30 sec), an annealing step (at 62°C for 30 sec) and an elongation step (at 72°C for 36 sec). A polishing step was included in the reaction (at 72°C for 6 min)

- 30 After the PCR, 1 µl of the reactions were diluted 1:20 in water, 1 µl of the dilutions were mixed with 12.5 µl formamide and 0.5 µl of fluorescent molecular length standards (Genescan 500, ROX, Applied Biosystems, Nærum, Denmark). The mixtures were denatured 2 min at 94 °C and put on ice. The resulting mixtures were analyzed on an ABI3730 unit (Applied Biosystems) with registration of both fluorophores.

DNAs from peripheral leukocytes from 434 breast cancer patients and a similar number of healthy controls, all part of the prospective study "Diet, Health and Cancer" (see previous examples) were analyzed.

The following sizes of PCR fragments were found:

Length	Number of chromosomes
230	1
238	993
239	7
259	2
263	1
264	409
269	49
274	28
279	2

5

We consider the difference between PCR fragments of length 238 and 239 non-significant. These are probably the same molecular entities, and were treated as such in the further analyses. The same goes for the fragments of length 263 and 264.

10

For sequencing the same PCR reactions were performed with non-fluorescent primers on select DNAs, preferentially from homozygotes. The primers RAI3'd1- f2: 5'-CCA -AGA TCG TGC CACRGT and RAI3'd1-r2: 5'- TTT GGA CTG GGT AAG AAT TTC C

15

were used for priming a big-dye dideoxynucleotide sequencing reaction, which was then analysed on a ABI 3130.

Examples of sequences observed, wherein the repeats are underlined/double underlined:

20

Sequence corresponding to 4 repeats:

238

GAAAATCATCTCAAAAATAATAAAATAAAATAAAATAAAATATAAACCGTCAGTT

25

Sequences corresponding to 8 repeats:

259

GAAAATCCATCTCAAAAATAATTAAATAAATAAATAAATAAATAAATAAAT
TAAATATAAACCAGTCAGTT

5

264

GAAAATCCATCTCAAAAATAATTAAATAAATAAATAAATAAATAAATAAAT
TAAATATAAACCAGTCAGTT

Sequences corresponding to 9 repeats:

10

269

GAAAATCCATCTCAAAAATAATTAAATAAATAAATAAATAAATAAATAAAT
TAAATATAAACCAGTCAGTT

274

15

GAAAATCCATCTCAAAAATAATTAAATAAATAAATAAATAAATAAATAAAT
TAAATATAAACCAGTCAGTT

We conclude that the difference between the major forms (length 238 and 264) is an increase in the number of repeats of the sequence TAAAA (complementary sequence ATTTT). Of the minor forms length 259 also seems to fit into this system, while length 269 and 274 seem identical to length 264 in the region investigated and may have differences elsewhere in the PCR fragment.

For the analysis of the association with breast cancer we categorized the fragments into short fragments (S: fragments 230 and 238) and long (L: all other sizes). We then calculated the relative risk of breast cancer as well as 95% confidence interval for homozygous SS and heterozygous SL individuals relative to the individuals with the LL genotype. This calculation was performed both on the material as a whole and after age-stratification. For each age group we finally calculated a p-value for trend in the relative risk among the three genotypes. The results are shown in table 21.

Table 21. Association of the polymorphism RAI3'd1 with risk of breast cancer.

Age group/ Genotype	N cases		N controls	RR (95% CI)	P trend
All					0.0008
	LL	38	59	1	
	SL	127	165	1.39 (0.81-2.38)	
	SS	200	150	2.44 (1.41-4.23)	
<=55 yr					0.01
	LL	5	15	1	
	SL	18	25	1.46 (0.33-6.47)	
	SS	33	14	6.29 (1.49-26.61)	
55-60 yr					0.19
	LL	12	17	1	
	SL	45	63	1.72 (0.63-4.73)	
	SS	61	49	2.57 (0.88-7.48)	
>60 yr					0.12
	LL	21	27	1	
	SL	64	77	1.09 (0.52-2.60)	
	SS	106	87	1.73 (0.83-3.64)	

It is obvious that persons homozygous SS are at greater risk of developing breast cancer.

This is apparent in the combined group as well as in the younger persons. In the two older age groups the tendency is suggestive. The relative risks in the combined group and among the younger patients are higher than we have observed with any other polymorphism. We conclude that RAI3'd1 is a superior marker for risk of breast cancer.

EXAMPLE 12

Interaction between RAI3'd1 and markers in the second haplotype block.

We have previously described that the markers investigated form two haplotype blocks, one ranging from XPD exon 23 to RAI intron1-3 and the other ranging from RAI5' to ERCC1 exon4. The marker RAI3'd1 is located in the first block. We found it of interest to look for interaction between RAI3'd1 and markers of the second haplotype block, because this might indicate a possibility for strengthening the association with breast cancer risk further.

Table 22 to 25 shows testing for interaction between RAI3'd1 and four markers located in the second haplotype block. We pooled the heterozygotes with the low risk homozygote for all markers. Using pooled groups as references tends to reduce the

relative risks, but the larger group sizes improve the statistical power. Indicated in each quadrant of the tables are the number of individuals in each subgroup, the relative risk for breast cancer and the corresponding confidence intervals. The p-value for interaction between the markers and the magnitude of the interaction are indicated below each table. It is clear that the middle two markers ASE1 exon1 and ASE1 exon3-2 show a statistically significant interaction with RAI3'd1, while the outer markers RAI5' and ERCC1 exon4 show tendencies in the same direction, but do not achieve statistical significance.

10 Table 22. Interaction between markers RAI3'd1 and RAI5' in relation to breast cancer

RAI3'd1/RAI5'	AA	AG+GG
LL+SL	n=279 1	n=107 1.24 (0.74-2.05)
SS	n=241 2.23 (1.47-3.37)	n=104 1.39 (0.82-2.35)
P(interaction) = 0.18. Magnitude of interaction = $1.39 - 2.23 - 1.24 + 1 = -1.08$		

15 Table 23. Interaction between markers RAI3'd1 and ASE1 exon1 in relation to breast cancer

RAI3'd1/ASE1 exon1	AA + AG	GG
LL + SL	n=109 1	n=277 0.70 (0.42-1.16)
SS	n=105 1.08 (0.58-1.99)	n=239 1.78 (1.08-2.04)
P (interaction) = 0.02. Magnitude of interaction = $1.78 - 1.08 - 0.70 + 1 = 1$		

20 Table 24. Interaction between markers RAI3'd1 and ASE1 exon3-2 in relation to breast cancer

RAI3'd1/ASE1 exon3-2	GG	AG+AA
LL+SL	n=311 1	n=76 1.44 (0.84-2.49)
SS	n=272 2.31 (1.56-3.43)	n=77 1.22 (0.67-2.22)
P(interaction) =0.02. Magnitude of the interaction = 1.22 – 2.31 – 1.44 + 1 = -1.53		

Table 25. Interaction between markers RAI3'd1 and ERCC1 exon4 in relation to breast cancer

RAI3'd1/ERCC1 exon4	GG + AG	AA
LL + SL	n=257 1	n=127 0.89 (0.54-1.46)
SS	n=160 1.55 (0.98-2.45)	n=184 2.16 (1.36-3.43)
P (interaction) =0.18. Magnitude of interaction = 2.16 – 1.55 – 0.89 + 1 = 0.72		

Inspired by these results we subdivided the patients and controls into three groups: those that were homozygote for the haplotype RAI3'd1^S ASE1 exon1^G, those that were sure or possible heterozygotes for the same haplotype, and those that were definite non carriers of the haplotype.

Table 26 lists the relative risk with confidence interval and the P value for each group.

Table 26. Risk of breast cancer in relation to presence of the haplotype RAI3'd1^S ASE1 exon1^G

	RR (CI)	P
Non-carriers (117)	1	
Heterozygotes (374)	1.31 (0.81-2.12)	0.26
Homozygotes (239)	2.51 (1.50-4.20)	0.0005

We made a similar calculation using only those breast cancer cases occurring before age 55 and the corresponding controls. The results are shown in table 27.

5

Table 27. Risk of breast cancer before age 55 in relation to presence of the haplotype RAI3'd1^S ASE1 exon1^G

	RR (CI)	P
Non-carriers(22)	1	
Heterozygotes (55)	1.20 (0.37-3.88)	0.77
Homozygotes (32)	6.78 (1.56-29.48)	0.01

10 It is apparent that the combined use of the two markers improves discrimination and increases the relative risk in the homozygote group.

15

Claims

1. A method for estimating the disease risk of an individual comprising

- in a sample from said individual assessing in the genetic material at least
5 one sequence polymorphism
- in a region corresponding to SEQ ID NO: 132, or a part thereof, or
- in a region complementary to SEQ ID NO: 132, or a part thereof, or
- in a transcription product from a sequence in a region corresponding to SEQ
10 ID NO: 132, or a part thereof, or
- or translation product from a sequence in a region corresponding to SEQ ID
NO: 132, or a part thereof, and optionally a further sequence polymorphism
- in a region corresponding to SEQ ID NO: 133, or a part thereof, or
- in a region complementary to SEQ ID NO: 133, or a part thereof, or
- in a transcription product from a sequence in a region corresponding to SEQ
15 ID NO: 133, or a part thereof, or
- or translation product from a sequence in a region corresponding to SEQ ID
NO:133, or a part thereof,
- with the proviso that when using SEQ ID NO: 132 only the at least one se-
20 quence polymorphism is in a region corresponding to at least one region
within SEQ ID NO: 132 selected from the regions flanked by and including
the sequence polymorphism RAI3'd1 (SEQ ID NO: 19) and rs2377329 (SEQ
ID NO: 22), or
- in a region flanked by and including the sequence polymorphism positioned
25 at 18149120 in Contig NT_011109 (SEQ ID NO: 3) and sequence polymor-
phism positioned at 18149154 in Contig NT_011109 (SEQ ID NO: 4), or
- the sequence polymorphism positioned at 18150815 in Contig NT_011109
30 (SEQ ID NO: 5), or
- the sequence polymorphism positioned at 18151158 in Contig NT_011109
(SEQ ID NO: 6)
- obtaining a sequence polymorphism response,
- 35

- estimating the disease risk of said individual based on the sequence polymorphism response.

- 5 2. The method according to claim 1, wherein the cell sample is a blood sample, a tissue sample, a sample of secretion, semen, ovum, a washing of a body surface, such as a buccal swap, a clipping of a body surface, including hairs and nails.
- 10 3. The method according to any of the preceding claims, wherein the cell is selected from white blood cells and tumor tissue.
4. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least one mutation base change.
- 15 5. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least two base changes.
6. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least one single nucleotide polymorphism.
- 20 7. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least two single nucleotide polymorphisms.
8. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least one tandem repeat polymorphism.
- 25 9. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least two tandem repeat polymorphisms.
- 30 10. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least one deletion polymorphism.
11. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least one insertion polymorphism.
- 35

12. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least one dinucleotide polymorphism.

5 13. The method according to any of the preceding claims, wherein the disease is cancer.

10 14. The method according to any of the preceding claims, wherein the cancer is selected from skin carcinoma including malignant melanoma, breast cancer, lung cancer, colon cancer and other cancers in the gastro-intestinal tract, prostate cancer, lymphoma, leukemia, multiple myeloma, pancreas cancer, head and neck cancer, ovary cancer and other gynecological cancers.

15 15. The method according to any of the preceding claims, wherein the cancer is selected from skin cancer, lung cancer, colon cancer, breast cancer, and multiple myeloma.

20 16. The method according to any of the preceding claims, wherein the cancer is selected from skin cancer, lung cancer, colon cancer and breast cancer.

17. The method according to any of the preceding claims, wherein the cancer is selected from skin cancer, lung cancer, and breast cancer.

25 18. The method according to any of the preceding claims, wherein the cancer is lung cancer.

19. The method according to any of the preceding claims, wherein the cancer is lung cancer in an individual with marker XPD exon 23^{AA}.

30 20. The method according to any of the preceding claims, wherein the cancer is selected from skin cancer and breast cancer.

21. The method according to any of the preceding claims 13-16 and 19, wherein the skin cancer is basal cell carcinoma.

35

22. The method according to any of the preceding claims, wherein the cancer is multiple myeloma.

23. The method according to any of the preceding claims, wherein the at least one sequence polymorphism is assessed in a region of SEQ ID NO: 132 flanked by and including the markers RAI 3'7 (SEQ ID NO: 21) and RAI 3'4 (SEQ ID NO: 26).

24. The method according to any of the preceding claims, wherein at least one sequence polymorphism is assessed in a region of SEQ ID NO: 132 flanked by and including the markers RAI 3'3 (SEQ ID NO: 23) and the marker sequence RAI 3' d1 GGTTTAT[ATTTT]Ntgagatggatttt (SEQ ID NO: 19).

25. The method according to any of the preceding claims, wherein at least one sequence polymorphism is the marker sequence RAI 3' d1 GGTTTAT[ATTTT]Ntgagatggatttt (SEQ ID NO: 19).

26. The method according to any of the preceding claims, wherein at least one sequence polymorphism is RAI 3'3 (SEQ ID NO: 23).

27. The method according to any of the preceding claims, wherein at least one sequence polymorphism is assessed in a region of SEQ ID NO: 133 flanked by and including the polymorphisms RAI 5' (SEQ ID NO: 4 and ASE 1 exon 3-2 (SEQ ID NO: 142).

28. The method according to any of the preceding claims, wherein at least one sequence polymorphism is assessed in a region of SEQ ID NO: 132 flanked by and including the regions flanked by and including the sequence polymorphism RAI3'd1 (SEQ ID NO: 19) and rs2377329 (SEQ ID NO: 22), and at least one sequence polymorphism is assessed in a region of SEQ ID NO: 133 flanked by and including the polymorphisms RAI 5' (SEQ ID NO: 4 and ASE 1 exon 3-2 (SEQ ID NO: 142).

29. The method according to any of the preceding claims, wherein at least one sequence polymorphism is assessed in a region of SEQ ID NO: 132 flanked by and including the regions flanked by and including the sequence polymorphism positioned at 18149120 in Contig NT_011109 (SEQ ID NO: 3) and sequence polymorphism positioned at 18149154 in Contig NT_011109 (SEQ ID NO: 4), and at least one sequence polymorphism is assessed in a region of SEQ ID NO: 133 flanked by and including the polymorphisms RAI 5' (SEQ ID NO: 4 and ASE 1 exon 3-2 (SEQ ID NO: 142).
30. The method according to any of the preceding claims, wherein the sequence polymorphism is assessed in SEQ ID NO: 132 said polymorphism is positioned at 18150815 in Contig NT_011109 (SEQ ID NO: 5), and at least one sequence polymorphism is assessed in a region of SEQ ID NO: 133 flanked by and including the polymorphisms RAI 5' (SEQ ID NO: 4 and ASE 1 exon 3-2 (SEQ ID NO: 142).
31. The method according to any of the preceding claims, wherein the sequence polymorphism is assessed in SEQ ID NO: 132 said polymorphism is positioned at 18151158 in Contig NT_011109 (SEQ ID NO: 6), and at least one sequence polymorphism is assessed in a region of SEQ ID NO: 133 flanked by and including the polymorphisms RAI 5' (SEQ ID NO: 4 and ASE 1 exon 3-2 (SEQ ID NO.: 142).
32. The method according to any of the preceding claims, wherein the sequence polymorphism assessed in SEQ ID NO: 132 is RAI3' d1 (SEQ ID NO:19), and the sequence polymorphism assessed in SEQ ID NO: 133 is ASE 1 exon 3-2 (SEQ ID NO: 142)
33. The method according to any of the preceding claims, wherein the sequence polymorphism assessed in SEQ ID NO: 132 is RAI3' d1 (SEQ ID NO: 19), and the sequence polymorphism assessed in SEQ ID NO: 133 is ASE 1 exon 1 (SEQ ID NO: 43)
34. The method according to any of the preceding claims, wherein the assessment is conducted by means of at least one nucleic acid primer or probe, such as a

primer or probe of DNA, RNA or a nucleic acid analogue such as peptide nucleic acid (PNA) or locked nucleic acid (LNA).

- 5 35. The method according to claim 34, wherein the nucleotide primer or probe is capable of hybridising to a subsequence of the region corresponding to SEQ ID NO: 132, or a part thereof, or a region complementary to SEQ ID NO: 132 and optionally further a subsequence of the region corresponding to SEQ ID NO: 133, or a part thereof, or a region complementary to SEQ ID NO: 133.
- 10 36. The method according to claim 34 or 35, wherein the primer or probe has a length of at least 9 nucleotide or peptide monomers.
- 15 37. The method according to any of the preceding claims 34-36, wherein at least one nucleotide primer or probe is capable of hybridising to a subsequence of SEQ ID NO: 132 selected from the group of subsequences
- | | |
|---|--------------|
| 1. TTTTAGTAGAGACATGGTTCGCCA[C/T]GTTGCCCAGGCT
GGTCTTGA ACTCC | SEQ ID NO:21 |
| 2. GGTTTAT[ATTTT]Ntgagatggatttt | SEQ ID NO:19 |
| 3. ctggggaggctgaggcaggagaatc[A/G]ctgaaacgggaggcggagggtgt | SEQ ID NO:22 |
| 4. actaaaaataaaaaataaaaaaa[-/AA]atagccgagcatggtggtgggtgcc | SEQ ID NO:23 |
| 5. GATTGTCATGT[G/T]ACATCAGCCAATACT | SEQ ID NO:2 |
| 6. AAAAACTAAAGTGGGGTTTGC GGG[G/T]-
AGTGGGAGGGCCCTTCCTGCTAGGT | SEQ ID NO:24 |
| 7. caggcggatcacaaggtcaggagt[C/T]-gagaccagcctggccaacacagtga | SEQ ID NO:25 |
| 8. CACAGTGAAAC[C/T]CCATCTCTACTAAA | SEQ ID NO:3 |
| 9. AAAATTAGCCGG[A/G]CGCCATGGCGGGAG | SEQ ID NO:4 |
| 10. CCCTATGTTGTCCAAGCTGGCA-
GAG[A/G]TTTTTGT TTTGTTTGT TTAGAGAGGA | SEQ ID NO:26 |
| 11. AGCCTGGCCAACATG[C/G]TGAAACCCCGTCTCT | SEQ ID NO:5 |

12. ctcgggaggctgaggcaggagaatc[A/G]cttgaactcaggaggcagaggttc

SEQ ID NO:27

13. AAGTTTCTCTATT[G/T]TGTTTATAAACA

SEQ ID NO:6

or to a sequence complementary to any of the subsequences.

- 5 38. The method according to claim 34, wherein at least one nucleotide primer or probe is capable of hybridising to a subsequence of SEQ ID NO: 132 selected from the group of subsequences

GGTTTAT[ATTTT]Ntgagatggatttt

SEQ ID NO:19

actaaaaataaaaaataaaaaaa[-/AA]atagccgagcatggtgggtgggtgcc

SEQ ID NO:23

AAAAAACTAAAGTGGGGTTTGC GGG[G/T]-AGTGGGAGGGCCCTTCCTGCTAGGT

SEQ ID NO:24

AAAAT-

TAGCCGG[A/G]CGCCATGGCGGGAG

SEQ ID NO:4

CCCTATGTTGTCCAAGCTGGCAGAG-[A/G]TTTTTGTGTTGTTGTTGAGAGGGA

SEQ ID NO:26

or to a sequence complementary to any of the subsequences.

- 10 39. The method according to claim 34, wherein at least one nucleotide primer or probe is capable of hybridising to a subsequence of SEQ ID NO: 132 selected from the group of subsequences

GGTTTAT[ATTTT]Ntgagatggatttt

SEQ ID NO:19

actaaaaataaaaaataaaaaaa[-/AA]atagccgagcatggtgggtgggtgcc

SEQ ID NO:23

AAAAAACTAAAGTGGGGTTTGC GGG[G/T]-

AGTGGGAGGGCCCTTCCTGCTAGGT

SEQ ID NO:24

AAAATTAGCCGG[A/G]CGCCATGGCGGGAG

SEQ ID NO:4

CCCTATGTTGTCCAAGCTGGCAGAG-

SEQ ID NO:26

[A/G]TTTTTGTGTTGTTGTTGAGAGGGA

or to a sequence complementary to any of the subsequences.

- 15 40. The method according to claim 34, wherein at least one nucleotide primer or probe is capable of hybridising to a subsequence of SEQ ID NO: 132 selected from the group of subsequences

GGTTTAT[ATTTT]Ntgagatggatttt

SEQ ID NO:19

ctggggaggctgaggcagga-

gaatc[A/G]cttgaaaccgggaggcggaggtgt

SEQ ID NO:22

or to a sequence complementary to any of the subsequences.

41. The method according to claim 34, wherein at least one nucleotide primer or probe is capable of hybridising to a subsequence of SEQ ID NO: 132 selected from the group of subsequences

CACAGTGAAAC[C/T]CCATCTCTACTAAA

SEQ ID NO: 3

AAAATTAGCCGG[A/G]CGCCATGGCGGGAG

SEQ ID NO: 4

5 or to a sequence complementary to any of the subsequences.

42. The method according to claim 34, wherein the at least one nucleotide primer or probe capable of hybridising to a subsequence of SEQ ID NO: 132 is

AGCCTGGCCAACATG[C/G]TGAAACCCCGTCTCT

SEQ ID NO: 5

or to a sequence complementary to the subsequence.

10

43. The method according to claim 34, wherein the at least one nucleotide primer or probe capable of hybridising to a subsequence of SEQ ID NO: 132 is

AAGTTTCTCTATT[G/T]TGTTTATAAACA

SEQ ID NO: 6

or to a sequence complementary to the subsequence.

15

44. The method according to claim 34, wherein at least one nucleotide primer or probe is capable of hybridising to a subsequence of SEQ ID NO: 133 selected from the group of subsequences

AGAACCTGTTTCAGGCTGGCGGCTCA[C/T]TTGGATGAACAGGGAGTGTGTGAC
CCCCCTTCTTAGGACGCATGGGGGT[G/T]GAGAGAACGGGGAGATAGACAGAG

SEQ ID NO: 41

SEQ ID NO: 42

TGCGAGCAGCCCGGGCTACAGGGTT[A/G]CCTGAGGTGTGGGTCCAGGATGG

SEQ ID NO: 43

GGCGCCTCAACAGCCAGAAGGAGCG[A/G]AGCCTCAGGCCAGGCAGCTCTGG

SEQ ID NO: 44

AGAAAGAAAAACAGCAA A/G ATGCCACAGTGGAGCCAGAG

SEQ ID NO: 142

or to a sequence complementary to any of the subsequences.

20

45. The method according to claim 34, wherein at least one nucleotide primer or probe is capable of hybridising to a subsequence of SEQ ID NO: 133 selected from the group of subsequences

CCCCCTTCTTAGGACGCATGGGGT[G/T]GAGAGAACGGGGAGATAGACAGAG SEQ ID NO:42

TGCGAGCAGCCCGGGCTACAGGGTT[A/G]CCTGAGGTGTGGGTCCCAGGATGG SEQ ID NO:43

GGCGCCTCAACAGCCAGAAGGAGCG[A/G]AGCCTCAGGCCAGGCAGCTCTGG SEQ ID NO:44

AGAAAGAAAAACAGCAA A/G ATGCCACAGTGGAGCCAGAG SEQ ID NO: 142

5 or to a sequence complementary to any of the subsequences.

46. The method according to claim 34, wherein at least one nucleotide primer or probe is capable of hybridising to a subsequence of SEQ ID NO: 133 selected from the group of subsequences

AGAACCTGTTTCAGGCTGGCGGCTCA[C/T]TTGGATGAACAGGGAGTGTGTGAC SEQ ID NO:41
TGCGAGCAGCCCGGGCTACAGGGTT[A/G]CCTGAGGTGTGGGTCCCAGGATGG SEQ ID NO:43

GGCGCCTCAACAGCCAGAAGGAGCG[A/G]AGCCTCAGGCCAGGCAGCTCTGG SEQ ID NO:44

AGAAAGAAAAACAGCAA A/G ATGCCACAGTGGAGCCAGAG SEQ ID NO: 142

10

or to a sequence complementary to any of the subsequences.

47. The method according to claim 34, wherein at least one nucleotide primer or probe is capable of hybridising to a subsequence of SEQ ID NO: 133 selected from the group of subsequences

15

TGCGAGCAGCCCGGGCTACAGGGTT[A/G]CCTGAGGTGTGGGTCCCAGGATGG SEQ ID NO: 43

AGAAAGAAAAACAGCAA A/G ATGCCACAGTGGAGCCAGAG

SEQ ID NO: 142

or to a sequence complementary to any of the subsequences.

48. The method according to claim 34, wherein the at least one nucleotide primer or probe capable of hybridising to the subsequence of SEQ ID NO: 133 is

GGCGCCTCAACAGCCAGAAAGGAGCG[A/G]AGCCTCAGGCCAGGCAGCTCTGG

SEQ ID NO: 44

5

or to a sequence complementary to the subsequence.

49. The method according to claim 34, wherein the at least one nucleotide primer or probe capable of hybridising to the subsequence of SEQ ID NO: 133 is

AGAAAGAAAAACAGCAA A/G ATGCCACAGTGGAGCCAGAG

SEQ ID NO:
142

10

or to a sequence complementary to the subsequence.

50. The method according to any of the preceding claims, wherein at least two different probes are used, one or more probes being selected from the probes as defined in any of claims 34-49, and one or more probes being capable of hybridising to a sequence different from SEQ ID NO: 132 and optionally SEQ ID NO: 133, or a part thereof, or to a sequence complementary to a region different from SEQ ID NO: 132 and optionally SEQ ID NO: 133, or a part thereof.

15

51. The method according to claim 1, wherein the translational product from a sequence in a region corresponding to SEQ ID NO: 132 and optionally SEQ ID NO: 133, or a part thereof, is an antibody, such as a monoclonal or polyclonal antibody.

20

52. A method for estimating the disease prognosis of an individual comprising

- in a sample from said individual assessing in the genetic material at least one sequence polymorphism
- in a region corresponding to SEQ ID NO: 132, or a part thereof, or
- in a region complementary to SEQ ID NO: 132, or a part thereof, or

25

- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 132, or a part thereof, or
- or translation product from a sequence in a region corresponding to SEQ ID NO: 132, or a part thereof, and optionally a further sequence polymorphism
- 5 - in a region corresponding to SEQ ID NO: 133, or a part thereof, or
- in a region complementary to SEQ ID NO: 133, or a part thereof, or
- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof, or
- 10 - or translation product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof,

- with the proviso that when using SEQ ID NO:132 only the at least one sequence polymorphism is in a region corresponding to at least one region within SEQ ID NO:132 selected from the regions flanked by and including
- 15 the sequence polymorphism RAI3'd1 (SEQ ID NO: 19) and rs2377329 (SEQ ID NO: 22), or

in a region flanked by and including the sequence polymorphism positioned at 18149120 in Contig NT_011109 (SEQ ID NO: 3) and sequence polymorphism positioned at 18149154 in Contig NT_011109 (SEQ ID NO: 4), or

20

the sequence polymorphism positioned at 18150815 in Contig NT_011109 (SEQ ID NO: 5), or

25 the sequence polymorphism positioned at 18151158 in Contig NT_011109 (SEQ ID NO: 6)

- obtaining a sequence polymorphism response,

- 30 - estimating the disease prognosis of said individual based on the sequence polymorphism response.

53. The method according to claim 52, wherein the method has any of the features as defined in any of the claims 2-51.

54. A method for estimating a treatment response of an individual suffering from cancer to a disease treatment, comprising
in a sample from said individual assessing in the genetic material at least one sequence polymorphism

5 in a region corresponding to SEQ ID NO: 132, or a part thereof, or

- in a region complementary to SEQ ID NO: 132, or a part thereof, or
- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 132, or a part thereof, or

10 or translation product from a sequence in a region corresponding to SEQ ID NO:132, or a part thereof, and optionally a further sequence polymorphism

in a region corresponding to SEQ ID NO: 133, or a part thereof, or

- in a region complementary to SEQ ID NO: 133, or a part thereof, or
- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof, or

15 - or translation product from a sequence in a region corresponding to SEQ ID NO: 133, or a part thereof,

- with the proviso that when using SEQ ID NO:132 only the at least one sequence polymorphism is in a region corresponding to at least one region within SEQ ID NO:132 selected from the regions flanked by and including the sequence polymorphism RAI3'd1 (SEQ ID NO:19) and rs2377329 (SEQ ID NO: 22), or

20 in a region flanked by and including the sequence polymorphism positioned at 18149120 in Contig NT_011109 (SEQ ID NO: 3) and sequence polymorphism positioned at 18149154 in Contig NT_011109 (SEQ ID NO: 4), or

25 the sequence polymorphism positioned at 18150815 in Contig NT_011109 (SEQ ID NO: 5), or

30 the sequence polymorphism positioned at 18151158 in Contig NT_011109 (SEQ ID NO: 6)

- obtaining a sequence polymorphism response,

- estimating the individual's response to the disease treatment based on the sequence polymorphism response.

55. The method according to claim 54, wherein the method has any of the features
5 as defined in any of the claims 2-51.

56. A primer or probe for use in a method as defined in any of the claims above,
said primer or probe being selected from the group of nucleotides consisting of
agt gca gcc tca act tcc (SEQ ID NO: 96)
10 cca gtc caa aca ata tga tcc (SEQ ID NO: 97)
AAA AAA ATA GCC GAG CAT GG (SEQ ID NO: 134)
TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)
cat gat tca ctg cac cca acc (SEQ ID NO: 98)
ttt cac tct tgt tgc cca agc (SEQ ID NO: 99)
15 ttt tca cac agg tcc aat cc (SEQ ID NO: 104)
act gca acc tcc atc tcc (SEQ ID NO: 105)
atg ttg ggg aga ctg agg (SEQ ID NO: 124)
ccg cat cta act tat tct gg (SEQ ID NO: 125)
aac tac ctc tgc aaa ccc agc (SEQ ID NO: 126)
20 ttg gaa tgg agg gat tct acc (SEQ ID NO: 127)
ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
gga cag atg gca atg atg g (SEQ ID NO: 130)
tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)
25 acc atg gcg cct caa ca (SEQ ID NO: 136)
gaa ttg gct cag tca ctg tgt ga (SEQ ID NO: 137)
55. The primer or probe for use in a method as defined in any of the claims above,
said primer or probe being selected from the group of nucleotides consisting of
agt gca gcc tca act tcc (SEQ ID NO: 96)
30 cca gtc caa aca ata tga tcc (SEQ ID NO: 97)
AAA AAA ATA GCC GAG CAT GG (SEQ ID NO: 134)
TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)
cat gat tca ctg cac cca acc (SEQ ID NO: 98)
ttt cac tct tgt tgc cca agc (SEQ ID NO: 99)
35 ggt ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
gga cag atg gca atg atg g (SEQ ID NO: 130)
tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)
acc atg gcg cct caa ca (SEQ ID NO: 136)
5 gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:137)

57. The primer or probe for use in a method as defined in any of the claims above,
said primer or probe being selected from the group of nucleotides consisting of
agt gca gcc tca act tcc (SEQ ID NO: 96)

10 cca gtc caa aca ata tga tcc (SEQ ID NO: 97)
AAA AAA ATA GCC GAG CAT GG (SEQ ID NO: 134)
TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)
cat gat tca ctg cac cca acc (SEQ ID NO: 98)
ttt cac tct tgt tgc cca agc (SEQ ID NO: 99)
15 ggt ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
gga cag atg gca atg atg g (SEQ ID NO: 130)
tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)
acc atg gcg cct caa ca (SEQ ID NO: 136)
20 gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:137)

58. The primer or probe for use in a method as defined in any of the claims above,
said primer or probe being selected from the group of nucleotides consisting of
agt gca gcc tca act tcc (SEQ ID NO: 96)

25 cca gtc caa aca ata tga tcc (SEQ ID NO: 97)
AAA AAA ATA GCC GAG CAT GG (SEQ ID NO: 134)
TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)
ggt ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
30 acc atg gcg cct caa ca (SEQ ID NO: 136)
gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:137)

59. The primer or probe for use in a method as defined in any of the claims above,
said primer or probe being selected from the group of nucleotides consisting of

35 AAA AAA ATA GCC GAG CAT GG (SEQ ID NO: 134)

TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)

ggt ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

- 5 60. The primer or probe for use in a method as defined in any of the claims above,
said primer or probe being selected from the group of nucleotides consisting of

AAA AAA ATA GCC GAG CAT GG (SEQ ID NO: 134)

TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)

acc atg gcg cct caa ca (SEQ ID NO: 136)

- 10 gaa ttg gct cag tca ctg tgt ga (SEQ ID NO: 137)

61. The primer or probe for use in a method as defined in any of the claims above,
said primer or probe being selected from the group of nucleotides consisting of

AAA AAA ATA GCC GAG CAT GG (SEQ ID NO: 134)

- 15 TT TGG ACT GGG TAA GAA TTT CC (SEQ ID NO: 135)

ggt ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

acc atg gcg cct caa ca (SEQ ID NO: 136)

gaa ttg gct cag tca ctg tgt ga (SEQ ID NO: 137)

20

62. A primer or probe for use in a method as defined in any of the claims above as
the other probe said primer or probe being selected from the group consisting of

agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)

ggt tcc gcc acg ttg cc (SEQ ID NO: 54)

- 25 tcg gct att ttt ttt att ttt tta tt (SEQ ID NO: 55)

att aca ggc acc cac cac cat g (SEQ ID NO: 56)

cct gga caa cat agg gag acc ctg tgt (SEQ ID NO: 138)

caa aca aac aaa aac ctc tgc ca (SEQ ID NO: 139)

atg ttg ggg aga ctg agg (SEQ ID NO: 124)

- 30 ccg cat cta act tat tct gg (SEQ ID NO: 125)

aac tac ctc tgc aaa ccc agc (SEQ ID NO: 126)

ttg gaa tgg agg gat tct acc (SEQ ID NO: 127)

ggt ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

- 35 gga cag atg gca atg atg g (SEQ ID NO: 130)

tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)
acc atg gcg cct caa ca (SEQ ID NO: 140)
gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:141)

5 63. The primer or probe according to claim 61 said primer or probe being selected from the group consisting of

agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)
ggg tcc gcc acg ttg cc (SEQ ID NO: 54)
tcg gct att ttt ttt att ttt tta tt (SEQ ID NO: 55)

10 att aca ggc acc cac cac cat g (SEQ ID NO: 56)
ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
acc atg gcg cct caa ca (SEQ ID NO: 140)
gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:141)

15

64. The primer or probe according to claim 61 said primer or probe being selected from the group consisting

tcg gct att ttt ttt ttt att ttt tta tt (SEQ ID NO: 55)
att aca ggc acc cac cac cat g (SEQ ID NO: 56)

20 ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
gga cag atg gca atg atg g (SEQ ID NO: 130)
tct tct tct tgg tgg atg tgg (SEQ ID NO: 131)
acc atg gcg cct caa ca (SEQ ID NO: 140)
25 gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:141)

65. The primer or probe according to claim 61 said primer or probe being selected from the group consisting of

agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)

30 ggg tcc gcc acg ttg cc (SEQ ID NO: 54)
ggg ttt ctg ctc tgc aca cg (SEQ ID NO: 128)
cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)
acc atg gcg cct caa ca (SEQ ID NO: 140)
gaa ttg gct cag tca ctg tgt ga (SEQ ID NO:141)

35

66. The primer or probe according to claim 61 said primer or probe being selected from the group consisting of

agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)

ggt tcc gcc acg ttg cc (SEQ ID NO: 54)

5 acc atg gcg cct caa ca (SEQ ID NO: 140)

gaa ttg gct cag tca ctg tgt ga (SEQ ID NO: 141)

67. The primer or probe according to claim 61 said primer or probe being selected from the group consisting of

10 agg ctg gtc ttg aac tcc tgg gct taa g (SEQ ID NO: 53)

ggt tcc gcc acg ttg cc (SEQ ID NO: 54)

ggt ttt ctg ctc tgc aca cg (SEQ ID NO: 128)

cct ttc tcc ttc cac caa cg (SEQ ID NO: 129)

15 68. The primer or probe according to any of claims 55 or 62, wherein the probe is operably linked to at least one label, such as operably linked to two different labels.

69. The primer or probe according to claim 68, wherein the label is selected from
TEX, TET, TAM, ROX, R6G, ORG, HEX, FLU, FAM, DABSYL, Cy7, Cy5, Cy3,
20 BOFL, BOF, BO-X, BO-TRX, BO-TMR, JOE, 6JOE, VIC, 6FAM, LCRed640,
LCRed705, TAMRA, Biotin, Digoxigenin, DuO-family, Daq-family.

70. The primer or probe according to claim 68, wherein the primer or probe is operably linked to a surface.

25

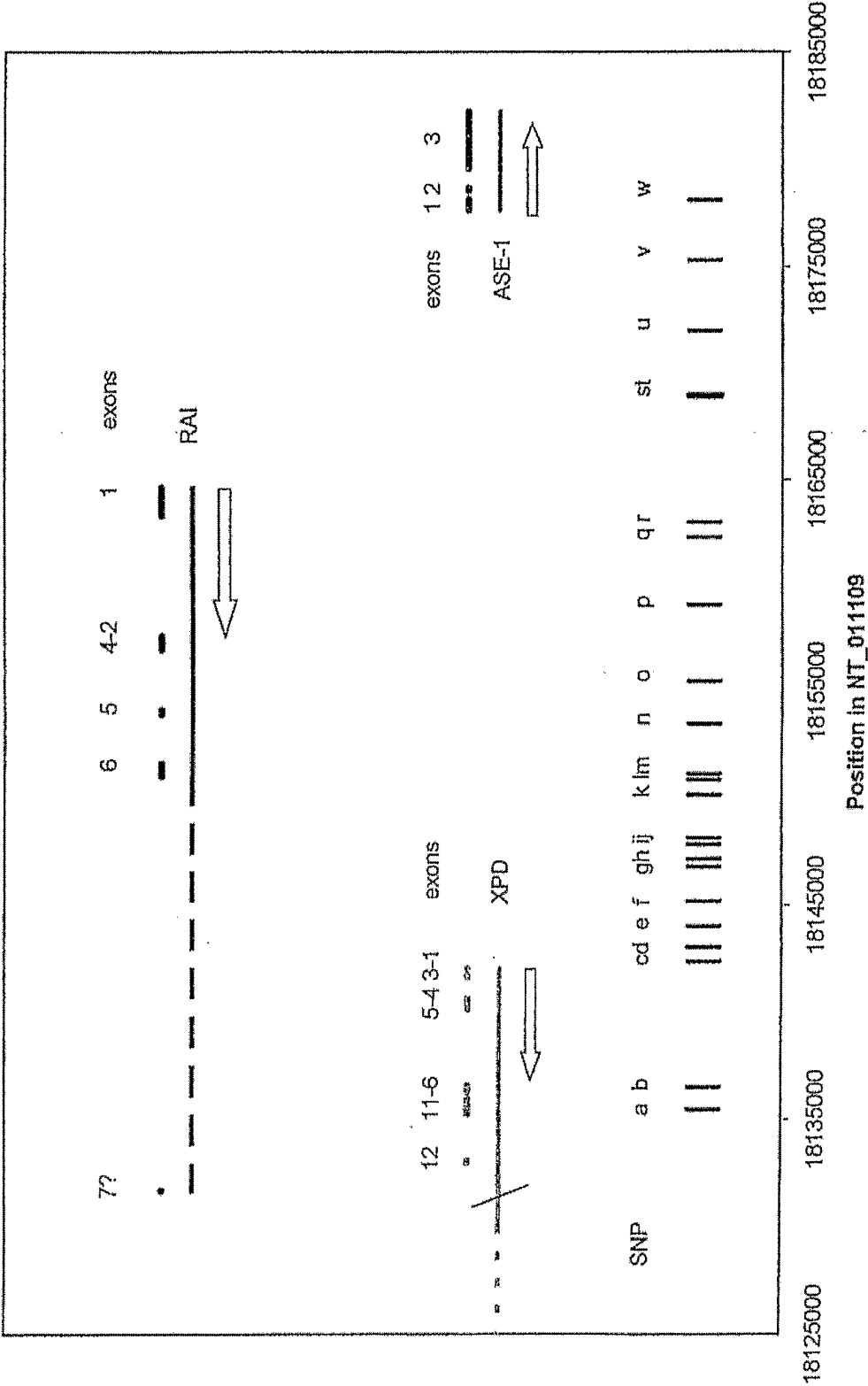
71. The primer or probe according to claim 70, wherein the surface is the surface of microbeads or a DNA chip.

72. An antibody directed to an epitope of a RAI gene product.

30

Figure 1

The genes RAI and ASE-1 and part of XPD



2/17

Figure 2

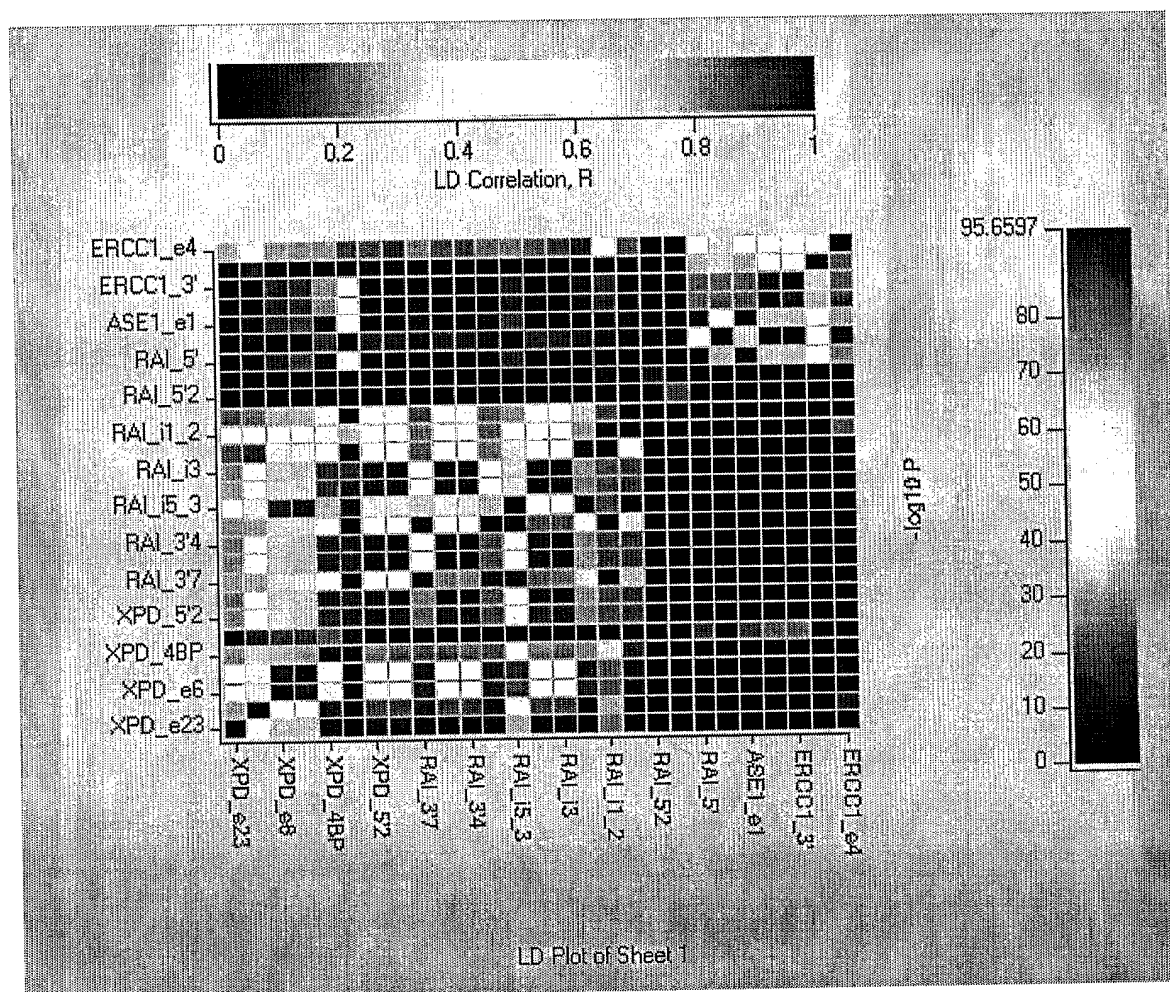
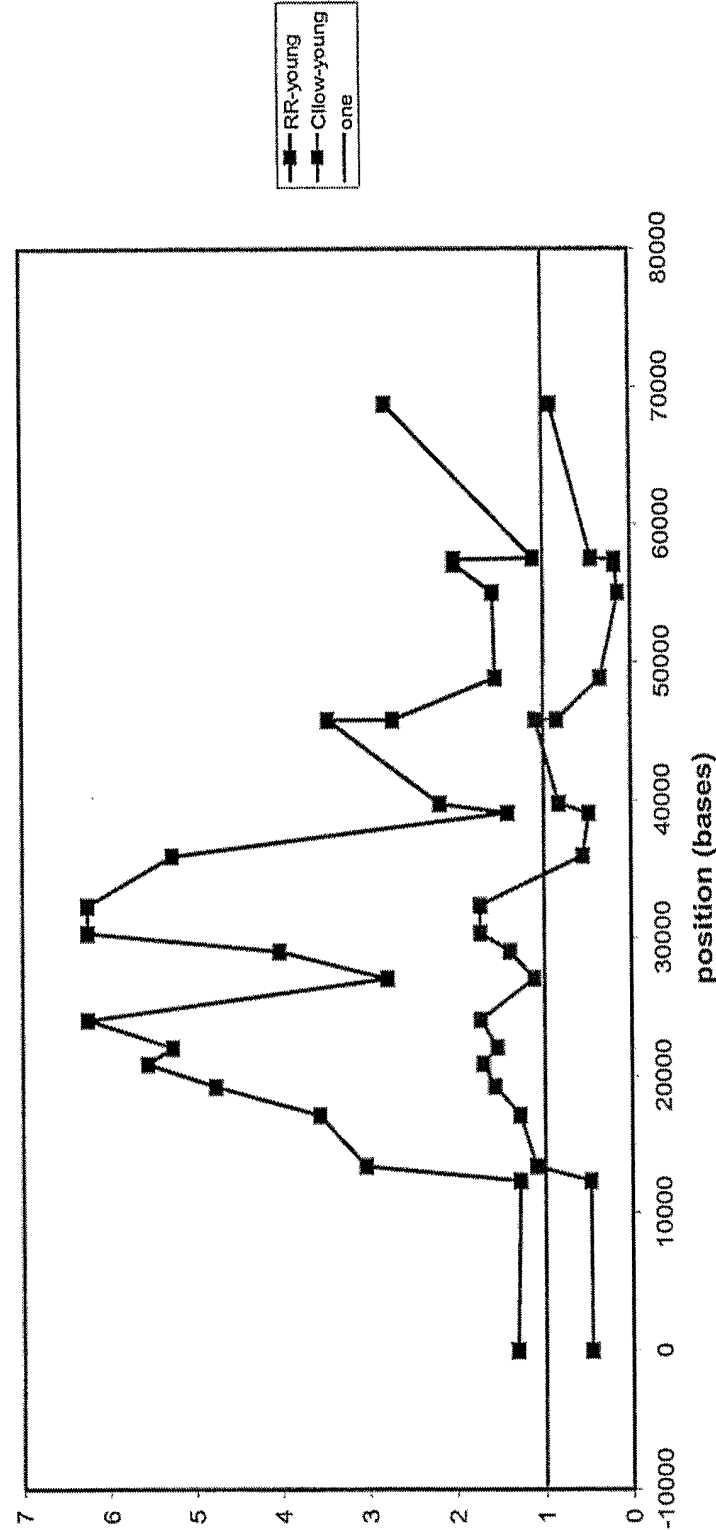


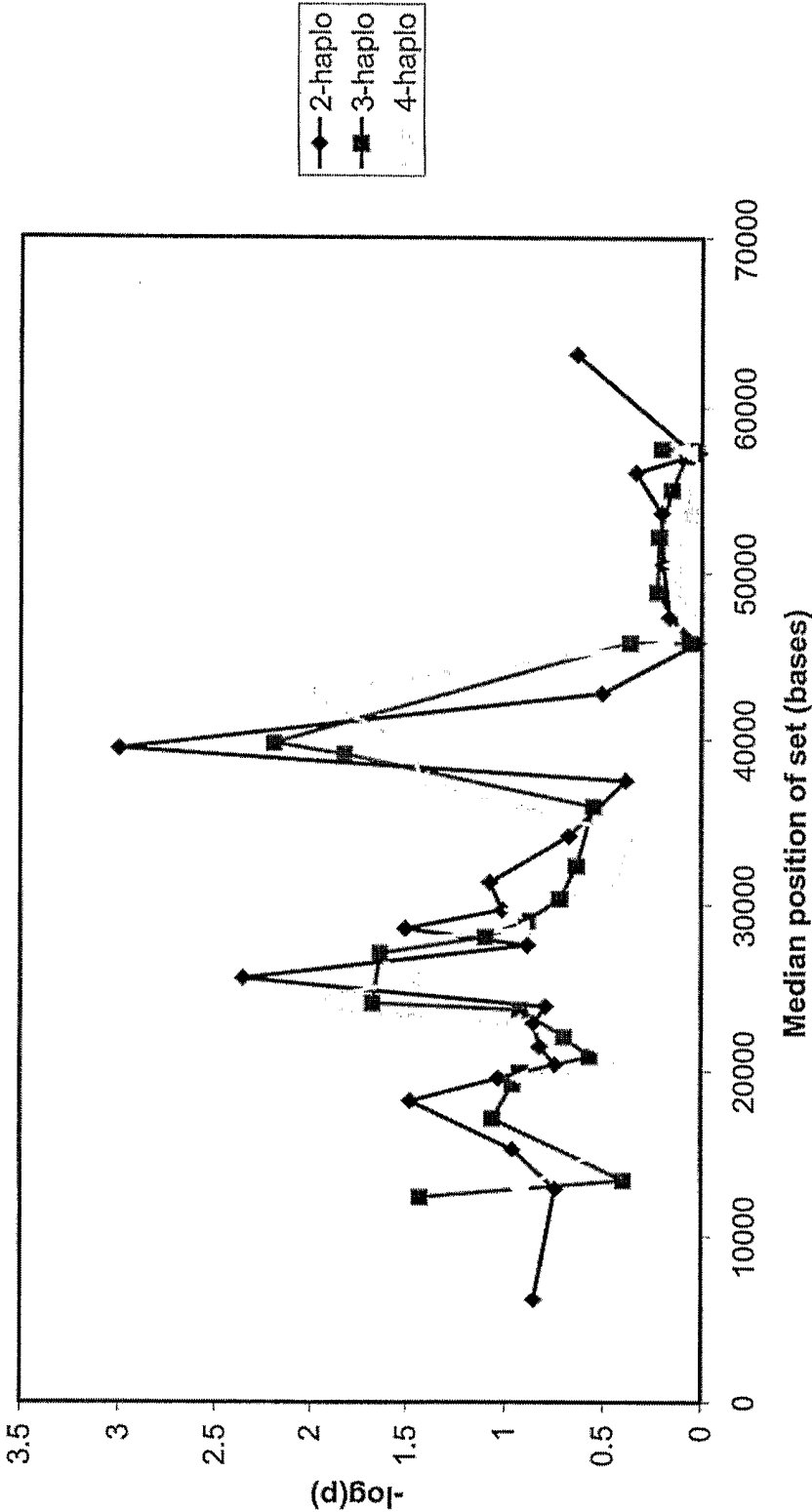
Figure 3

Relative risk and lower confidence limit for different markers in relation to early breast cancer



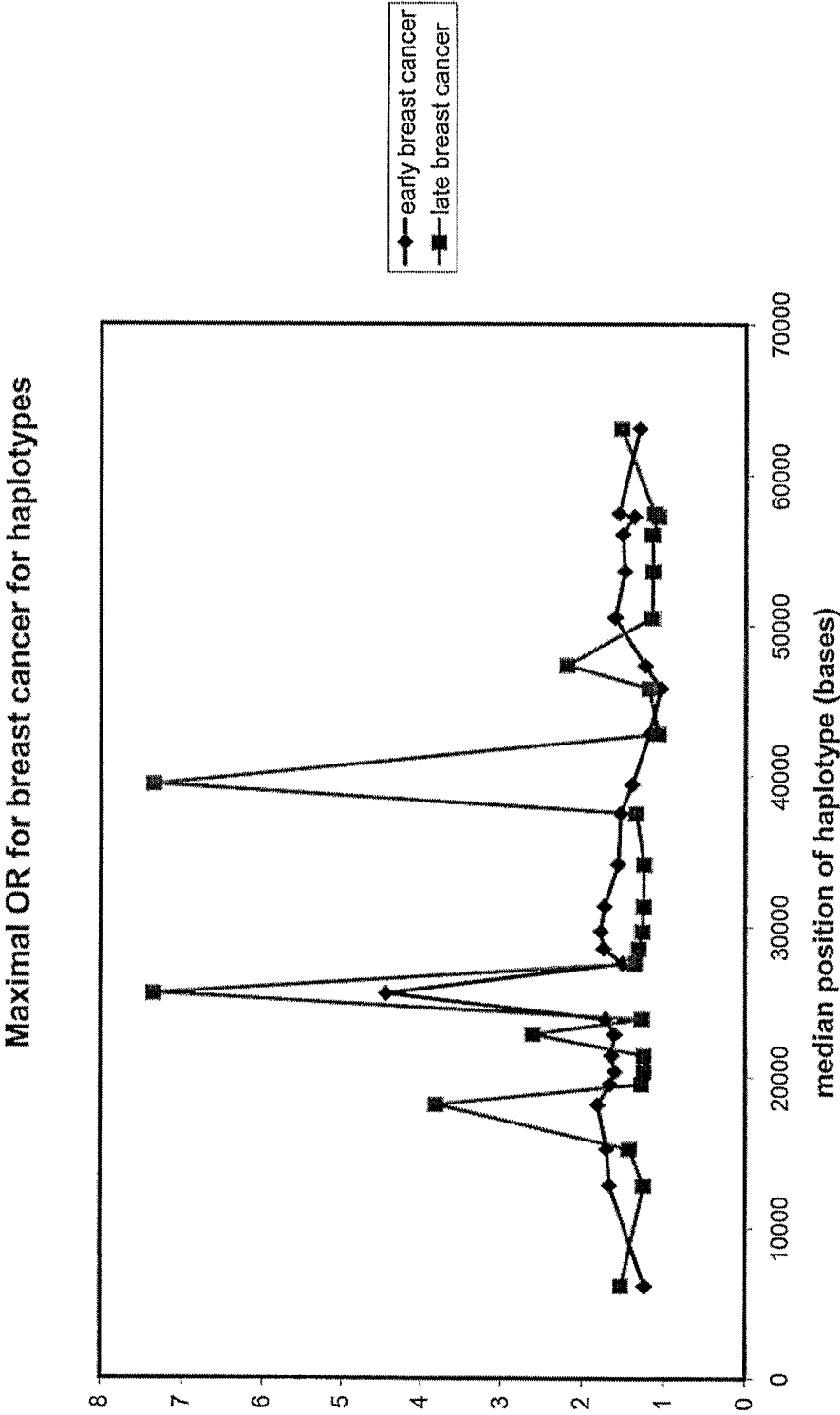
4/17

Figure 4
Association of breast cancer with sets of neighboring SNPs



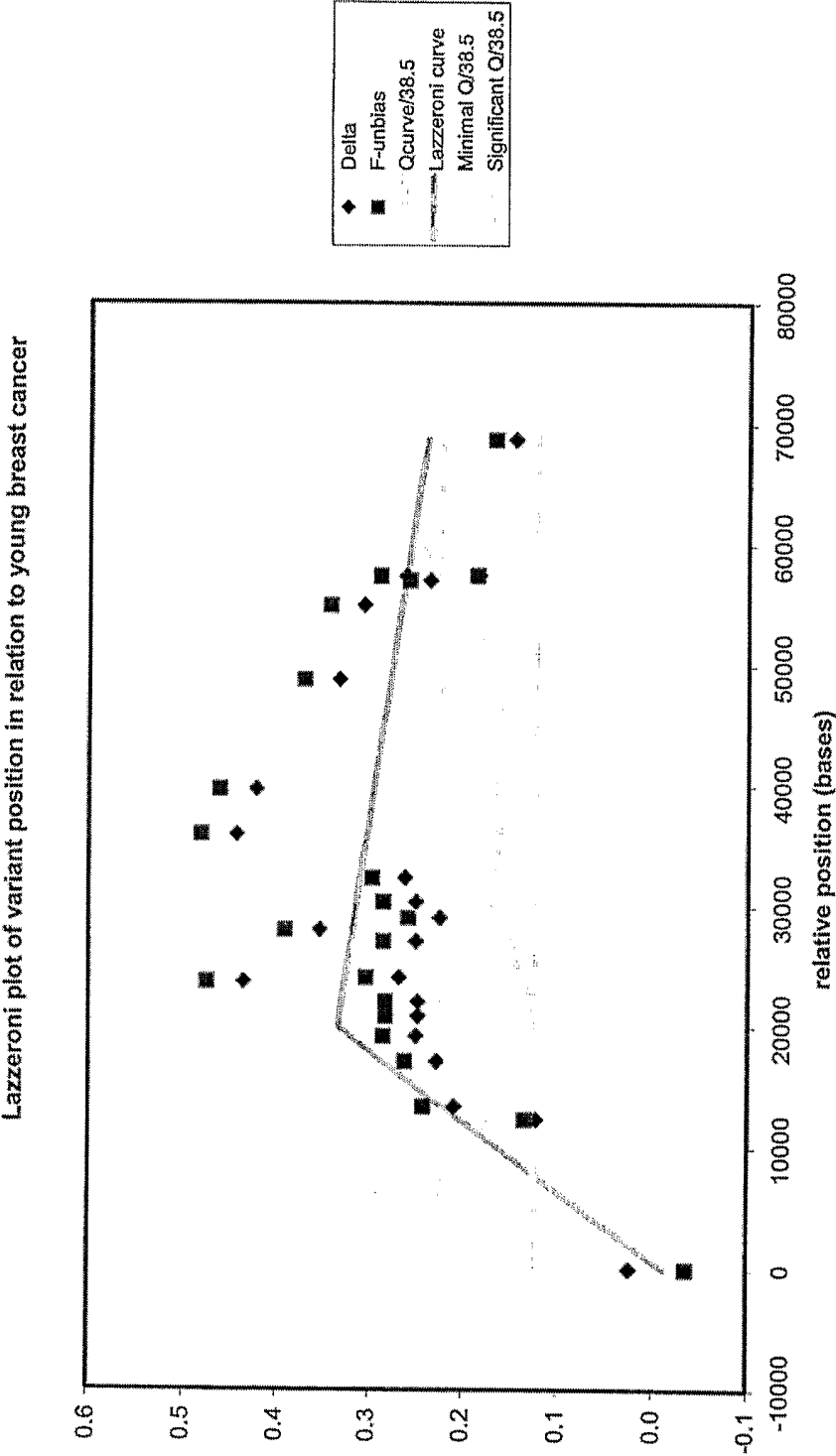
5/17

Figure 5



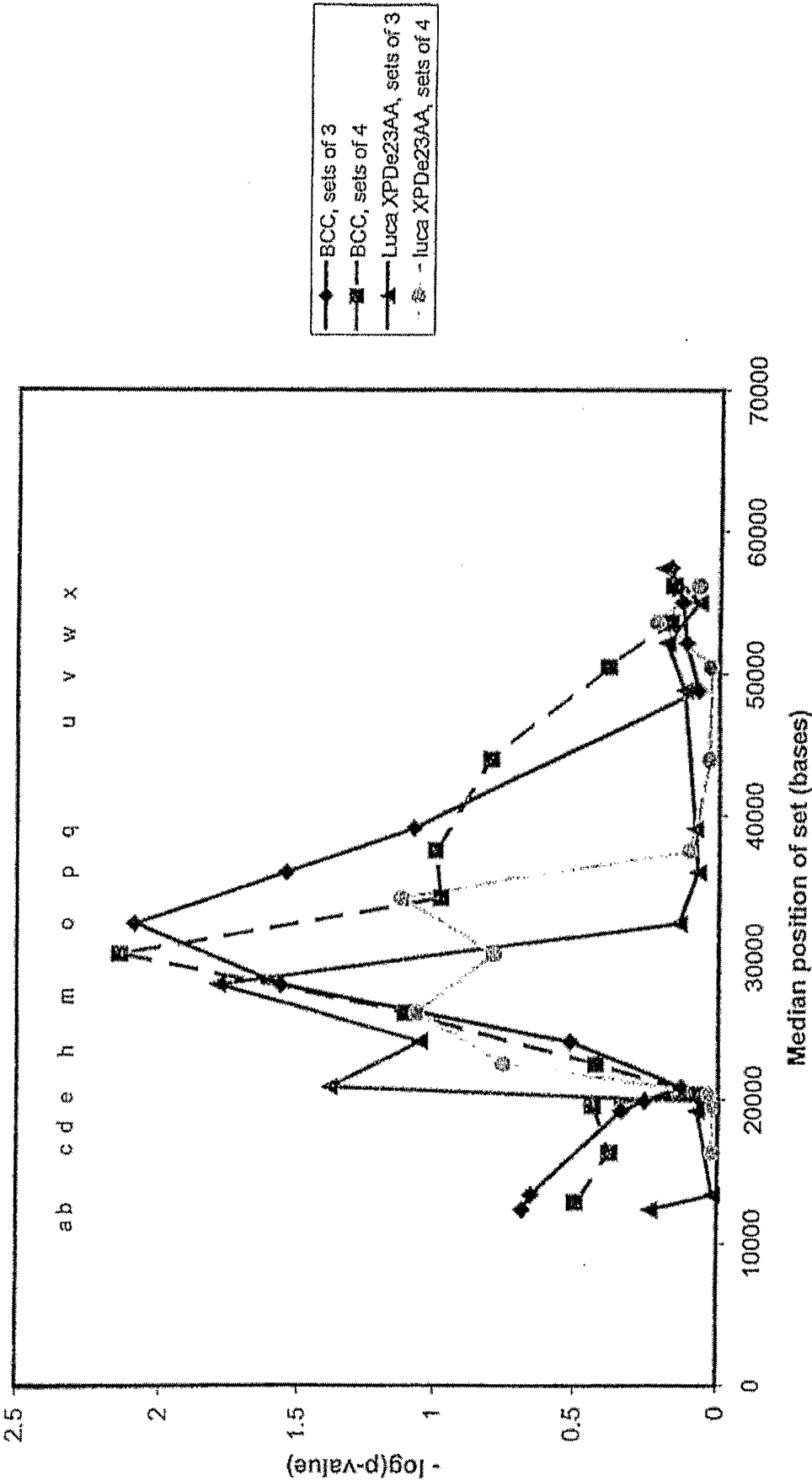
6/17

Figure 6



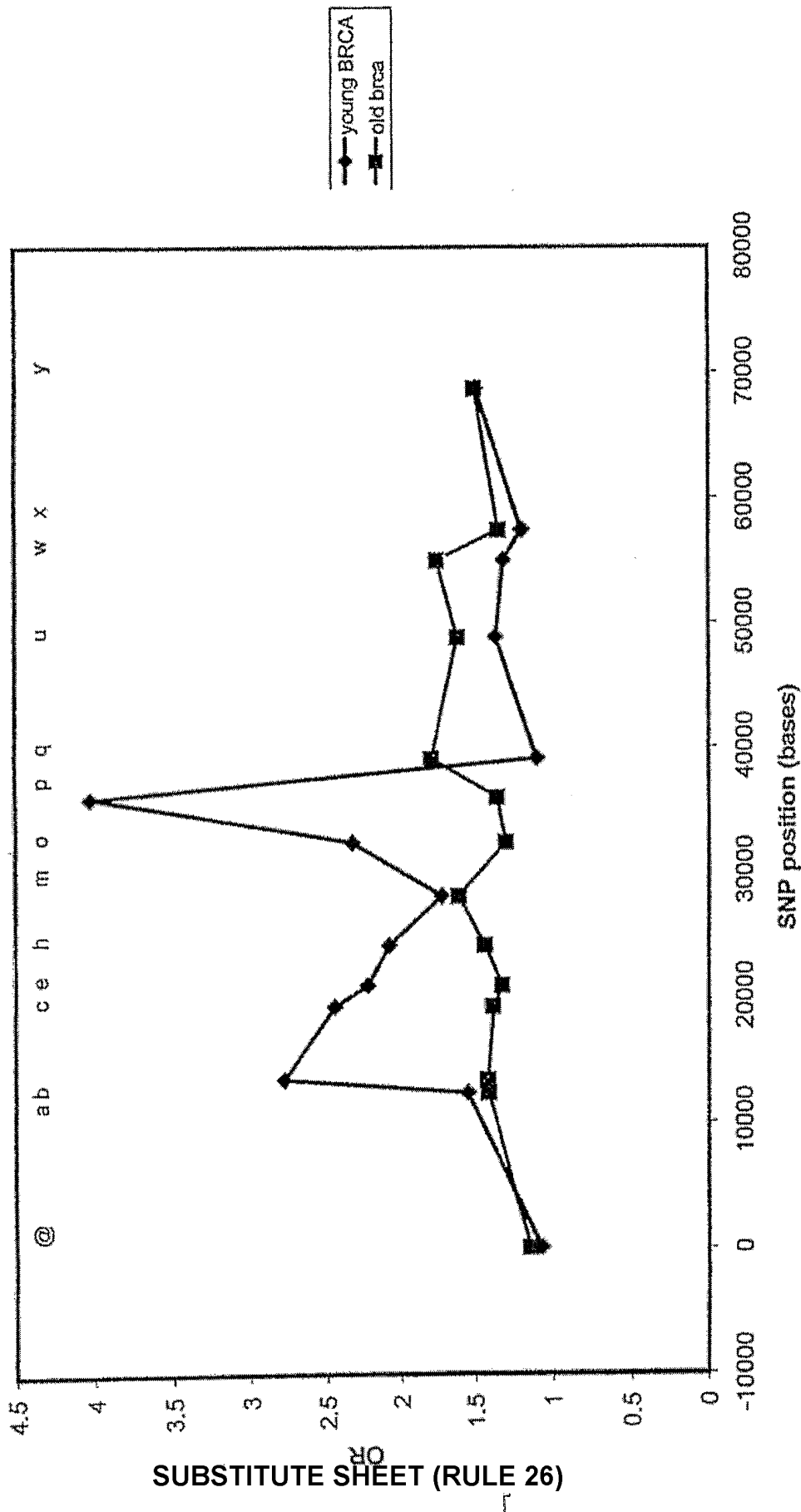
7/17

Figure 7 Overall association of cancer with sets of neighboring markers among young persons



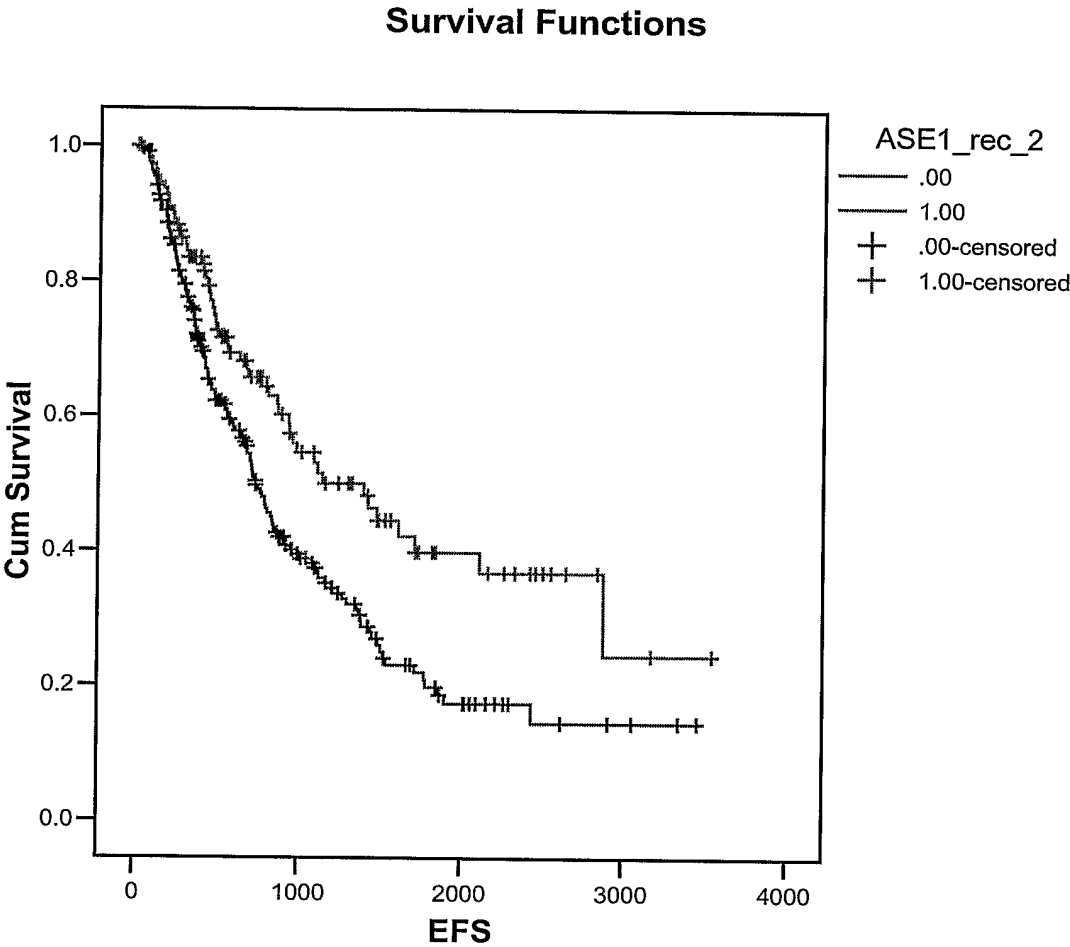
8/17

Figure 8
Oddsratio for cancer vs. control between the 2 homozygotes of each SNP in relation to location



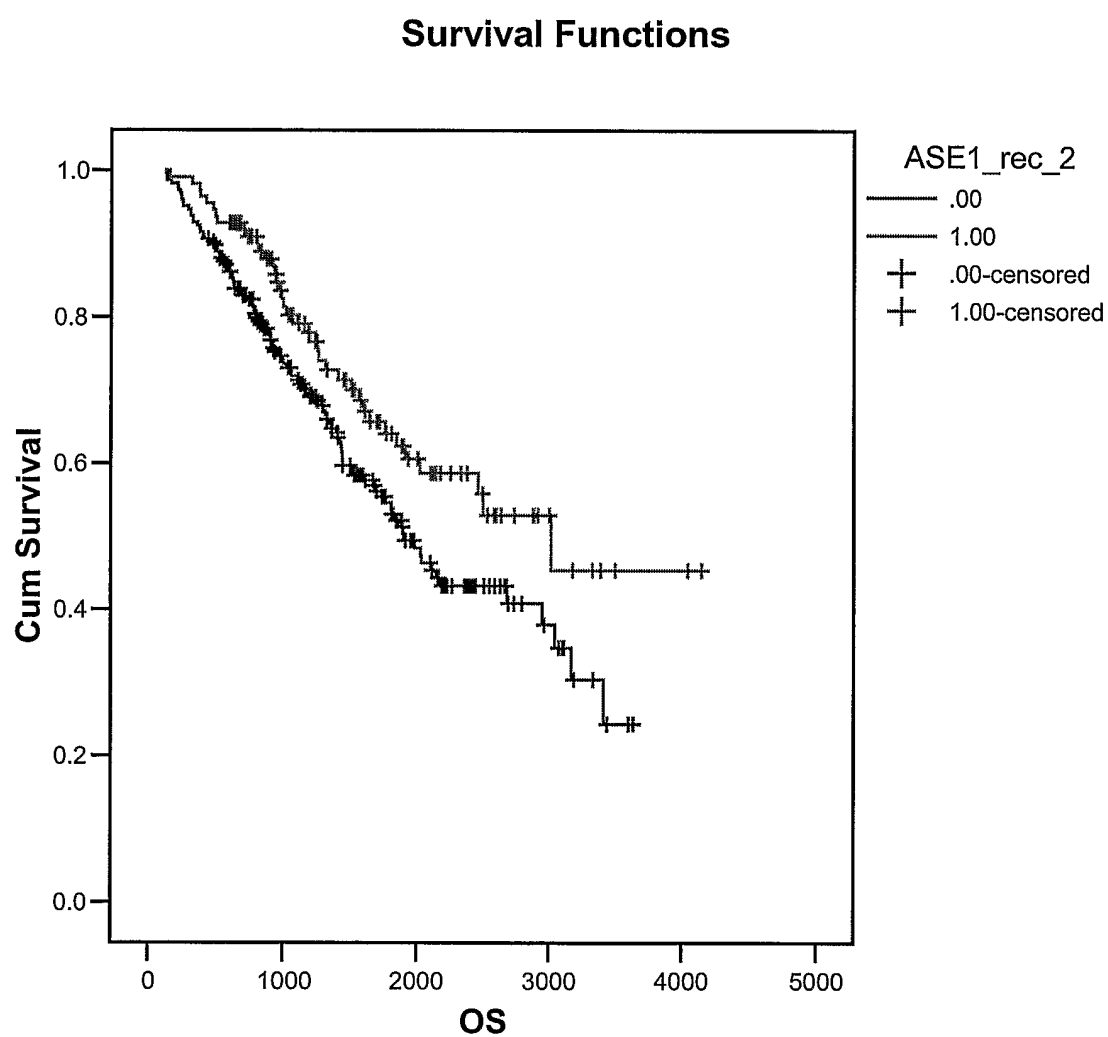
9/17

Figure 9



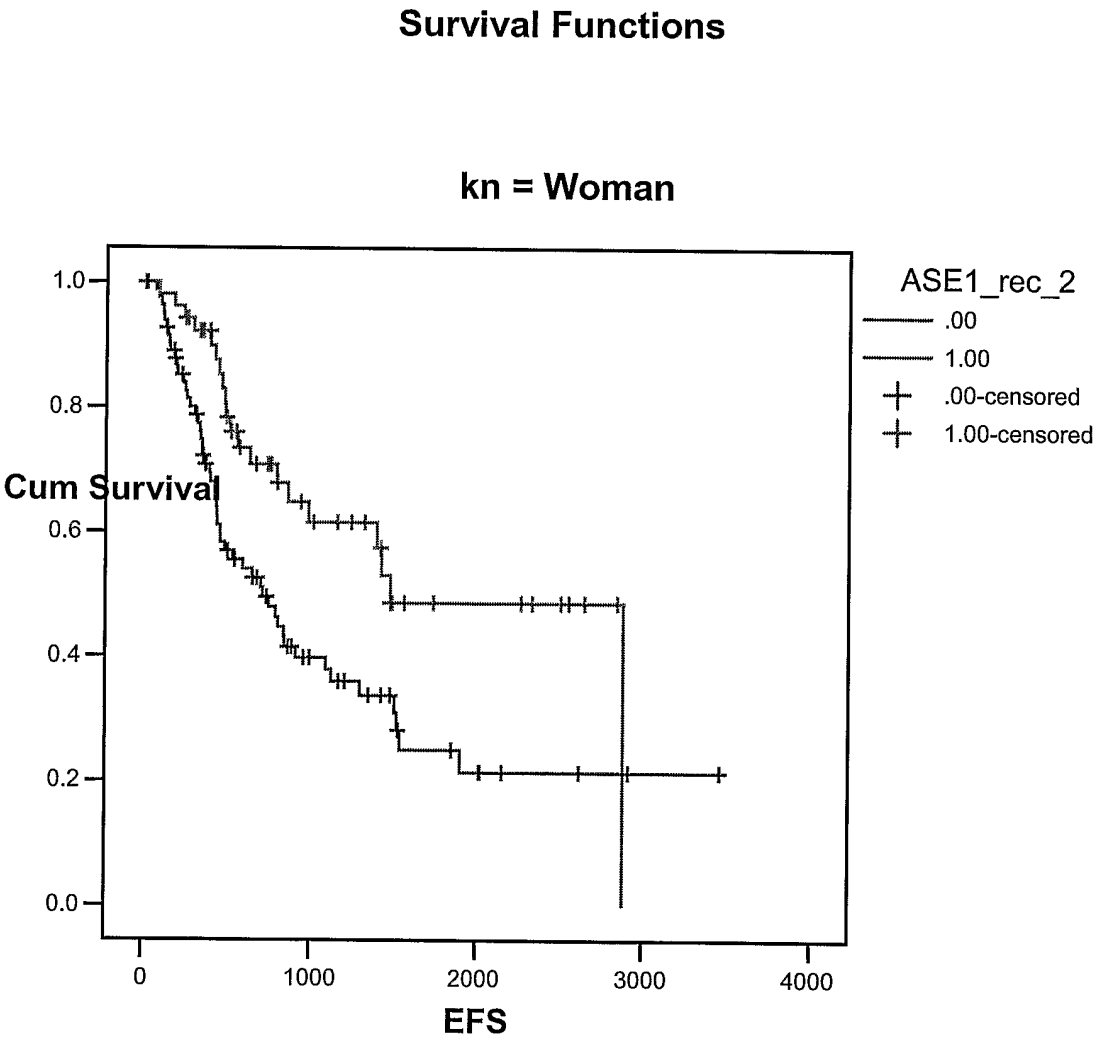
10/17

Figure 10



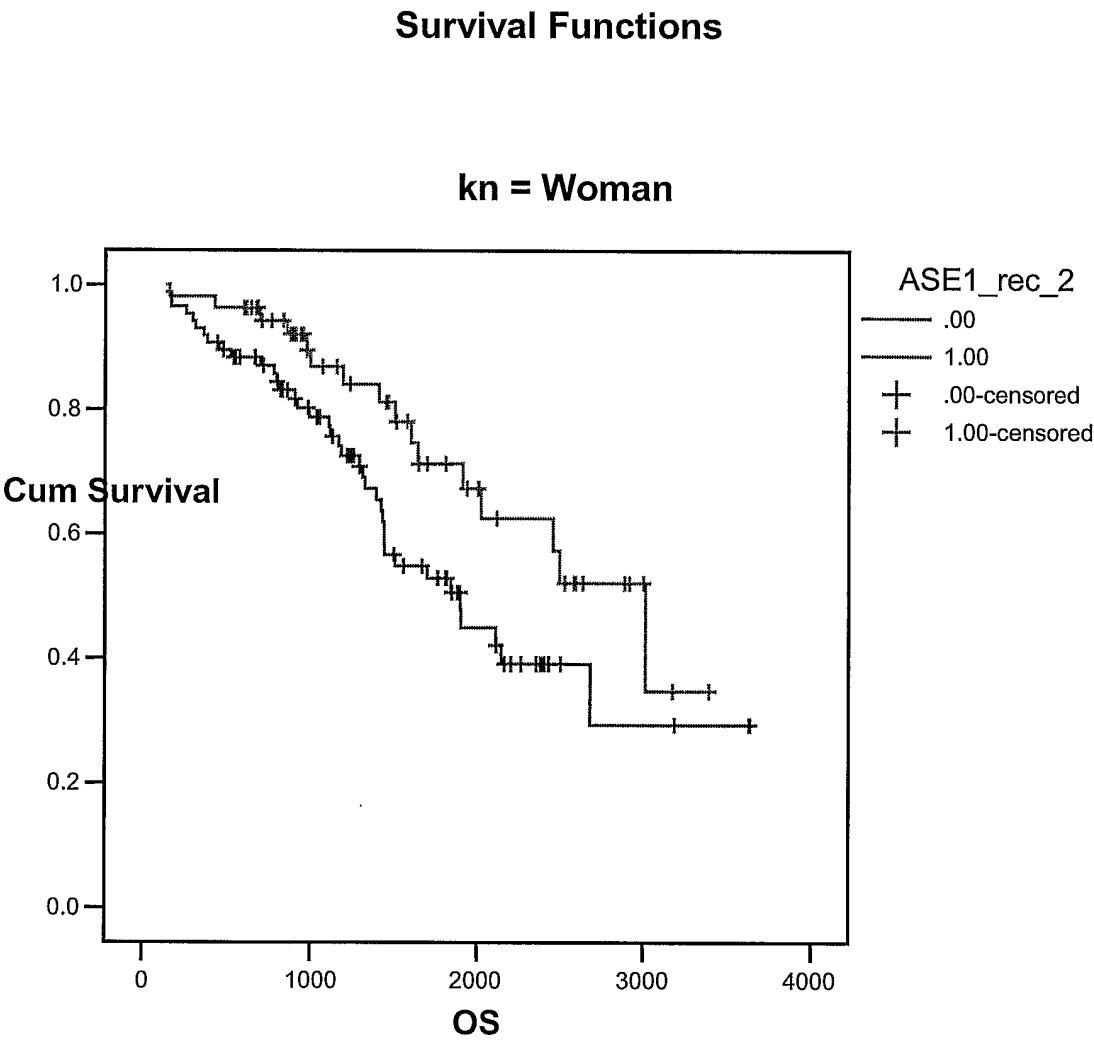
11/17

Figure 11



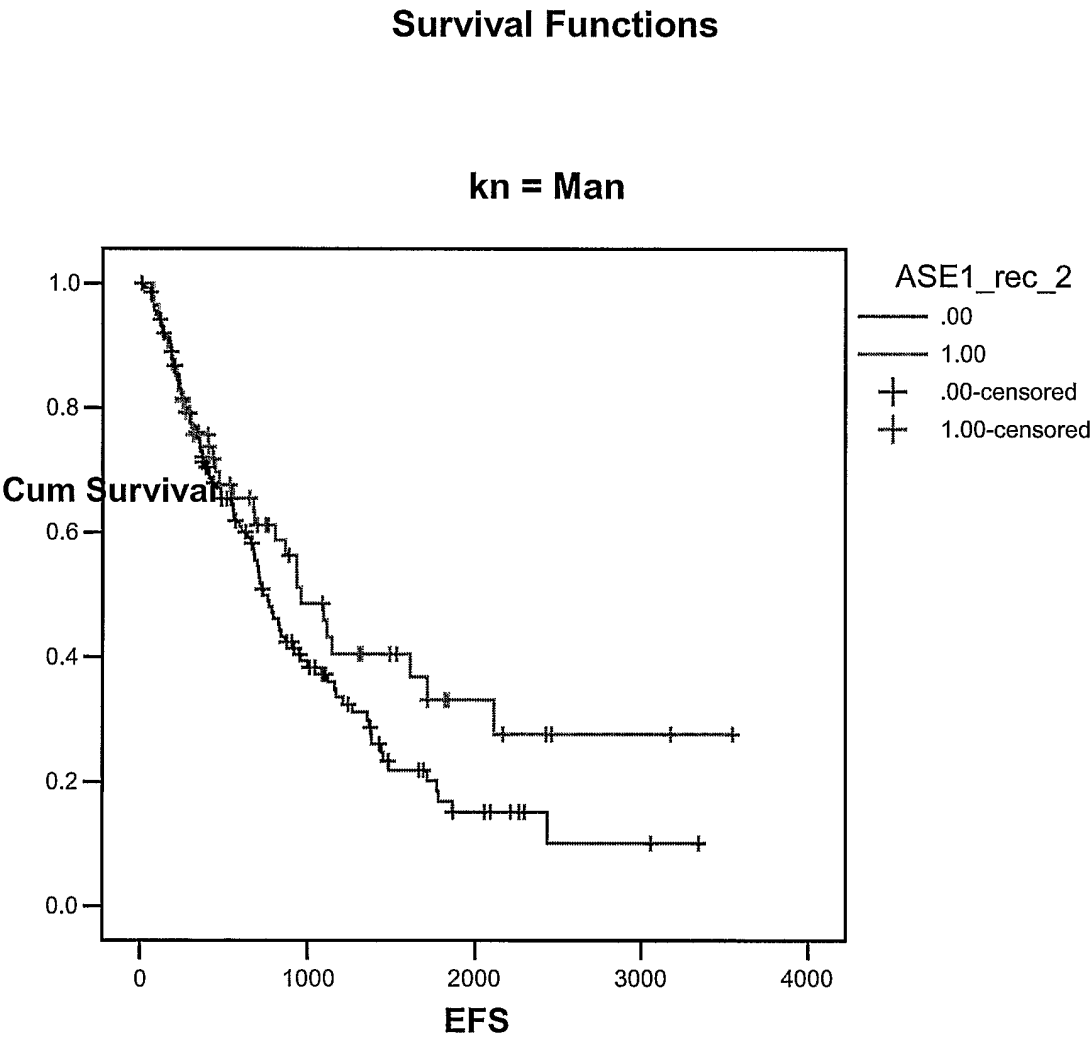
12/17

Figure 12



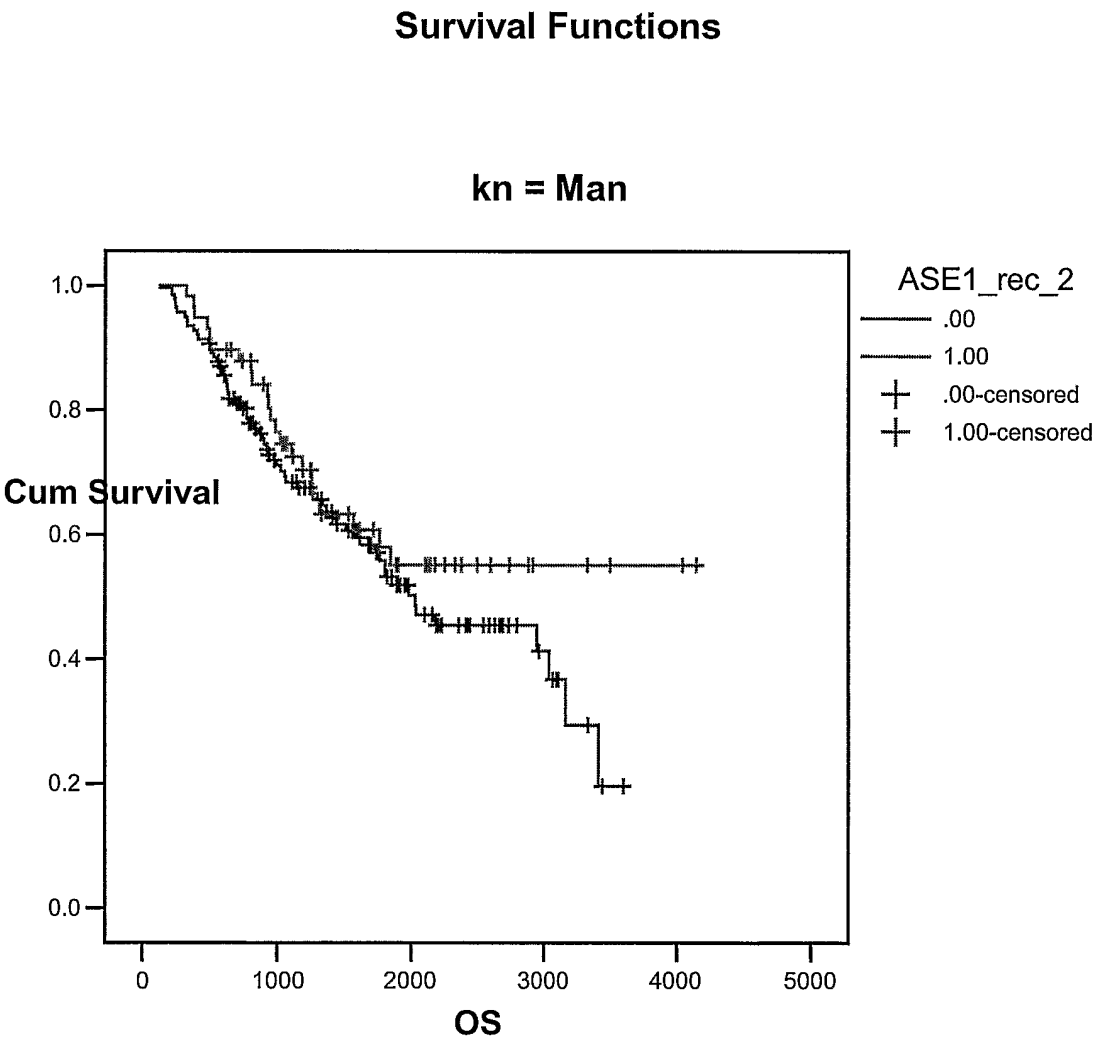
13/17

Figure 13



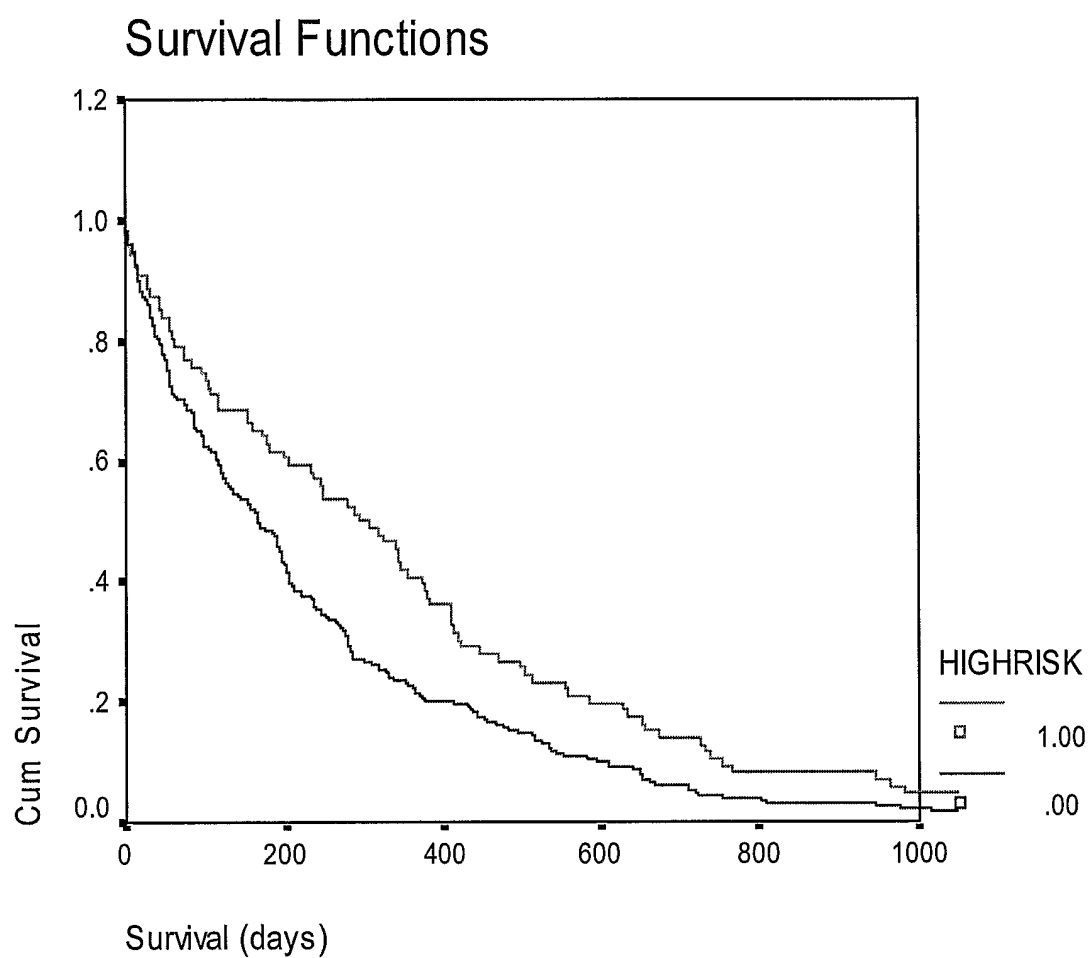
14/17

Figure 14



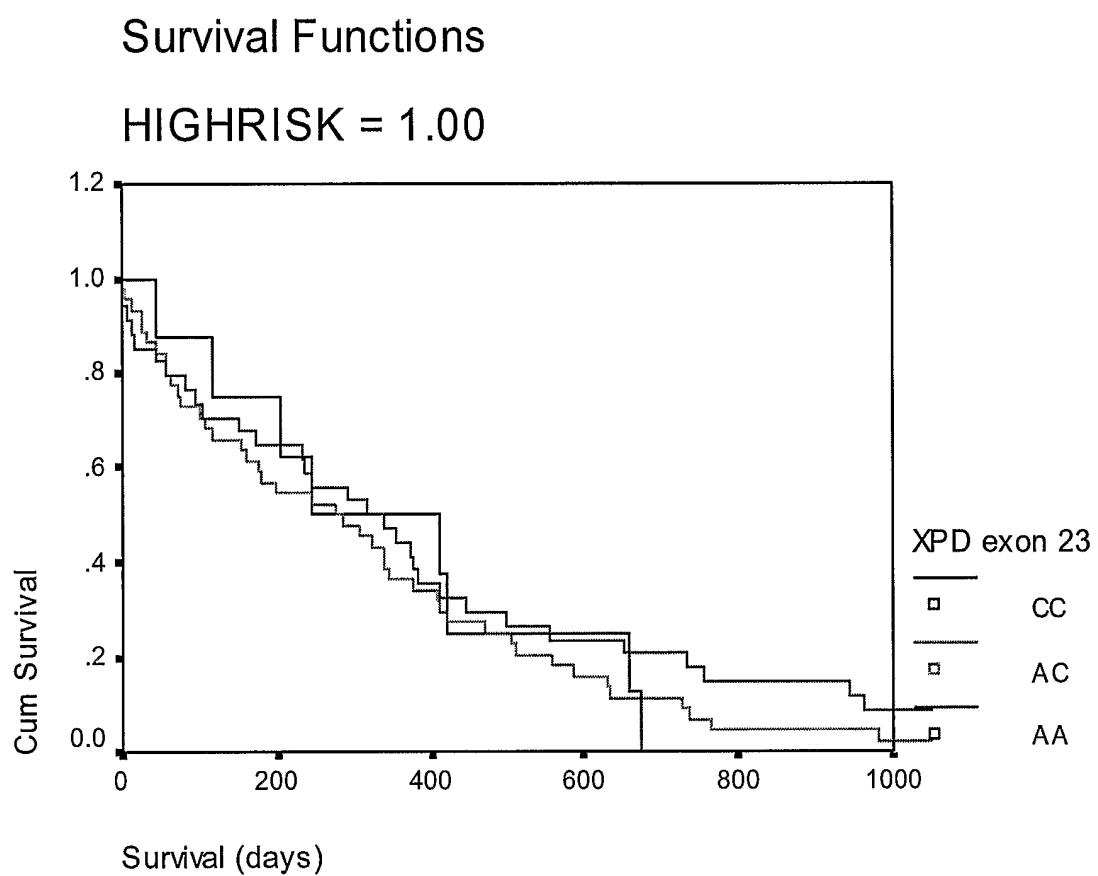
15/17

Figure 15



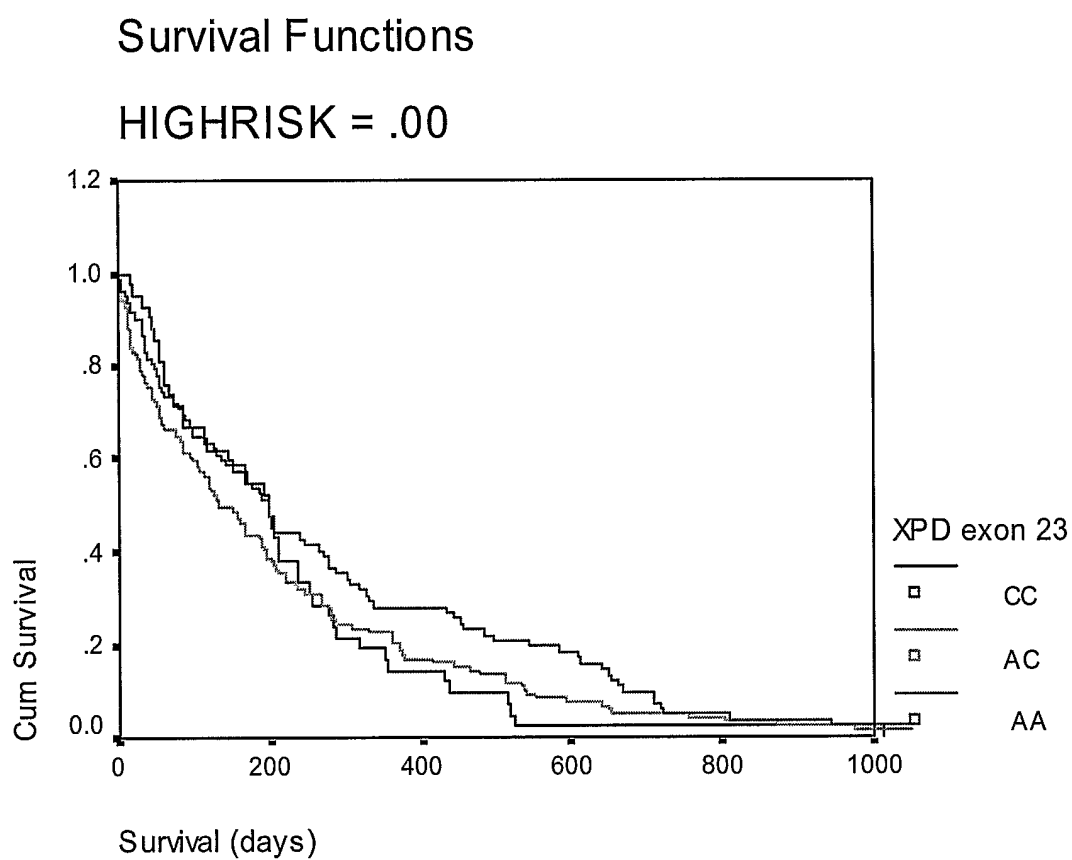
16/17

Figure 16



17/17

Figure 17



1
SEQUENCE LISTING

<110> Aarhus Universitet og Arbejds milj instituttet

<120> Disease risk estimating method using sequence polymorphisms in a specific region of chromosome 19

<130> P954PC01

<160> 142

<170> PatentIn version 3.3

<210> 1

<211> 37790

<212> DNA

<213> Homo sapiens

<400> 1

agaacccccg cccctccacc tcgtctcaaa aaaaaaaaaa aatcgtctca gtagcgaata	60
gtctaacgga gaatgacagg gaaattggtg atcctttctg ggccaagag ttagaaatgg	120
ctttgcaggc cgggcgcggt ggctcaagcc tgtaatccca gcactttggg aggctgaggc	180
agggtggatca cctgaggctg ggagttcaag accagcctga ccaacatgga gaaaacctgt	240
ctctactaaa gatacaaaat tagccggggc tgctggcaaa tgcttgtaat cccagctact	300
cgggaggctg aagcaggaga attgcttgaa cctgggaggc agaggttgca gtgagcagag	360
atggcgccgt cgcactctag cctgggcaac aaaagcgaac ctccatttca aatattaata	420
ataataacta ataaataaaa cataaatgct agcttttggt tgtttcttca acaaatagct	480
atgtggcatc taccatgtgt ctgatcctgt gctggcccct gggaacagaa aggtgaccat	540
gacagcctca gcacctgccc tcaagaaca gatttttttc cttgagacag ggtctttctc	600
tgtcgccaaag gctggagtgc agtggcacag tcacagctca ctgcagcctc cacctcttgg	660
gctcaagcga tcctccacc tcagcttcca gagtagctgg gaccacaggc gtgcaccacc	720
aagcccagct aagttttatt ttttaaat ttttagagac gaggtctcac cacgttgccc	780
aggctgggta aactcgcagg ttcaagtgat cctctcccct cagccttca aattgttggg	840
attacagggg tgaggcacca ggcctggcct caaagaacag atattaaata taaaaatgaa	900
tatatgatta cagcctggag tgggtggctg tgctgtgggt tccaacactt tggaaggcca	960
aggcgagtac attgcttgag ctgaggagct agagaccagc ctgggcaaca tggtgaaaac	1020
ccgtctctac aaaaaatgca aaaattagct gggcgtggtg gcgtgcacct gtagtcccag	1080
atactcagga ggctgagggt ggagaatcac ctgggcctgg gaggcagagg ttgcaatggg	1140
cagtgattgt gccactgcac tccagcctgg gcaacaggag tgaaaaccta tctcaaatgt	1200
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgc gcacgtgtat aatcacaagt acaaaagtgc	1260
tgtgaaggaa aacttcaagt caccataaag attgattatg ggctgggtgc agtggctcat	1320
gcctgtaatc ccagcacttt gggaggccaa ggcagatgga tcacgaggtc aggagttcaa	1380
gaccagcctg gtcaacatgg tgaaacccta tctctactaa aaaaaaaaaa aaaaaaaaaa	1440
aagccaggca tagtggcatg catctgtaat cccatctact cgggaggcta aagcaggaga	1500

attgcttgaa cccaggaggc agaagtgagc caagatcacg ccactgcact ccagcctgcg 1560
 tgacagagca agactccgtc ccagaaaaag aaaaaaaaaa aagacttatt atgacaggat 1620
 gtctactgtc aactgtgggg tgtgagtgtt ggccaagtga tcagagaagg cttcgtggaa 1680
 gaagcgaggt ttgagtagag ccagaaaata attagaagag atcaaccagc aagaggggat 1740
 ggatgagaga agtgagaaag gtgttccagg gagagagacc atcatacaca aaagctctag 1800
 gccagaagaa agctgaggcc tgtgagtgtc gaaaggaagc ctgtgggggt ggagctctga 1860
 gttgagcaca gggagcagag aaagggcagc tggaggggaa ggcaggggca gatcgaaatc 1920
 tcttttttaa attaatat tcttaattta tttatttttg agacaaggtc tctactcttc 1980
 gccagactg gagtacagt gcacaatctc agcgcaccgc aacctctgcc acccaggctc 2040
 aagcaattct ctggcctcag cctccctagt agctgggatt acagggtgcgc accactactg 2100
 cccagctaatt ttttatactt ttagtagaaa cggggtttca ctatgttggc caggctggcc 2160
 tcaaaactct gacctcaaaa gatccacca cttcagcctc ccaaagtgtc gggattacag 2220
 gtgtgagcca cccttcccgg ctgtattttt ggagacagag tcttgctctg tcccagcctg 2280
 gagtatggtg gtgtgaattt ggctcattgc caccttgacc tccagggctc aagtgatcct 2340
 cccacctcag cctcctgagt agctgggact gcgggtacac gacaccacgc ctggttaatt 2400
 ttttttaatt ttttgtagag acgaggggtat ctactatgt tgtccaggct ggttgaactc 2460
 ctgagctcaa gcaattctcc cacctcagcc tcccaaagtg gtgggattac agacgtgagc 2520
 cactgtgccc ggcttaattt atttacataa atttttttat gtttactttt ctatctccta 2580
 caggaagaaa atatattttg ttattgacag ggtctcgtc tgttgcccag gctggtattg 2640
 ggtcaagcc atcctgttcc ctacgcctcc caaagtactg ggattacaag cgtgagcctc 2700
 tgcattccagc ccagatccaa aatctttact gtcacctaca gagtcctctg taactagctt 2760
 actgtcatc atccccatac caaccacct tactgtctctg atctcctcct ctctctcccc 2820
 cagctcattt tgtttcagct atgctggtct ccttgctgtc tctaaaacat aacaagcaca 2880
 tcccatctca gggcctttgc accagctatt ttgtctgcct ggaatgctgt tccccctgat 2940
 agccatgtgg ctgacacact cacctccctc agctctttgc tcaattgtca acttctcggc 3000
 ccggcatggt ggctcacacc tgtaatccta ccactttggg aggctgaggt gggcagatca 3060
 cctgagatca ggagttcgag accagcctgg ccaagatggt gaaatcccgt ctctactaaa 3120
 aatacaaaaa ttggcaaagc atggtagcac ataccagtaa tcctagctac ccgggaggct 3180
 gaggcaggag aattgctgga acccgggagg cagaggctgc agtgagccaa gatcatgcc 3240
 ctgtactcca gcctgggtga caaagcaaga ctctgtctca aaaaaaaaaa agtctccttc 3300
 tcaatgaggg cttcctgacc accaaattaa atctacctcc tagacacaca cacacacgca 3360
 cgcacgcacg cacacacaca cacgcacgca cgcacacaca cacacacaca cactatat 3420
 cccctttccc tgctttattg ttcttgagag ctcatTTAAC catgtgacat gctgaatatt 3480
 ttacttattt attttgttta gaaagctcct ggctgggcgc gggggctcac gcctgtaatc 3540

ccagcactttt gggaggctgg aacaggtgga tcatgtgagg tcaggagtgc cagaccagcc 3600
 tgaccaacac ggtgaaacct catctctatt aaaaatgcaa aaattagctg ggtgtggtgt 3660
 cgcattgcctg taatcccaac tactcagaag gctgaagcag gagaatcgct tgaacctggg 3720
 aggcagaggt taacgctgag ccgagatcgc gccattgcac tccagcctgg gcaacaagag 3780
 tgaaactctg tctcgaaaaa aacaaaagtc agctccatgg caggagtgat ggctcacgcc 3840
 tataatccca gcactttgtg aggccgaggc gggcggatca cttgagggtca ggagttggag 3900
 accagcctgg ccaacatggt gaaacctcat ctctactaaa aatacaaaaa ttagccgggc 3960
 gtggtgacac atgtctgtag tcccagctac ttgggagggt gaggctggag aatggcttga 4020
 acctgggagg tagaggttgc agtaagccaa gatcgcgcca ttgctctcca tcctgggcaa 4080
 cagactccgt ctcagaaagg aagaaagaag gaaagagaga aagagagaaa gagacagaga 4140
 gagagagaga aaggggagaaa gagagaaagg atggaaggac cctgacaagc actgttgcac 4200
 aaaagtttct tttctctctc tttttttttt tttttttttt ttgagacagg gtctcacttc 4260
 tgttgctcca gctgaagtgc agtggtgaga acatgggtca gtgcagcctc aacttcccag 4320
 gcttaagtga tcctgccacc tcagcctcct gagtagctgg gactgtaggt gtgcaccacc 4380
 gtgcctagct aattttttgt attttttagta gagacatggt tccgccacgt tgcccaggct 4440
 ggtcttgaac tcctgggctt aagggatctg cccgccatgg cctcccaaag tgctgggatt 4500
 accagcgtga gccactgtac ccagcctgag tataggtttc tgataaattt taggatcata 4560
 ttgtttggac tgggtaagaa tttccagaac tctaataaag aaactgactg gtttatattt 4620
 tattttattt tattttatta tttttgagat ggattttcac tcttggtgcc caagctggat 4680
 tgcagtggca cgatcttggc tcaccacaac ctccgcctcc cggtttcaag tgattctcct 4740
 gcctcagcct ccccaggagc tgggattaca ggcaccacc accatgctcg gctatttttt 4800
 tttttatttt ttatttttta gtagagacgg ggtttcacca tgttggccag gctggtctcg 4860
 aactcctgac ctgaggtgat ccacctgcct tggcctccca aagcgtggg attacaggca 4920
 tgagccactg tgcaaggcct aggctggttt ataaaattgc taaaccaagc agaacatgaa 4980
 ttaaatacca aggaaatact ctcttagatt gtcattgtac atcagccaat actaaaattg 5040
 tcaagataca caatttgaat gaactccatg gtccaagtgc aattatctat gatattacc 5100
 atctaataaa cagcactatg tcccttaatg ggagaaaaag ttggagaatt taagagaata 5160
 tcaatccaat gttggttggg tgcagtgaat catgtctata ttcccagcac tttgggaggc 5220
 caaggcagga ggatcacttg agcccaggaa ttcaaggcca gcctcggcaa cacggtgaga 5280
 tcctgtctct acggaaaatt aaaaaaaaaa aaagagagag attagtggga tgtggtgcct 5340
 atagtcccag ctacttggga ggctgaggcg ggaggatcat ttaagcctgg gacgttgagg 5400
 ttgcagtga ccatgagtga gactcatctc aaaaaaaaaa aaaaaatggc gatcactaga 5460
 ggaaaaaaaa actaaagtgg ggtttgcggg tagtgggagg gcccttcctg ctaggttgca 5520
 ctatgatctc caggagggt ccacgggaga atcatttcct tgtctttttc agtttctaga 5580

gccaaattct ttgcatacct tgcattcctt ggctcggaac cccttccta accttcaaag 5640
ctggcagcta gcctctggct caagtgtcac atggcctgtc tctgtcttcc tatccaatct 5700
tcctcttata agaacattgg agccaggcat ggtggctgac gcctgtaatc ccagcacttt 5760
gggagaccga ggcaggcgga tcacaaggctc aggagtctga gaccagcctg gccaacacag 5820
tgaaaccccg tctctactaa aaaaatacaa aaaagtagcc gggcatgggtg gcagggtgcct 5880
gtaatcccag ctacttgaga ggctgaggca ggagaatcgc ttgaacctgg gaggcagagc 5940
ttgcagttag ccgagatagt gccaatgcag tccggcctgg gcgaaacagc gagactccgt 6000
cgcaaaaaaa aaaaaataat aataaataat aaataaaaaat aaaaataaaa taaaaaaata 6060
aaaataataa aataaataaa aattattttg agacaaagtc tattctgtgg cagaggctgg 6120
aatgcagtgg cgtgatcaca gcttactgca gcttctacct cctgagctca agcgatcctt 6180
ccaccttggc ttcctgagta gctgggacct cagggtgtaca ttaccacgct cagctaatta 6240
tttattttatt tattatattt ttgtgacgga gtttcgctct tgttgcccgg gctggagtgc 6300
aatggtgcta tctcagctca ctgcaacctc tgcctcctgg attccagtga ttctcctgtc 6360
tcagcttcct gagtagctgg gattacaggt acatgccatc acgcccagct aatttttgta 6420
tttttagtag agacgggggt tcatcatatt ggtcaggctg gtctcgaact cctgacctca 6480
ggtgatccac ctgccttggc ctcccaaagt gctgggatta caggcgtgag gcaccacgcc 6540
cggcaatttt ttttttcttt ttttttttc agacagagtc ttgctctgtc acccaggctg 6600
gagtgcagta gcgtgatctc ggtttactgc aacctccatc tcccgggttc aagcgattct 6660
cctttctcag cctcccaagt agctgggact acagggtcac accaccacgg cgggctaatt 6720
tttgtatttt tagtagacac cagggttcac catattggct agactggtct caaactcctg 6780
acctcagggtg atccatctgc ctacgcctcc caaattgctg ggattacaag cgtgagccac 6840
acacctggct taattttttt atttttgatc gacacaggggt ctccctatgt tgtccaagct 6900
ggcagagatt tttgtttgtt tgtttgagag ggaattttgc tctttagacc caggctggag 6960
tacaatgggtg caatcttggc tcaccacaac ttccgcctcc cgggtttaac agattctcct 7020
gcctcagcct cccaagtagc tggaactaca ggcacctacc accacaccag gctaattttt 7080
gtgcttttta gtagagatga ggtttcacca tgttggccag gctggtctta aactcctggc 7140
ctccagtgat ccaccgcct tgacctcca aagtgtgaa attacaggcg tgagcaccgc 7200
gcctggcctc tcaacctaca atttcaacac ccaaggaaac agcccaccat gagtgagaac 7260
cagcagacac aacaaactat aggattagct gcctccaaac ttcagggtgat agattatcag 7320
gcatgtactt gaaactaaag gacacaaaag aagaatccga aatataaaat aaaggattgg 7380
acttgtgtga aaagaatccc ttagaaaggg ctactttcag gctggccatg gtggctaattg 7440
gcctgtaatc ccagcacttt ggaaggccga ggtgtgtgga tcacctgagg tcaagagttc 7500
aagaccagcc tggccaacat ggtgaaacct cgtctctact gaaaatacaa aaattagcca 7560
ggtgggggtg cagatgcctg taatcccagc tactcgggag gctgaggcag gagaatcgtc 7620

tgaactcagg aggcagaggt tgcagtgagc tgagattgcg ctatcgtgcc ccagcctggg	7680
cactagagtg agatcaaaaa aaaaaaaaaa aaaagaagaa gaagaagaaa gggctacttt	7740
cagactgcct tgccaaaaat cataaccaca atgatgagca tgtattgagt caaacagaa	7800
tcaaaagaga agaaagtcaa tttctgtgca aactactttt atttataagg aaagtttctc	7860
tattttgttt ataaacatta aaccagtgtc gtgtgaaggc acttaattgg ggagagggtg	7920
ggcagggatc ctggtagaga ccaatgtttc ccaccagac cccaagactg ctgggagaga	7980
tggtgtcagc agtgactccc aggaatatcc agtggtgtgg tggcccatcc caggcccggc	8040
tgggcaggtg gctggcttgc tgggggatgt gatgatggtg gtaggcatgg gaggcacttt	8100
ggacgggatc tgatttggca aaaggaagtg gtttctgtc cccagtgatt tccagccctt	8160
cccagacctc ccaaggctaa ggagattac taaatttaag gctggggccc tccttcttcc	8220
ctggacttcc aggagaacag agaaccggtg gcaaggacca ccaccagcag ggtgaggggt	8280
gcagataaag gcagcaaaaa acagaggag aggtctggag ggaaggcagg aatgcttgtt	8340
tctgtcagcc tcagaaacct ctttctatcc tgctagactt tactccttg aggcttcacc	8400
ctggggaaca gctggggaga gacaggatct tcagacatca ggagctcca cctcctcatc	8460
ccacatgcaa atccgctgcc tgtctctatc ctcccacccc ttcctaaggg gacctctcag	8520
cacctcccaa actgtctcag aatccaagtt ctgtgtcacc tccaagaacc agatggaacc	8580
ttccaatcag agcctccact gatgaaatgg aatatttcca gtgtctccta actgccataa	8640
ggagaagccc acctctctct aacaccttgg ttgtcttttt ggggtcccacc tccatattta	8700
aaaaatctcc tctctcaggg ccgggagcag tgggtcacac ctataatccc agcagtttgg	8760
gaggccgagg tgggtggatg acctgagctc aggagttcaa gacaagcctg gtcaacatga	8820
cgagaccctg tctctactaa aaacacaaaa aattagctgg gcgtgggtgg gcatgcccg	8880
aatcccagct acttgggagg ctgaggcagg agaatcactt gaatccggga ggtggaggct	8940
gcagtgagcc aagatcgcg cactgcactc cagcctgggc gacgcagctg aagctgtgtc	9000
tccaaaaaca aaacacacac acacacacac acagaaaaaa aaaacaaaaa taaaaaatc	9060
tcccttctca ggaatgtaac ggaatcttcc ttgccttctc ccctaaccct aatagagaat	9120
tttctcagt tacactgtaa ttttattaat ggatttttcc tcattctgcc caatgcagtg	9180
taatgaaagc ttcctctcca tctgttatat tatatataaa tatatattat atatttatat	9240
attatatatt tatatataac atataatttt attgtcaccc aggctggagt gcagtggcac	9300
catcagggct cactgcagga tcaatctccc aggcctaagc gattctcctg tgtcagcctc	9360
ctgatgagct gggattacag gcacccgcca ccacacccgg ctaacttttt tttttgtat	9420
ttttagtaga gatggagttt caccatgttg gccaggctgg tctagaactc ctgacctcag	9480
gagatccgcc cgccttggcc tcccaaagtg ctgggattac aggtgtgagc cacctggccg	9540
ggccctccac ttccttcttg tacattgctg aatccctgtg tcagccctag aggtccagtc	9600
ttttgcctc tcccagcctt aatctacaat tctgtaaccc acccaccatc attaaaaatga	9660

gattcttctt tgctgcttcc cttggctaaa atggattatt ctttaacctc tccaccaata 9720
caaccagggg tgataataaa aacattggat tgagcagaaa ccaatcaa atactagtaag 9780
gcagtactgg cgagcacctt acatcctgac agctttataa agggcgcttc cagccagggtg 9840
cggtggcaca tgcctgtaat cccaggactt tgggaggctg aggcgggcag gtcacctgag 9900
gtcaggagtt caagaccagc ctggccaacg tgatgaaacc ctgtctacac aaaatacaaa 9960
aaaaaaaaaa aaattagccg tgcgtggtgg catgcgcttg tcatcccagc tactctggag 10020
gccaaggagg gaggatcact tgagcccggg aggcagaggt tgcagtgagc ccacatctta 10080
tactgacact ccagtctggg tgacaaagca agactccatc tcaaataaat aaatacaaat 10140
tggccgggtg cggtggctca tgcctgtaat cccagcactt tgggagacca aggcagggtg 10200
atcatttgag gtcagtagat caaaaccagc ctggccaaca tgggtgaaacc ccgtctctac 10260
taaaaaataca aaaagtagcc gggcggtggtg gtggtgggcg cctgtaatcc caggcaggag 10320
aactggttga gcccgggtgg ggggggcccc aggttgacgt gagcacagat ggcgccattg 10380
cactccagcc tgggcgacag agcgagactc cgtttcagaa ataaataaat aaaataaaaa 10440
taaaaaataaa aaaataatag aaatttaaaa ataaaataaa gggcttttcc tcacctactc 10500
cactaactat aaggggaccct tacccccagc attactatta aatataacgg acttttcgtc 10560
tcctcccat gagcaataat gagcttttca gacctccctc tcccaatata acggtttggt 10620
cctgttgctt cttctttttc ctgtgggatc ccccttttcc ccaaccccca actgtcggga 10680
ggtcccatg acttctcccc tgggctcacc ccgaagtagt tccgcggcac gtagccctcc 10740
tggccgtgca gcgcggccca ccaccagtcg gtctcctccg gcccgctcct ccgcagcacg 10800
gtgaccgact cgccctcgcg gaaggacagc tcgtccccga actcggcgct gtagtcccag 10860
agagcgtaca ctgccccgct gttcatcagc ccatactct gctcgacgtc tgaaacatgc 10920
cacggagggg aaggtgagag cctggcccag ggggtccagg aacaggggccc acgtgggggtc 10980
caggacagac cctggaattt ggcgcctgtc ccagcaacca cctgaaatgt tgtgtgtgcc 11040
catggctgtg gatgggaacc ggagctggag tcagatgccg ggactggccg tctttgagcg 11100
ttcgaggaaa ctgggggagg catgccagtg ggccaccac tcccgaggca gggtcagagg 11160
ctcccatttc ttttctttct tttttttttt tttttgagac agagtctcgc tctgtcgcgc 11220
aggctggagt gcagtggcac gatctcggct cactgcaacc tccgcctccc gggttcacac 11280
cattctcctg cctcagcctc ccgagtagct gggactacag gcgcccgcca ccacgcctgg 11340
ctaatttttg gtatttttag tagagtcagg gtttcaccgt gttagccagg atggtctcga 11400
tctcctgacc ttgtgatccg cccacattgg cctcccaaag tgctgggatt acaggcgtga 11460
gccaccgctc ccggcctttt tttttttttt tttttttttg agatggaatt tcgctcttgt 11520
cgcccaggca ggagtgcaat ggtgcggtct cactgcaacc tccgcctccg gagttcgagc 11580
cattctcctg cctcagcctt ccaagtagct gggattacag gtgtgcgcca ccatgcctgg 11640
ccaatttttg tatcttttagt agagacgggg ttccaccatg ttggtcaggc tgggtatcaaa 11700

ctcttgacct caagtgatcc acccgctctg gcctcccaaa gtgctgggat tacaggcgtg 11760
 agccacctgg cccggccctc atttccttct tgtacattgc tgaatgcccg tgtcaaccct 11820
 agagggtccag tcttttgccc taccctggcg cttagcttaa gtggtacagt ctctaaggaa 11880
 gattcgacc ttccttgaat gataggggtcc ttttaagttgg ctcatctgcc tctttctttt 11940
 cttttctttt cttttctttt tggagacgga gtcttgctct gtcgcccagg ctggagtga 12000
 gtggcgcat ttcggctcac tgcaacctcc gcctcctggg ttccagcaat tctcctgcct 12060
 cagcctccaa agtagctggg actacaggcc cacgccgcta caccggcta aattgtttta 12120
 tatttttaat agagacgggg tttcaccgtg ttgccaggc tgggttgga atcctgagct 12180
 catgcaatcc gcccgctctg agcctccaa agtgctagga ttacaggcat gagccaccgc 12240
 gcctggcttt cttttcttt tctttcttt tttttttca gacaaggct cactctgcca 12300
 cccaggctgc gggagtgcag tggtgagatc aagcttactg cagcctcgaa cttccagatt 12360
 caagcaatcc tcctgcctca gcctcctcct gattctttat gttattatta aatattttgt 12420
 aggccgggca cagtggctca cacctataat cacagcactt tgggaggcca aggcaggcgg 12480
 atcctctgag gtcaggggtt tgagaccagc ctggccaaca tggcaaaacc ccgtctctac 12540
 taaaaataca aaaaaaaaaa aaaaaaagt tagcgggccg tggggccctt gcctgtaatc 12600
 ccagttactc gggagcctga ggcaggagaa tcgctttcac cgaggaggca gaggtttag 12660
 tgggctatgg tgccattgca ctccagcctg ggtgacagag caagactctg tctcaaaaaa 12720
 taaataaata aaaataaata aatatttcgt agaggctagg tgtggtggct cacacctgaa 12780
 tcttagcact ttgggaggcc aagggtggga gattgcctga gctcaagagt tcgggaccag 12840
 cctgggcaac actgcaaaac cccttctgta ctaaaaatac aaaaaaatga gtcgggcatg 12900
 gtggtgagca cctgtagtcc cagctactca agaggctgag gcagagaatt gcttgaatcc 12960
 aggagggtga ggttgacgtg agccgagatt gagccactgc actccagcct gggtgacagt 13020
 gagactctgt ctcaaaaata ataataaata aatatttgta gagacagggg gtctctacaa 13080
 tgtctttag cctgaccagg ctacactttc aaatatataa ccctctgtct caccataag 13140
 tcctaggacc tgcctcactc caactctccg tgaagttcct tgcccacacc gagatacaac 13200
 tggctcctcc aggtgtgaaa tgaccctgtg cacaatcccc gtggcacagc ctacttcgcc 13260
 ctgcccgtcg gggaaccagg tgatgtagcc tgccccctgg agagataggg tacagccttg 13320
 tgtcttccta caagcccctt tctggcagct gtagcctgct cacctgccag tgggtgtggca 13380
 atgcctctcc cacaagtggc agagcccacc tgcccagagc cctatgccag gtagatggca 13440
 ggggtgaaac gttcagctcc tcacccttga agatgtgaaa ggtgagcaga ccaatcttca 13500
 cagccactct cctccccaaa ggtgtccagc tcgcatagca cagcctccat gtcccccttt 13560
 cccttaggag ggcatagtcc cccaccccc gcaagcggtc catccctcat cctcctctc 13620
 ggcaatcctg ccaagtgggt ggtacagccc ccataaccct ctctccctag tagggggtag 13680
 ttgctcccc ccccgctcct gcgcaccgc caggtagcca ggcgccagca gccctgcctc 13740

gcacctgcca ggtaggtggc gcagtcagca taaccctcgc ggtaaggggtc gcacttctcg 13800
aaggcgggtgg cgccgtcgct gagcgtgggtg gcgaagattg cagcgccgtg ctgcaccagc 13860
gccatgcaga tgactgtgtc gttgcacgac gccgcgcagt gcaaggggtgt cctaggcgtg 13920
gggggtggggg gttgcgggga acgatgcgtg agaggctgctg cgtccgcca cgggggaccc 13980
agcccaccgc gcgggtcggg gctcaccagc cgtggctgtc gggggagttg acattggcac 14040
ccgcggtgat gaggaatcc acgatagagt agttggcgcc gcagatggcg ttgtgcaagg 14100
cagtgatgcc ctctcgttg ggctggctcg ggtcgttcat ctgagtgcac cgggggaggg 14160
ggaagactca gtcccgcggc tggcatctgc gatgcccccg ccgtgccac ctcccgtca 14220
gcagcgtca ctctcttcac cgcctgctgc accacctcca gctccccggt cagcgccgcg 14280
tccaggagga gcaccagagg gttgaggcgc gcgcggcggg ccttgccgcg ggagcccgc 14340
ttccgcagca cagagcgcat ctctggggg acaggcgca gaggtcagcg acttgagggg 14400
attgttagta tatccatgat ctagagtagg aaacagagggt ccagggactt gtggcaccca 14460
tctagacagg ggtagaactg ggattccctc gggatgggggt gagggggtgc cttcgatctc 14520
ctcctagagc ctccagttcc ctgcataga cagggaatcc tgtgatttga gaatcttggg 14580
ccctgaaact tgggagaaaag ctggggggcc atgggattgg tggcaaagta attctatcag 14640
ttcaaaacaa tgattgtgga agccagttat gcaattcaca cacagtctca catttctttt 14700
gttaataatg aatgcaatga gacacacatg acaaaatgtt accaggagtg ttcatcccg 14760
atgtttggaa tttgagcatt ttattattcc ttgtatttcc cttttctttt tctctttttt 14820
tttttttttt tgagatggag tctcgctctg tcaccaggc tggagtgcag tgcagtgggtg 14880
tgatctcagc tactgcacc ctccatcccc caggttcaag caattctcct gcctcagcct 14940
cctgagtagc taggattaca ggcattgcgc actatgcctg gctaattttc atatttttag 15000
tagagacagg gttttgtcat gttgtccagg ctggtctcga actcctgacc tcagggtgatc 15060
caccacctc agcctcccaa agtgctagga ttacagggtg gagccactgt gccagcctc 15120
atgggctttc ttatttttaa ttttctctct gtaagattca tttattctgg gctgggcgag 15180
gtggctcatg tctgtaatcc tagcactttg ggaggctgag gtgggaggat cacttgagcc 15240
caggagttcg agaacagctt gggcaatata gtgagacca gtctctacaa aaaataaaaa 15300
attagcctga catggtggcg cacaccgctc gtcccagcta cttgggaggc tgaggcagga 15360
ggattacttg aatggaagag aaggaggctt cagtgaacca tgatcatgcc actgcactct 15420
agcctgggca acagagtga acccagctc aaaagaaaaa aaaatgcatt tatttattcc 15480
aagtgtgtga gtgcatagca tttgtgattc tgggtcttgc tgtttccaga gtttcagtga 15540
ttttaagatt ctggaattca gagatcccaa cagccactga attcaaaatt cccagatgct 15600
cagttatttc aagtttccaa tatgttgtga ttgcagaaat gctaggctgt gctatttcaa 15660
attgctgagg ggccaggact ttggaatcca aagattctat gatggagaac tttaatattt 15720
ttctgttaga atttcttttt tttgttggtt tttttgagac agagtctcgc tctgtcggcc 15780

aggctggagt gcagtgggtgc gatctcagct cactgcaagc tccgcctccc gggttcaggg 15840
 cattctcctg cctcagcctg ccaagtagct gggactacgg gcgcccgccca ccacgcctgg 15900
 ctatttttga ttttttagtaa agatgggggtt tcaccgtgtt agccaggaag gtcttgttct 15960
 cctgacctcg tgatccgccc acctcggcct cccaaagtgc tgggattaca ggtgtgagcc 16020
 atcatgcctg acctagaatt tcatttttaa agactagaag gaaatggctg ggtgcggtgg 16080
 ctcatgtgtg taatctcagc actttgggag gctgaggaga gtggatcacc tgaggtcagg 16140
 caggagttca agaccagcct ggccaacgtg gtgaaaccct gtcttacta aaaatacaaa 16200
 aattaggtgg ccgtgggtgg gcacgcctgt aatcccagct actcaggagg ccgtggcatg 16260
 agaatcactt gaaccaggga ggcacagtta tagtgagctg agatggcacc atcgactcc 16320
 agcctgggtg acagagttag actccatctc aaaaaaggaa aaaaaaaga aagactagaa 16380
 ggaaatattc aaaatgttaa tgatgggtcc ctgtgagtggt tgtgattttg tcctctttct 16440
 tctattttta tttattttcc ccaagctctc tatgggtgtg gtgtatttct ctatagtgga 16500
 atgtgtaaat ttaaagtata aatctcagct gggcacagtg gctcatgcct ggtttgagac 16560
 cagcctggac aacataatga gaactgtctc tactgaaaat gttaaattt atctgggagt 16620
 ggtggtgcat gcctgtagtc ccagccatag gggaggctga ggcatgagga tcaattgagc 16680
 ccagtaggtg gaggctgcag tgagccatga tcttgccact gcactccagc ctgggcaaca 16740
 gagttagact ctgtctcgat aataataacc ctctattaca acatatcagt gcatgaattt 16800
 gtgattttat aattcaaaat atgagcatct ttaattgtca gatttggtga cttcaagaat 16860
 cagtaataat cagtctatga tactaacttt ataattattt tttttaagag aagagtttcc 16920
 ttttatttta ttttatttga gacagagttt ctctctgttg cccaggctgg agtgagctgg 16980
 cgcaatctcg gctcactgca gcctctgtct cctaggttca agcaattctc ctgcctgagc 17040
 ctcccagta gctgggatta caggcatgca ccaccaggcc cagctaattt ttgtattttt 17100
 agcagagacg gggtttcacc atgttggcga ggctagtctt gaactcctga cctcaagtga 17160
 tccaccgccg tcggcctccc aagggtgctgg gattacaggc atgagccacc gtgcccagcc 17220
 taactttata attctaagat cgtgttcaaa cttttaaatg ctctagggct ctaaaatgtt 17280
 actatcctaa gacggtgaca cttagcgttg attcttacat tctatgattt ttttaagttt 17340
 tctgtggcca ggactctgtg attctacaat gggatgctca gccatttcaa catgttggtt 17400
 ttcatccctt cttgatttca aaatcctgag cctcaagggt ccttgccctt actttcagga 17460
 gggcctagga ataggcattt tgggggggtc cacctgacct ctgcttctct gagaagtgat 17520
 ctcttcccg cgtctacgca cacggagtgt tcaggactgt tccatgtggc tacaaccctc 17580
 ttcccagtc agatgcaggg accaagatca gcaggagacc atcccctggg ccaatgggtga 17640
 caacagtaag agcagttaac agttatgtgc caggatttat gctaagcact acattaatgt 17700
 attaatctt ggcgggggtg ggtggctcac acctgtaatc ccagcacttt gggaggccag 17760
 ggcgggcaga tcacttgagg tcaggagttc aagaccagcc tagccaacac agtgaaaccc 17820

catctctact aaaaatacaa aaattagcca agcgtggtgg catatgcctg taatcccagc 17880
cacttgggag actgacgcag gagaatcact ttaaccacagg aggtggagtc cagcacccag 17940
ccgagactca cttgttttta tttattttatt tatttttttt tttttttatt ttttttgaga 18000
cggaatcttg ctctgtcacc caggctggag tgcagtggcg cgatctcagc tcaccacaag 18060
ctccgcctcc cgggctcacg ccattctcct ctacgcctcc agagtagctg ggactacagg 18120
cgcccgccac cacccccagc taatttttgt attttttagta gagacggggt ttcaccgtgt 18180
tagccaggat ggtcttatct cctgacttcg tgatccgccc gcctcggcct cccaaaatgc 18240
tgggattaca ggcataaacc accacgcccg gcctatttat ttattttatt agagatggag 18300
tcttgctctg tcgcccaggc tggagtgcag tgggtgcagtc ttggctcact gcaacctccg 18360
ccttcgggt ttaagcgatt ctcttcctc agcctcctga gtagctggga ttggaatgag 18420
accaccactt ctctgttgt ccttcccagc ttctcccca cctcccctt tccctagttt 18480
ataagacagg aaaaaaagg agaaagcaa acgctggaaa aaaacagaag tacgataaat 18540
agctagatga ccttggcgcc accatctggt cctggtggtt aaaataataa taataatatt 18600
aatccctgac caaaactact ggtgttatct gtaaattcca gacattgtat gagaaagcac 18660
tgtaaaacgt tttgttctgt tagctgatgt ctgtagcccc cagtcacgtt cctcacgctt 18720
acttgatcta tcgtggccct ttcacgtgga ccccttagcg ttgtaagccc ttaaaagtgc 18780
taggaatttc tttttcggg agctcggctc ttaagacgct gatgctccc gccgaataaa 18840
aacctcttcc ttctttaatc cgggtgtctga ggagttttgt ctgtggctcg tcctgctaca 18900
gaattacagg cacgcgccac cgctccgggc taatttttgt attttttttag tagacagggg 18960
gtttcaccat gttggtcagg ctggacttga acctctgacc tcatgatcca cccacctcgg 19020
cctcccaaag tgctgggatt acaggcgtga gccaccgcgc ccggccgaga ctactattt 19080
tataagagga gagagcaaag ccaggaacag tggtcatgc ctctaactgc agcaatttgg 19140
gaggctgagg cagggtggatc atttgaagtc aggagtgtga gaccagcctg gccagcatgg 19200
tgaaacctca tctctactaa aaatacaaaa attagccagg agtgggtggca tacacttata 19260
atcccagcta cttgggaagc taaagcggga ggatggcttg aacctgggag gcggagggtt 19320
cagtgaagcg aggtcaagcc actgcactcc agcctgagtg atggagcaag actctgcctg 19380
gaaaaaaaaa aaaaatagag gagagagcag agcagacaca agagacacag agacagagag 19440
ggagagaaga gaggggtgact gctttgattc aggcaagact tctcagtccc agaataaacc 19500
cactgttgtg ccaagactca gtcatgtcca ggtgtatgac tcgagattgc tgaaggaatg 19560
cccggggagc ggcacaggca caggttattg gagagaagga gcagagaaca tctctatgtg 19620
gccaagactc ccagatggcc ctccatatag tcacacacag ctatcctaaa gactacattt 19680
cccagcatcc cattgcaatg aggctcctgg ccagtgggag caggcagagt gatgtatgga 19740
actcccaggt tctgcctgaa acaggaaagg gcactttctc ttcttctttc tctcttcctg 19800
gctggagggc agacttggtg acagccatct aggaccatga aggcaggctt actccccgat 19860

ggatggcaga gccccaggta gatagagcct gggtcctgac tccagtgagg tgcctacagt 19920
 cctgggctgc aaactcttgg acttctactc aaaagaggag aaaacttcga tctcatctaa 19980
 gccactatat ttgggggggct ctttgctaca gtccttgat tcatgtagca aacatacccc 20040
 ggtttccctcc tgtattactt accatgctct gcggctgctc tgggtgggctg ctctgggacg 20100
 gggccggggg tggaatggga gctgggtggg caggagcagg gggccctgcc ctggcctcag 20160
 atccctcagt gatgggggac agctctggct ccggccccc gggccctggc ccccatgac 20220
 gatggaagag gcggctgatg atctgctggt actgtttctt gtgggtaggg ggcaggggcca 20280
 cagcaggggc ctgctccatg gagcccctgc gtttgagggg ccggggaatt tccgccaaca 20340
 cccgtgccac ctctccagc tcgggcaccg actgtgcctc cgggtggcagt gctggctgca 20400
 gcctcgtggg gctgagaggc ctgctacag ggccttcac cacatcgcca gcctccagca 20460
 ctgggtgtcag cagcccctct atctccggct caggctccag ctcggtgggg gggttggggg 20520
 gtcctagccg gaacaagagc ccatcagagg acaggtcccc aggagacacc caaactccc 20580
 tctccacaac ttccagggca tacaaccagc acatgatttt ctgtgtgacc tcaggggaagt 20640
 tccttgccct ctctgggcta cactttcctt gggctgtgaa taatatacaa ttatgatgcc 20700
 tcccatttat tgagcagtta gtatgtgcct ggcgctttac atgcctacct tattgtaatc 20760
 tcaccactgc tttgtgaggt agatacactg ccattctccac attaccgaaa gggaatctgg 20820
 gcctcagaga ggacaagtca gttgcccaca gccatgcagt tgggacttga actcagttct 20880
 ggctgactct agaactctact tctaccaacc gtgatagatg tgattttctg agatcctgag 20940
 agtttctct cctaactct caggcagaaa actccagcag gaagtagaat cctgggtgtt 21000
 aatgattttct tctctgtctt actcattctg acagtaaagc aggtggaaat aaaaatatgc 21060
 attattggct gagtcgagt gctcacacct gtaatcccag aactttggga ggccgaggca 21120
 ggagatctc ttgagatcag gagtttgaga ccagcctggc caacatggta aaaccctgtc 21180
 tctactaaaa atacaaaaa aaaaaaaaaa aaaaaaaat tagctgggcg tgggtggcaca 21240
 tgctgtaat cccagctact cggaaggctg aggcacagga atcgcttgaa cccaggaggc 21300
 ggaggttgca gtgagccgag attgcaccac tgcaccactg cactccagcc tgggcaaaag 21360
 agtgagattt catctcaaaa tatatatata tacacacaca cacacaaaca cacacacaca 21420
 ttatatatat agtgtatata tatttttata tagtatgcat atacatataa ataatacaca 21480
 cacacacaca cggctgagca tgggtggctca tgctgtaat cccagcactt tgggaggctg 21540
 aggtgggtgg atcacctgag gtcaggggtt cgagaccagc ctggccaaca tggcaaaacc 21600
 tcatctctac taaaaacaca aaaaattagt tgggtgtggt ggtgcatgcc tgtaaccca 21660
 gctacttggg aagctgaggt aggagaatcg cttgaacctg ggaggtgtag gatgcagtga 21720
 gctgaaacct caccactgca ttccagcctg ggcaagaaga gtgaaactcc atcttggctg 21780
 ggcacgggtg ttacgcctg taatcccagc actttgggag gccgaggtgg gcagatcatg 21840
 aggtcaggag atcgagacca tcctggctaa catgatgaaa cccgtctct actaaaaata 21900

caaaaattag ctgggggtgg tgggtgggcgc ctgtagtccc agccactcgg gaggctgagg 21960
 caggagaatg gcgtgaaccc gggaggcgga gcttgcaagt agcaagcacc actgcactcc 22020
 aacctggaag aaagagcgag actctgtctc aaaaaaaaaag agtgaaactc tgtctcaaaa 22080
 ataaataaat aaataaaccc caaaacacac acacatacac attatttcat tgaatccccg 22140
 tcacaattct atagggtaga tattattaat ctctcttcac agacgggaaa cagagtttcg 22200
 gacaagtaat ttatcttcag tcacacagca agttagcagt gaagagagac tccagcccat 22260
 ctgcttaact cactgatctc acacctcaa atattaataa attattataa ctaatatggt 22320
 agctatttat ttgagactgg gtctcactct gtcaccagg ctggagtgc gtggcgctat 22380
 cacagctcac tgcagcctgg atctcccagg cttaaagat cctcccacct cagcatcctg 22440
 agtagctggg actacaggcg cccactacca tgcccggcag attttttgta cttttatttt 22500
 tagtaaagtc tatttttagt tcactatgtt gccaggctg gtcttgaact ccagagctca 22560
 agcaatcctg tctgcattag cccaccaaac tgctaggatt acaagggtga gccacggtgc 22620
 ctggctaata tggtagctat tgatagctta ctatgtatca gatcctattt atttatttat 22680
 ttttgagaca gagtctcacc ctgtcacctg tgctggagtg cagtggcatg atcttggctc 22740
 actgccacct ccgcctcctt ggctcaagct gagtagctag gactacagt gtgagccacc 22800
 atgccagct aatTTTTTTTT tttttttttt tttttgatag agatgggatt tcatcatgtt 22860
 gtccaggctg gtcttgaact cctgacctca agtgatctgc ccacctcggc ctcccaaagt 22920
 gctgggatta cagggtgtgag caactgcacc tggcccatca ggtgctgttt taaaggcttt 22980
 atatgaattt aataacatat gtcaatagga tcgattctat cattatttgc cttttttttt 23040
 tttttttttt ttgaggcaga gtctccccgt caccaggat ggactgcagt ggcgcaatct 23100
 cggctcactg caacctccac ctcccgggtc caagtgattc tcctgcctca gcctcccaag 23160
 tagctgggac tacaggcgcc cgccaccatg cctggctaatt ttttgtattt ttagtagaga 23220
 tggggtttca tattggccag gctggtctcg aacttctgac tttgtgatcc gccgcctcg 23280
 gcctcccaaa gtgtgaggat tacaggcatg agccaccgtg cccggcccat tatttcctt 23340
 ttacactcaa gaaaattgag gccagtgag gttaagtgc ttgccaagg tcacacagcg 23400
 tggaaccagg cagtctggct tcagggtcca cacttaacct ttgagctatc cctggctcct 23460
 acccaaattc ccaaactcac ctggcctagc tctctgcagg gacagtgcct gtaaagaggc 23520
 atttggctgt gatctcccca cctcccagg ctggtctggt cccctgccca tttgtcctcc 23580
 cttcaccag tcctctagg ccctcattgc tgactcacct tcgttcacag gggccatgtc 23640
 tgttggggat gctggggggc tggggtaggg gtttgggggt gggctctggg ctgtgggggc 23700
 agctggggct gtggttgtga ttgtggctgg ggctgtggt gtggttggg ctgcagctta 23760
 ggcgggggtg ctcggtgaa gaggggggac ccaggagca tggcgggct ggccccgtgc 23820
 tcccagaagg cgttctgcag cttgaagatc atgctgaggg ggatgggacg ctggcgggg 23880
 gccccgcggg gctgggggct ggaggggggc atggggatgc ggctgacggg ctgccagctg 23940

cgaggcaaaag tgccccgacgg ccccgcgagg cccagcgagc gccggtagct gcccgcgctct 24000
 gaacgccggt cgctggccag aggagagacc ttgtaattgc gcggcaggggt ggcgctagtg 24060
 aggttgctct ggggaagagg gaagggagaa ggggatcggg tgagagaggg aaggtggagg 24120
 ggaggtaaag acaaaagacg agaagggaga ggaggtgagg gaagccctgg gagtgagggg 24180
 gaagaaaggg tgaggaagga gcagaaaccc agcacagtga agggagagcg tgggaacggg 24240
 cgccgagacc cagatcgag ccccgagggg gagactggcc ttgaccccg cccccaccc 24300
 cactcctcga ctttccccag cctctcctcc ccaggcgctc cctcctcacc ttgccggtgc 24360
 cccccagtcc atccaggctg ctctccctcc aaggcaacag ctgcaggctc ggcgagggcag 24420
 gccttgcgaa gacgtccagg cctgcggggc gggaatcatt agggctctgt gggctgcctc 24480
 tcctccgggt cctccattcc ccgggcctcc accactcacg ttcatagtct gctgtctgcg 24540
 aaggcttctt ctctacgcc acgtccagg cagactcgtt ccaggctttc ggaggccgcc 24600
 ggcgcagcgt caggtcgtct ggggagaagt ttccaggag gatgagacgg gaggggtggc 24660
 gagccccga tcctgcccgc tttgaccccg cgagtcгааг gcccgcgag gggcccctgg 24720
 gttcaccttg cgcgcgaga ggcgggcgga atgcgctgcc gccggagcct agcagggagc 24780
 tcccgaaggc ggacgtggc gcgtcgtagg ctgtggcagg gggcgcggt gacggcccac 24840
 gctcggggaa gaaggcctgg ggcccctcc ccagggggct gccgcggggg gagcctgcgc 24900
 ggcccaggaa gtcgaaaggc gtggggggac cctgctggcg gagcgggcct ggcccgggcc 24960
 gcggggaggg cgacggccg agggagctgc ctgcgccatc gaaggcgcg ggccggggcg 25020
 aggtcgcgcg gtccaggctg ccgtaggcgt ccggctgcag gtagagcggg gtgcgcggcg 25080
 acgacggccg tcccttgggg gacagcgggc tgtaggggtg tagggttggg gcactctctg 25140
 atcgtccgaa cggggtgtct gcgccgtcg tggccgcctt ccggggggac cctcggctgc 25200
 cgaagggtc agggatcgag ctggagctgt accggggcg ctgtggggag gccagggcat 25260
 tgagggatgg atcaaaggag acattagtgg aagggttgg gtgtgggcgg ggggtgtcaag 25320
 agagatcact ggaggtcaac ccagaggagg ctgaccggcc atggaaattc aggcacagag 25380
 agcccagggt agtagtggtg gggagacagc cctgaatcag cactgtggct agcccattac 25440
 tctatgtcac ctttatgcca cttaggtaaa cacctctttc cttctgaggg tccctttaga 25500
 tgtccacttc cactggtccc ctcttttcta tttctttctt tctttctttc tctctctttc 25560
 tttctttctt tttttcctc tctctccttc cttcctttct ctctctctcc ttcctcctc 25620
 ccctccctcc ctgcttgctt gctttctctc tctctcttc tttctttctt tctttctttc 25680
 tttctttctt tctttctttt ctatctcggc tcattgcagc ctcaacctcc ctggcttagt 25740
 gtgatcctcc cacttcagcc tccaagtag ctgggattac aggtatgcac caccacacct 25800
 ggctaacttt tgtattttta gtagagacag ggtttcacca tgtagccag gctggcttta 25860
 aactcctgac ctcaagtgat ccgcctgtct ctgaaagtgt tgagattaca ggcgtgaacc 25920
 accgtgcca gccagatttt taaaaaatca tttgtagagg ctggctctca actcttagtc 25980

tcaagcaatt ctctcacctc gccttccaaa gtgctgggat tccaggctctg agccatcgcg 26040
 cctggccttg tccccTTTT tcaagttccc ttgaagagcc cacaacctgc ataactatat 26100
 ggggcaattt tgcttgaat ccaggcctct ggtctggact gtggcgagag gctggccttg 26160
 gagatcaagg tgggaaccag gcttacccta gaaggggggtc cggcctgcgg gccaggaggc 26220
 gcgggagagt ctgaccacag cgactccagc tgcttgggtca gttcatccac cttggccgcc 26280
 gccgtgtcca gctccatctg cttcagatcc atgtgtttca tggccagcgc tgggaagggtg 26340
 ggagtggagg taaggacctg gcctcctggc agggggccggc ctgagcacc ctcgcccgt 26400
 gccgagggtcc ccgcctcgcc agccccgccc cctactccag cttacactgg aagtccatgt 26460
 ccagaaagtc ccgcgcgctc tggaatgcct cgctgtccat ggtgccggcc ggagcgggcg 26520
 cctgcatggt ggggagggag ggagctggct aagacccgc ccctctagac cccgccctca 26580
 gggagtcaga cgccgtcagg agcgggacaa cgcctcaact cagttccttc ccctggaagc 26640
 cctttaccct ttcacctccc cagctgggaa atgccaaact ctccaaagcc aagtccatgc 26700
 gccacggaga agtccaaacc cagtctaaaa cctccggaat tactttctc tttctttttt 26760
 tctttctttt tttttttttt ttttgtgtat gtgtgtgaga cagagtctcg ctctgtcgcc 26820
 caggcgggag tgcaatgacg cgatcttggc tactgcaac ctccgcctcc cgggttcaag 26880
 caaatcttct gcctagctgg gactacaagc gcgcgccatt atgccggct aatttttgta 26940
 gttctgggat tacaggagt agtctccgcg cccggccgtg tccatctctt tatctcagtc 27000
 ctaagacctg aatcactcct tgaacaatta tctattgatc acctacaatg tgccggtaaa 27060
 cataggatgg aataactatg aattactgaa tgtttactag ggaccaggac gactgtgct 27120
 agatcctgtt tttgtttgtt tttgagatgg tgtctcgcat tttcgcccag gctggagtgc 27180
 agtggcgga tctcggtca ctgcaagctc cgcctccagg gttcatgcca gtctcctgtc 27240
 tcagcctccc gagtagctgg gactacaggc gcctgccacc atgcctggct aaattttgt 27300
 attttttagta gagacggggt ttcaccgtgt cagccaggat ggtctcgatc tcctgaccgc 27360
 gtgatccatc tgctcggcc tcccaaagt ctgggattac aggcgtgagc caccgcgcc 27420
 ggcccttggt tttgtttttt aataataatt ctgctgtctg ctgtgtacta gaacctatgc 27480
 ctactgcttg gggataatg tagtaaatgt agtaaaaaca atatccgccg ggcgcgggtg 27540
 ctacgcctg taattccagc actttgggag gccaaggagg gcggatcacg aggtcaggag 27600
 agcgagacca tcctggctaa catggtgaaa cccgtctct actaaaaata ccaaaaatta 27660
 gccaggcgtg gtgatggacg cctgtagtcc cagctactcg ggaggctgag gcaggagaac 27720
 ggcgtgaacc cgggagggtg agcttgaact gagcggagat cgcgccactg cactccagcc 27780
 tgggagacag tgagagactc cgtcttaaaa caaacaata aataaatatg tttaaaacaa 27840
 caacaacaat aaccagccag gcgcgggtgt tactcctgt aaccgagca ctttgggagg 27900
 ccgagggtga tggatcgctt gaagccagga gaccagcctg gccaatatgg tgaaaccccg 27960
 tctctacaaa aaaatacaaa agttagctgg gcatggtggc atgtgcctgt aatcccagct 28020

actcaggagg ctgaggcaca aggctcactt gaacctggga ggcacagggt gcagtgaagca 28080
 tagattgtgt cactgcactg cagcttgggt gacagagcga ggctctatattt aaaaaaaaaa 28140
 aaattaattg agggggccact cccttctaga gtggtgagaa atgccgtgca ccgaaagctt 28200
 catttgatgg tcaaaaccac cctagcaggc aagaaagcat ggctcagaaa catatgttca 28260
 aggtcacctt gcaagaagtc ggtagtaatc gggttcacac ccgcatctaa cttattctgg 28320
 gtcattctta ccagattaga ggggtcctag aggggaagcga ctgctcagct tcctttccct 28380
 aggggtcccca ttcagtggag gtctggctct cactgaccca ttgttagcaa gaggaacagg 28440
 gaggtggcca ggggtggagg ggcagctgtg gtcactggcc cagtgggagg gagctaggcc 28500
 actaggaacc ggtcaggcca gcaccatccc tatcccatg ctagccacca caccaccag 28560
 ctctgccacc tccctgctgc atcgaccact tagctctggc agtataggca gcagggcagg 28620
 ctggggcatg ctgatacccg cctctgtctg ggaagtcgaa ggaacagaac ctgttcaggc 28680
 tggcggtca tttggatgaa caggagtggt gtgaccttgg gcgttgagtc ctctccactc 28740
 cctgggcctc agtctcccca acatcaaaga agaaggcaaa tcaccttttt tttttttttt 28800
 gagataggggt ctgctctgtt aaccagggt acaattgtga ctactacag cctcttgacc 28860
 tcccagctca agtggctctc ccacctcagc ctctgagta gctgagacta taggtatagc 28920
 ctctgaccac cacaccagc taattttttt tttttttttt tttttttttt tttttttgag 28980
 acggagtctt gctctgtcgc ccaggctgga gttcagtggc gggatctcgg ctactgcaa 29040
 gctccgcctc ccgggttcac gccattctcc cgcctcagcc tccaagtag ctgggactac 29100
 aggcgcccgc cactacgccc ggctaatttt tgtatttttag tagagacggg gtttcaccat 29160
 tttagccggg atggtctcga tctcctgacc tcatgatccg cccgcctcgg cctcccaaag 29220
 tgctgggatt acaggcgtga gccaccgccc ccggccaccc agctaatttt ttaaaaacat 29280
 tttgtacact ttgggagggt aaggcgggag gatcacgagg tcaggagctc gagaccatcc 29340
 tggctaacac aggtgaaacc ctgtctctac taaaaaatac aaaaaatta gctgggcgtg 29400
 gtggcgggcg cctgtagtcc cagctactcg ggaggctgag gcaggagaat ggtgtgaacc 29460
 agggaggcgg agctttcagt gagccgagat cgcgccactg cactccagcc tcggagacag 29520
 agcgagactc cgtcccaaaa aaaaaaaaaa aaaaaatttg tagagacaga tcaagtctca 29580
 ctttgttgct caggctggtt ttgaactcct gggctcaagc aatcctccc cctcagcctc 29640
 ccaaagtgt gagattacag gcatgagcca ccacacctgg ccaaatcagc tattctgaaa 29700
 ggccccttta atctctatga gcccagact ttcaaactgt aaggacctta ggactgtaac 29760
 taaagtctta cagagcctaa acccctcagc taaagagcct attgttgga agttctgagt 29820
 ccaagattct atctttggaa cattctagaa ttctccaatt tgtctaacc agaattctga 29880
 gtctttctgt accacattct acctaacca gggttgact gctctggaag tctagatgga 29940
 tggatatgtg cagctggtaa aagcatgagt aagaagtcag acttcaaaaa ttcaaatctg 30000
 agggccgggc atggtagctt ctgcctgtaa tccttgact ttgggaggcc gaggggggag 30060

gatcacttga ggccaggagt tcaagaccaa catggccaac acaatgagac cccattttctt 30120
aaaaaaaatt aaaataaaat catcaaactt ggcagcacca ccgtccaacc ctgaccacag 30180
tacctcagtc tcgtaatccg taaaatgggg atgaaagtcc acctcatagg actactgtaa 30240
gaatccacct gggtcagaagg tgcaggaaga attcagagct ctgagaattg aggcctcagg 30300
aagaagagac tacaggaata aaaactcggg catttagaat ttcagagata cacaacaat 30360
actttgttaa ctgttaaaat agataaatga gcaagtctgt gcagccctaa tgccagctgt 30420
aagtgactct ttttttttct tttggtagag atttagtctc tctcgcgcct gtggttaggc 30480
tggtctcgaa ctccatagcct catgggatcc tccccggctc gatctcccaa agtattggga 30540
ttacaggcgt gagcacggcg ccatgatccc caaatttcca agattctcag attccatact 30600
gacatttctt ggctctcagg aaatgccaac cctgggtgtg gggctgtcgc ggggacaggc 30660
gggtggggacg tcggagccac cagggggcgg tcacgcccgg acccccgcca ggagggcgga 30720
ctgcgcctga gctcaggccc ggggaatgcg cagcgggccc gggcagggtg tgtacatccc 30780
ggggcaaggg agctgggccc ggcggggtac aaggggcggg cgcggggggtg gcgcggggcg 30840
tgtgtctgtt cccaggcctc tgcccctgac ctctgcctcc gagtcctctc ccatgtgctc 30900
ccctctagct ctagctccga gctctccgc gggctctggg ccagccgag gtactctccc 30960
ctgggctcct ctctccgctc caccctggc tctcctccc tggcctcctc tgcaccccag 31020
ccaggttctt tagggctaag gatcctgtgg acttctgga ggagtcactc tcagtaggaa 31080
ccgggtcaga gagccagact gagctgggaa caccaggct ggactcctac agccctgtcg 31140
gggtcacactg aatctggaga ggctccactg tctctgggac tcggtttcct cttttgtgga 31200
cgtctatgga atgggctagg gcctttcttg ctctaagcct ctacttgggc ttgttattta 31260
gcttctctgt gcctgtttcc tcatgtggac catgggaaga attaatacct tcgcctcaaa 31320
ggggtatgag gattgagtga cataatttat aagccgtgat tagaacaatg cagtgcgcga 31380
aataaagttc acacatacag gattcataat taccagatgt ccttggtgtg tcattataat 31440
aacacagggg ctggcaacag agtgaggggt ccagactcaa tgtaattttt ttttccccta 31500
aaagggccct ttcaactctt tctgagatca tacaagccct gagttttgac acccagggtc 31560
tcaacttcct gagcccttgc ctctcagagt cctaaatttc ccctgtacat tcctgagtct 31620
ggccagtgat caccctcagt cacttaggga cgggaggggt gggagagccc tggaagattc 31680
cagacagaag ctggcaaaaag cccaggggtg gggcaatatc cactctccag cctccgtttc 31740
tccactcgta atgaggagtc cttccctggg gtcagcaaac cttattcaaa gggagacctc 31800
tcagtcaccc aagattcctc tagacaatgc gagctttcct acctacctac ctaccagctc 31860
tgagcttggg acaccagag ccctgttttg gcaaccacgg ttattatttt taatttcatt 31920
tcaggttatc atcaaatgcc cttcaagccc agacattggg aaacactcct ctctcatcag 31980
atgctcgcct ccccatctt gtttttaatc ccccttctta ggacgcatgg ggggtgagag 32040
aacggggaga tagacagagg gaggtgcctg gtcctgccct ccccccgcct caaggacaga 32100

cagacacctc cagaattagc ctctgtccct ctttatctcc cacaataccc caggtcagac 32160
 agatgggcgt ggaggtgaca tttctcacct cagggtcagg gcaaggagcc ctgaggcaga 32220
 aggttagtca gaaaatctgg cgggggcgga tggaatcccg tccccagag agctgcagaa 32280
 gaaggaggag gcagaatcct gaccctacaa actctactgc ctgtgtgagc tccaagcctc 32340
 agtttacccc ttcctctccg tgtaatgggt aaatgcccgg ctatgcaaac ctcccagaat 32400
 ccaatagccg ctttccggaa ttctgcccctg gggttctagaa ctacctctgc aaaccagct 32460
 gtttcccacc ccataaggca ataggggagc ccacctccgc caggggggtgc cctagggcgg 32520
 atgtcccttc tctgggttagg caggtctgac gcccagggtta atgacatggt gggttcgctc 32580
 agcggcacag aggaggttgg agatctgcct cgggtgtttc tctcctacc cgccccatc 32640
 cccgagccga aaagtcgggg gagagccggg acacagcctc cggagggacc ccgggtacct 32700
 gtctgtctcc acttcaggaa cccaggctcc actatccctg cccaccctt aattctgctc 32760
 agagacctag aagatcggtc gagacagcag cttgaggctg gcaggggtgg caccattcc 32820
 accttgagcc ccaccagtct gaggctctca tttctgacca agactcgggg attcgaacct 32880
 ctatactacc caaagactcg gcttcctaga gccccccagt tcgagggact caggaattcc 32940
 agctccaacg tctccccggg atgaaggggt agaatccctc cattccaaga attcaggcat 33000
 ccgaaccgcg tttccttccc tccagtaaaa caggcaacgg agtttccttc taaggatcca 33060
 ggtgtcggcg cgccccaaat tccgccctgg gacctggcgt ccgagtcccc tccaatcct 33120
 cccagggacg cgggtgttgg gctttttcag ggcctctggg cccaggagg gtgaaactca 33180
 cggatccggg cagatcctgg cacctggggg cttcctccag ctcgggctcc ggcttgggga 33240
 gcggagaacg gggcggggca ggagctggga acaggttaga cgacgtgact tgggctggag 33300
 ggaggcgggt cccgggtggg agggggagcc aaggctgcct cgagcacctt gggacttgta 33360
 gtcccgagg gacaggacgt agcccaagac gatcccattt ggattcacc agagtccatt 33420
 tcacagacag gaaggcgag gcccagaagc cgagagcgac caggccaggg agatacagaa 33480
 gagccgagac gcctgcctcg ctgtggctgg agactgactc ctgagccctt gccccacccc 33540
 ttcaggcgca ctatcccctt tcctgatcag tatccccag ggtctctgag cccgaatctc 33600
 cccgtcgata aaaagcgcg gttggatctt caaaggatgt cccagcaaga gttcaaaatc 33660
 ttagtttggg ctacaacccc cagcagcctc cgcgaccgcc ctcgggcgac tctttgcctc 33720
 gggtcctgtg ggaattgtag tcctggagcc cgcagggtg caccgcgggtg tctctctcgc 33780
 ccacgcgaag gaaaccgtct ggagatcctg gataggggaa acatttcccc tttcccttga 33840
 ccctccctcc gctctggaaa gcctctccca cctggggaga aggggtgccc caattctgga 33900
 gtaggatcct aaatcttggc agagggggcg ggaagtggcg ctgacacact ggccaggaat 33960
 gcagtcgggt caccctgtct agccaccgtc tcgcggctcc aaccgccgcc caacgcgggg 34020
 cggccccagt gggaagggaa gtgggtgctg ccccaaatc tgtgtccacg tgccgtgtt 34080
 tacacgtctc ctggggcagg gaggagtcgc cgatcaggct cttcctgaa agtcatcgag 34140

gtttccacg catgagacta aacccccgag ggcattctaca agtcccattt gatccacaaa 34200
 cgctacaccg tgcccagcac cactccacgc gtgtggggct cctgggtccg aggctccgcc 34260
 ctcgagaacc acaagctcct cccctatgt ttcccgtcc cccggagtcc agaagccccg 34320
 cccctggctg gaacttcacg ccctccggac ggattgcccc tatttctcca ttttccgct 34380
 tctcccagtc aagtcttgaa cttgtgaggc atctgggcct cccagaaga catttaacac 34440
 agaaagcaca gccctactaa ctagtattct tacctgtctc ttcaagaatt tcagaccaat 34500
 cgaccgtcct gtctctttaa ggcttaggaa gagcagtgtg gctgcccctt taaggaggcg 34560
 ttgcaacaaa ccatattgga cagacgatgg gggcgaccca tcgggacccg acgggcctct 34620
 gactccagca atacagcgaa tcagcggctt tcgggaatac attttctgga aaaagacttc 34680
 ttctcgggtt ttctgctctg cacacgttga aattttcccc agtttttctt gcagatcggg 34740
 agtcgagcaa tgcctacccc cgcgtcccg caccagttgg gcgtcccg atgatgccct 34800
 acccctttgg atccacgtgg tctgcaacct ggtgcgagca gcccgggcta cagggttgcc 34860
 tgagggtgtg gtcccaggat ggaggagccc caggccggcg gtgaggggtg gggttgacgg 34920
 ggtgcggagg gtgcgttggt ggaaggagaa agggcgctcc gagaggggtt gggcggaaaa 34980
 ggaggcgtag ctgcaagcag gacttgcgaa gagcgtgcat tcccagtggg cgaacgggaa 35040
 ttcgaacgga gagaggggtt tctgtgggg ggctacccgt ggagagcaag gcgccccag 35100
 gggttggtc ggtgaaattg aggtcgcccc tggggaacag gtgggcagaa aggagaaacc 35160
 aggttgaggg gactggagtg ctcacaggt taagaccaat ggaccgatag gcgcgccctg 35220
 caagattgga cgggcaagga ggtgtcagtc gacccattt ccccttctgc tgcagatgct 35280
 gctcggttct cttgtcccc caactttacc gcgaagcccc cagcctcaga gtcccctcgt 35340
 ttctccttgg aggcgtgac gggccagat acggagctgt ggcttattca ggcccctgca 35400
 gactttgccc cagaatggtg agtggtcttg ttgacggaaa agaggggtccc ggtccagacc 35460
 ccaagagcgg gttcttgaat ttgtcacagg aaagaattag aggtgagtca cagagcacag 35520
 tgaaagaaac aagtttattg gaaactactc ctttacagag tagagtgtcc tcagaaagca 35580
 gggggagaaa cccacagccc tttgttagta tttctactta taagaaacta taaggaacta 35640
 tagttaaact tggagtgtgc agataagctc actaaaggta ggggctattg gtgttatcca 35700
 cgaccattaa tcctgcaacc taagcttgct ctttatgtt atatttaagt aatgggggct 35760
 gcattcttag gacatttgga cattctgcag gcttggtgga acatgttctg tatggccata 35820
 aatattctgt aattataatt ggtggtcagc ctgggatgtg gttattttca ggccataagc 35880
 atgaaccttg taagtgccta gctactcact ttaagatgga gtcactctag tcatgtttta 35940
 ttaaaaacca gaggccagcc aggcgcagt gctgggtgcct gtaatcccat cctttgggag 36000
 gccgaggcga gcagatcact tgaggtcagg agttcaagac cagcctggcc aacatagtga 36060
 aattgtctct actaaaaata caaaaattg ctgggcgtgg tggcagggtg ctgtaatccc 36120
 agctacttga gaggctgagg caggagaatc gcttgaaccc aggaggtgga cattgcagtg 36180

19

agccgagatc atgccactgc actccagcct aggcaacaga gcaagactct ctcaaaaaaa 36240
 aacaaaaaaa aaatcaaaaa accttccttc tcctgttcca cttaagcctc tgccctccct 36300
 gtttctctct gtagcttcaa tgggcggcat gtgcctctct ctggctcca gatcgtcaag 36360
 ggcaaatttg caggcaagcg gcaccgctat cgagtcctca gcagctgtcc ccaagctgga 36420
 gaagcgaccc tgctggcccc ctcaacggag gcaggagggt gactcacctg tgcctcagcc 36480
 cccagggca ccctaaggat ccttgagggt cccagcaat ccctgtcagg gagccctctg 36540
 cagcccatcc cagcaagtcc cccaccacag atccctcctg gcctgaggcc tcggttctgt 36600
 gcctttgggg gcaaccacc agtcacaggg cctaggtcag ccttggcccc caacctgtc 36660
 acctcagga agaagaaaaa ggagatgcag gtgacagagg cccagtcac tcaggaggca 36720
 gtgaatggg acggggccct ggaggtggac atggctttgg ggtcgccaga aatggatgtg 36780
 cggaagaaga agaagaaaaa aaatcagcag ctgaaagaac cagaggcagc agggcctgtg 36840
 gggacagagc ccacagtgga gacactggag cctctgggag tgctgttccc gtccaccacc 36900
 aagaagagga agaagcccaa agggaaagaa accttcgagc cagaagacaa gacagtgaag 36960
 caggaacaga ttaacactga gcctctagaa gacacagtcc tgtccccgac caaaaagaga 37020
 aagaggcaaa aggggacgga agggatggag ccagaggagg gggtagacagt tgagtctcag 37080
 ccacaggtga aggtggagcc actggaggaa gccatccctc tgccccctac gaagaagagg 37140
 aaaaaagaaa agggacagat ggcaatgatg gagccaggga cggaggcgat ggagccagt 37200
 gagccggaga tgaagcctct ggagtccca ggggggacca tggcgctca acagccagaa 37260
 ggagcgaagc ctcaggccca ggcagctctg gcagctcca aaaagaagac gaagaaagaa 37320
 aaacagcaag atgccacagt ggagccagag acagagggtg tggggcctga gctgccgat 37380
 gaccttgagc ctcaggcagc tcccatcc accaagaaga agaagaagaa gaaagagaga 37440
 ggtcacacag tgactgagcc aattcagcca ctagagcctg aactgccagg ggagggacag 37500
 cctgaagcca gggcaactcc gggatccacc aagaagagga agaagcagag tcaggaaagc 37560
 cggatgccag agacagtgcc ccaagaggag atgccagggc cgccactgaa ttcagagtct 37620
 ggggaggagg ctccacaggg ccgggacaag aagcggaagc agcagcagca gcagcctgtg 37680
 tagtctgccc ccgggaaact gaggaactaa agaaagctga aggtgcccac ctggggccacc 37740
 agaaggtgac acccccagaa tccctccca gagactgcac cagcgagcc 37790

<210> 2
 <211> 27
 <212> DNA
 <213> Homo sapiens

<400> 2
 gattgtcatg tkacatcagc caatact

27

<210> 3
 <211> 26
 <212> DNA

20

<213> Homo sapiens

<400> 3
cacagtgtaaa cyccatctct actaaa

26

<210> 4
<211> 27
<212> DNA
<213> Homo sapiens

<400> 4
aaaattagcc ggrcgcatg gcgggag

27

<210> 5
<211> 31
<212> DNA
<213> Homo sapiens

<400> 5
agcctggcca acatgstgaa accccgtctc t

31

<210> 6
<211> 26
<212> DNA
<213> Homo sapiens

<400> 6
aagtttctct attktgttta taaaca

26

<210> 7
<211> 31
<212> DNA
<213> Homo sapiens

<400> 7
ttctcctgac ctctgatcc gccacctcg g

31

<210> 8
<211> 41
<212> DNA
<213> Homo sapiens

<400> 8
gggattacag gcatgcrcca ccaggcccag ctaatttttg t

41

<210> 9
<211> 33
<212> DNA
<213> Homo sapiens

<400> 9
tccaatggtg acamcagtaa gagcagttaa cag

33

<210> 10
<211> 33
<212> DNA
<213> Homo sapiens

<400> 10
tccaatggtg acaasagtaa gagcagttaa cag

33

21

<210> 11
<211> 38
<212> DNA
<213> Homo sapiens

<400> 11
ttgcctcagc ctcctgakta gctgggattg gaatgaga 38

<210> 12
<211> 45
<212> DNA
<213> Homo sapiens

<400> 12
aaaataataa taataatatt aaycctgacc ttggcgccac catct 45

<210> 13
<211> 36
<212> DNA
<213> Homo sapiens

<400> 13
cctcatgagc cacccacytc ggcctcccaa agtgct 36

<210> 14
<211> 33
<212> DNA
<213> Homo sapiens

<400> 14
tgagccaccg cgcccrgccg agactcacta ttt 33

<210> 15
<211> 35
<212> DNA
<213> Homo sapiens

<400> 15
ggagagaagg agcagagaac mtctctatgt ggcca 35

<210> 16
<211> 35
<212> DNA
<213> Homo sapiens

<400> 16
tttcccagca tcccaytgca atgaggctcc tggcc 35

<210> 17
<211> 31
<212> DNA
<213> Homo sapiens

<400> 17
tcctgactcc agtgmgtgc ctacagtcct g 31

<210> 18
<211> 34
<212> DNA
<213> Homo sapiens

22

<400> 18
tccagcctgg gcaagaasag tgaaactcca gctt

34

<210> 19
<211> 21
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> n is ATTTT repeated any number of times

<220>
<221> misc_feature
<222> (8)..(8)
<223> n is a, c, g, or t

<400> 19
ggtttatntg agatggattt t

21

<210> 20
<211> 18
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> n is c or n is GACCTTGGCGCCACCATCTT

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or t

<400> 20
tacgataaat agctagan

18

<210> 21
<211> 51
<212> DNA
<213> Homo sapiens

<400> 21
ttttagtaga gacatgggttc cgccaygttg cccaggctgg tcttgaactc c

51

<210> 22
<211> 51
<212> DNA
<213> Homo sapiens

<400> 22
ctggggaggc tgaggcagga gaatcrcttg aaaccgggag gcggagggttg t

51

<210> 23
<211> 51
<212> DNA
<213> Homo sapiens

<220>

23

<221> misc_feature
 <223> where n is an AA insertion or no insertion

<220>
 <221> misc_feature
 <222> (26)..(26)
 <223> n is a, c, g, or t

<400> 23
 actaaaaata aaaaaataaa aaaaanatatg ccgagcatgg tgggtggggtg c 51

<210> 24
 <211> 51
 <212> DNA
 <213> Homo sapiens

<400> 24
 aaaaaactaa agtgggggttt gcgggkagtg ggagggccct tcctgctagg t 51

<210> 25
 <211> 51
 <212> DNA
 <213> Homo sapiens

<400> 25
 caggcgatc acaaggtag gagttagaga ccagcctggc caacacagtg a 51

<210> 26
 <211> 51
 <212> DNA
 <213> Homo sapiens

<400> 26
 ccctatgttg tccaagctgg cagagrtttt tgtttgttt tttgagaggg a 51

<210> 27
 <211> 51
 <212> DNA
 <213> Homo sapiens

<400> 27
 ctcgggaggc tgaggcagga gaatcrcttg aactcaggag gcagaggttg c 51

<210> 28
 <211> 51
 <212> DNA
 <213> Homo sapiens

<400> 28
 ggagcttgca gtgagctgag atcgcrccac tgcactccag cctgggacgac a 51

<210> 29
 <211> 51
 <212> DNA
 <213> Homo sapiens

<400> 29
 tacaggcgcc cgccaccacc cccagmtaat ttttgatatt ttagtagaga c 51

<210> 30
 <211> 51

24

<212> DNA
 <213> Homo sapiens

<400> 30
 tcgtcctgct acagaattac aggca ygcgc caccgctccg ggctaatttt t 51

<210> 31
 <211> 51
 <212> DNA
 <213> Homo sapiens

<400> 31
 taaagcggga ggatggcttg aacctr ggag gcggaggttg cagt gaggccg a 51

<210> 32
 <211> 30
 <212> DNA
 <213> Homo sapiens

<400> 32
 atcctaaaga ctacmtttcc cagcatccca 30

<210> 33
 <211> 36
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Where n is an insertion of ACACCTGTGGTCCCAGCTACTCTGG
 AAGCTGAGGTGGGAGGATCGCTTGAGCCCAAGAGGTGGAGGCTGCAGTGAGCTGT or

no
 insertion

<220>
 <221> misc_feature
 <222> (16)..(16)
 <223> n is a, c, g, or t

<400> 33
 gctgcagtga gctgtngact gtgccactgc actcca 36

<210> 34
 <211> 39
 <212> DNA
 <213> Homo sapiens

<400> 34
 tgacagtaga catcctgtca trataagtct ttttttttt 39

<210> 35
 <211> 41
 <212> DNA
 <213> Homo sapiens

<400> 35
 catccccata ccaaccacc ytactgctct gatctcctcc t 41

<210> 36
 <211> 40

25

<212> DNA
 <213> Homo sapiens

<400> 36
 attaagtgcc ttcacacagc atctggttta atgtttataa 40

<210> 37
 <211> 46
 <212> DNA
 <213> Homo sapiens

<400> 37
 ttgggagacc aaggcaggtg gatcattttg aggtcagtag atcaaa 46

<210> 38
 <211> 39
 <212> DNA
 <213> Homo sapiens

<400> 38
 gggaggctcg aggcgggcrq attgcatgag ctcaggatt 39

<210> 39
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 39
 ccctccctgc ttgcttgctt tctctytctc tctttctttc tttctttctt 50

<210> 40
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 40
 ccctgcttgc ttgctttctc tctctytctt tctttctttc tttctttctt 50

<210> 41
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 41
 agaacctgtt caggctggcg gctcayttgg atgaacaggg agtgtgtgac 50

<210> 42
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 42
 cccccttctt aggacgcatg ggggtkgaga gaacggggag atagacagag 50

<210> 43
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 43
 tgcgagcagc ccgggctaca gggtttrcctg aggtgtgggt cccaggatgg 50

<210> 44
<211> 50
<212> DNA
<213> Homo sapiens

<400> 44
ggcgcctcaa cagccagaag gagcgragcc tcaggcccag gcagctctgg 50

<210> 45
<211> 42
<212> DNA
<213> Homo sapiens

<400> 45
cttgctacag aattacaggc aygcgccacc gctccgggct aa 42

<210> 46
<211> 31
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Where n is insertion of A or no insertion

<220>
<221> misc_feature
<222> (13)..(13)
<223> n is a, c, g, or t

<400> 46
ctaaagacta cantttccca gcatccatt g 31

<210> 47
<211> 22
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 47
agtcacagct cactgcagcc tc 22

<210> 48
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 48
acctcttggg ctcaagcgat cctc 24

<210> 49
<211> 29
<212> DNA
<213> Artificial

27

<220>

<223> Synthetic construct

<400> 49

aaaaaaagac ttatcatgac aggatgtct

29

<210> 50

<211> 27

<212> DNA

<213> Artificial

<220>

<223> Synthetic construct

<400> 50

gcaagactcc gtcccagaaa aagaaaa

27

<210> 51

<211> 26

<212> DNA

<213> Artificial

<220>

<223> Synthetic construct

<400> 51

tcctctctct cccccagctc attttg

26

<210> 52

<211> 24

<212> DNA

<213> Artificial

<220>

<223> Synthetic construct

<400> 52

acccaccct actgctctga tctc

24

<210> 53

<211> 28

<212> DNA

<213> Artificial

<220>

<223> Synthetic construct

<400> 53

aggctggctct tgaactcctg ggcttaag

28

<210> 54

<211> 17

<212> DNA

<213> Artificial

<220>

<223> Synthetic construct

<400> 54

ggttccgcca cgttgcc

17

<210> 55

28

<211> 29
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 55
tcggctattt ttttttttat ttttttatt

29

<210> 56
<211> 22
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 56
attacaggca cccaccacca tg

22

<210> 57
<211> 36
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 57
acccctttt agtttttttt tcctctagt atcgcc

36

<210> 58
<211> 18
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 58
gccctccac tacccgca

18

<210> 59
<211> 23
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 59
agacgggggt tcactgtgtt ggc

23

<210> 60
<211> 22
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 60

aggctgggtct caaactcctg ac	29	22
<210> 61		
<211> 23		
<212> DNA		
<213> Artificial		
<220>		
<223> Synthetic construct		
<400> 61		23
caaacaaaca aaaacctctg cca		
<210> 62		
<211> 27		
<212> DNA		
<213> Artificial		
<220>		
<223> Synthetic construct		
<400> 62		27
cttgacaac ataggagac cctgtgt		
<210> 63		
<211> 22		
<212> DNA		
<213> Artificial		
<220>		
<223> Synthetic construct		
<400> 63		22
tgccctcagcc tcccagtag ct		
<210> 64		
<211> 23		
<212> DNA		
<213> Artificial		
<220>		
<223> Synthetic construct		
<400> 64		23
cctcctgagt tcaagcgatt ctc		
<210> 65		
<211> 26		
<212> DNA		
<213> Artificial		
<220>		
<223> Synthetic construct		
<400> 65		26
tgcccttcaca cagctctggg ttaatg		
<210> 66		
<211> 26		
<212> DNA		
<213> Artificial		

30

<220>
<223> synthetic construct

<400> 66
tgccttcaca cagcactggg ttaatg

26

<210> 67
<211> 23
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 67
accatgttgg ccaggctggg ttt

23

<210> 68
<211> 26
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 68
atctactgac ctcaaagat ccacct

26

<210> 69
<211> 15
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 69
tgcaatccgc ccgcc

15

<210> 70
<211> 27
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 70
ccaggctggg ttggaaatcc tgagctc

27

<210> 71
<211> 20
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 71
ctgagatcgc accactgcac

20

<210> 72

31

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

<220>
<223> synthetic construct

<400> 72
gggaggcgga gcttgcagtg a 21

<210> 73
<211> 21
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 73
gcgcatgcct gtaattctgt a 21

<210> 74
<211> 25
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 74
caggacgagc cacagacaaa actcc 25

<210> 75
<211> 19
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 75
tgcaatgagg ctcctggcc 19

<210> 76
<211> 22
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 76
actacaattt cccagcatcc ca 22

<210> 77
<211> 17
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 77

cctccctccc tccctgc	32	17
<210> 78 <211> 18 <212> DNA <213> Artificial		
<220> <223> synthetic construct		
<400> 78 tgcttgcttt cttcttct		18
<210> 79 <211> 22 <212> DNA <213> Artificial		
<220> <223> synthetic construct		
<400> 79 tccctgcttg cttgctttct ct		22
<210> 80 <211> 22 <212> DNA <213> Artificial		
<220> <223> synthetic construct		
<400> 80 tctctctttc tttctttctt tc		22
<210> 81 <211> 20 <212> DNA <213> Artificial		
<220> <223> synthetic construct		
<400> 81 tgttcatcca aatgagccgc		20
<210> 82 <211> 29 <212> DNA <213> Artificial		
<220> <223> synthetic construct		
<400> 82 agcctgaaca ggttctgttc cttcgactt		29
<210> 83 <211> 24 <212> DNA <213> Artificial		

33

<220>
 <223> Synthetic construct

<400> 83
 caagctgcta tctcgaccga tctt

24

<210> 84
 <211> 20
 <212> DNA
 <213> Artificial

<220>
 <223> Synthetic construct

<400> 84
 ggggtgaccac cctgccagcc

20

<210> 85
 <211> 21
 <212> DNA
 <213> Artificial

<220>
 <223> Synthetic construct

<400> 85
 cgggctacag gggtacctga g

21

<210> 86
 <211> 22
 <212> DNA
 <213> Artificial

<220>
 <223> Synthetic construct

<400> 86
 tctgcaacct ggtgcgagca gc

22

<210> 87
 <211> 20
 <212> DNA
 <213> Artificial

<220>
 <223> Synthetic construct

<400> 87
 tgaggctccg ctccttctgg

20

<210> 88
 <211> 21
 <212> DNA
 <213> Artificial

<220>
 <223> Synthetic construct

<400> 88
 agctgccaga gctgcctggg c

21

<210> 89

34

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

<220>
<223> synthetic construct

<400> 89
gcttgtaatc ccagctactc g 21

<210> 90
<211> 19
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 90
tgccctaccc ctgtaatcc 19

<210> 91
<211> 21
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 91
gcttgtaatc ccagctactc g 21

<210> 92
<211> 19
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 92
caacactcac accccacag 19

<210> 93
<211> 19
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 93
agatcacgcc actgcactc 19

<210> 94
<211> 20
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 94

ttgacaattg agcaaagagc	35	20
<210> 95 <211> 19 <212> DNA <213> Artificial <220> <223> synthetic construct <400> 95 ttggattaca gacgtgagc		
<210> 96 <211> 18 <212> DNA <213> Artificial <220> <223> synthetic construct <400> 96 agtgcagcct caacttcc		
<210> 97 <211> 21 <212> DNA <213> Artificial <220> <223> synthetic construct <400> 97 ccagtccaaa caatatgatc c		
<210> 98 <211> 21 <212> DNA <213> Artificial <220> <223> synthetic construct <400> 98 catgattcac tgcacccaac c		
<210> 99 <211> 21 <212> DNA <213> Artificial <220> <223> synthetic construct <400> 99 tttcactctt gttgcccaag c		
<210> 100 <211> 21 <212> DNA <213> Artificial		

36

<220>
<223> Synthetic construct

<400> 100
caccatgcct ggctccaatg t 21

<210> 101
<211> 17
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 101
agcccaggaa ttcaagg 17

<210> 102
<211> 17
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 102
agaacattgg agccagg 17

<210> 103
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 103
agaatcactg gaatccagg 19

<210> 104
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 104
ttttcacaca ggtccaatcc 20

<210> 105
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 105
actgcaacct ccatctcc 18

<210> 106

37

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

<220>
<223> Synthetic construct

<400> 106
caattaagtg cttcacaca gca 23

<210> 107
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 107
ggtcaagagt tcaagaccag c 21

<210> 108
<211> 17
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 108
ccctgcccc cctctcc 17

<210> 109
<211> 30
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 109
agtcaatttc tgtgcaaact acttttattt 30

<210> 110
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 110
gaggcaacag gaacaaacc 19

<210> 111
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 111

38

cattggaatg agcagaaacc

20

<210> 112
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 112
taacataaag aatcaggagg aggc

24

<210> 113
<211> 20
<212> DNA
<213> Artificial

<220>
<223> ssynthetic construct

<400> 113
agttggctca tctgcctctt

20

<210> 114
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 114
tggctaacac ggtgaaacc

19

<210> 115
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 115
ggaatccaaa gattctatga tgg

23

<210> 116
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic construct

<400> 116
actcctgact tcaaatgatc c

21

<210> 117
<211> 19
<212> DNA
<213> Artificial

39

<220>
<223> synthetic construct

<400> 117
tagccccag tcacgttcc 19

<210> 118
<211> 21
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 118
agaagtccaa gagtttgag c 21

<210> 119
<211> 21
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 119
ttctcagtcc cagaatgaac c 21

<210> 120
<211> 21
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 120
ccacttaggt aaacacctct t 21

<210> 121
<211> 21
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 121
ctgcaatgag ccgagataga a 21

<210> 122
<211> 21
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 122
ccacttaggt aaacacctct t 21

<210> 123

40

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

<220>
<223> synthetic construct

<400> 123
ctgcaatgag ccgagataga a 21

<210> 124
<211> 18
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 124
atggttggga gactgagg 18

<210> 125
<211> 20
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 125
ccgcatctaa cttattctgg 20

<210> 126
<211> 21
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 126
aactacctct gcaaaccag c 21

<210> 127
<211> 21
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 127
ttggaatgga gggattctac c 21

<210> 128
<211> 20
<212> DNA
<213> Artificial

<220>
<223> synthetic construct

<400> 128

41

ggtttttctgc tctgcacacg

20

<210> 129
 <211> 20
 <212> DNA
 <213> Artificial

<220>
 <223> Synthetic construct

<400> 129
 cctttctcct tccaccaacg

20

<210> 130
 <211> 19
 <212> DNA
 <213> Artificial

<220>
 <223> Synthetic construct

<400> 130
 ggacagatgg caatgatgg

19

<210> 131
 <211> 21
 <212> DNA
 <213> Artificial

<220>
 <223> Synthetic construct

<400> 131
 tcttcttctt ggtggatgtg g

21

<210> 132
 <211> 3475
 <212> DNA
 <213> Homo sapiens

<400> 132
 ttttagtaga gacatggttc cgccacgttg cccaggctgg tcttgaactc ctgggcttaa 60
 gggatctgcc cgccatggcc tcccaaagtg ctgggattac cagcgtgagc cactgtaccc 120
 agcctgagta taggtttctg ataaatttta ggatcatatt gtttggactg ggtaagaatt 180
 tccagaactc taatgaagaa actgactggg ttatatatta ttttatatta ttttattatt 240
 tttgagatgg attttctctc ttgttgccca agctggattg cagtggcacg atcttggctc 300
 accacaacct ccgcctccc gtttcaagtg attctcctgc ctgagcctcc ccaggagctg 360
 ggattacagg caccaccac catgctcggc tttttttttt tttttttttt ttttttagt 420
 agagacgggg ttccaccatg ttggccaggc tgggtctgaa ctcttgacct cagggtgatcc 480
 acctgccttg gcctcccaa ggcgtgggat tacaggcatg agccactgtg caaggcctag 540
 gctgggttat aaaattgcta aaccaagcag aacatgaatt aaataccaag gaaatactct 600
 cctagattgt catgtttacat cagccaatac taaaattgtc aagatacaca atttgaatga 660
 actccatggg ccaagtcgaa ttatctatga tattaccat ctaataaaca gcactatgtc 720

ccttaatggg agaaaaagt ggagaattta agagaatatc aatccaatgt tggttgggtg	780
cagtgaatca tgtctatatt cccagcactt tgggaggcca aggcaggagg atcacttgag	840
cccaggaatt caaggccagc ctcggaaca cggtagatc ctgtctctac ggaaaattaa	900
aaaaaaaaa agagagagat tagtgggatg tgggtgcctat agtcccagct acttgggagg	960
ctgaggcggg aggatcattt aagcctggga cgttgagggt gcagtgaacc atgagtgaga	1020
ctcatctcaa aaaaaaaaaa aaaatggcga tcactagagg aaaaaaaac taaagtgggg	1080
tttgccggta gtgggagggc ctttcctgct aggttgact atgatctcca gggaggctcc	1140
acgggagaat ctttccttg tctttttcag tttctagagc caaattcttt gcataccttg	1200
cattccttg ctcggaaccc cttccctaac cttcaaagct ggcagctagc ctctggctca	1260
agtgtcacat ggcctgtctc tgtcttccta tccaatcttc ctcttataag aacattggag	1320
ccaggcatgg tggctgacgc ctgtaatccc agcacttttg gagaccgagg caggcggatc	1380
acaaggctcag gagttcgaga ccagcctggc caacacagtg aaaccccgtc tctactaaaa	1440
aaatacaaaa aagtagccgg gcatgggtggc aggtgcctgt aatcccagct acttgagagg	1500
ctgaggcagg agaatcgctt gaacctggga ggcagagctt gcagtgaacc gagatagtgc	1560
caatgcagtc cggcctgggc gaaacagcga gactccgtcg caaaaaaaaaa aaaataataa	1620
taaataataa ataaaaataa aaataaaaata aaaaaataaa aataataaaa taaataaaaa	1680
ttattttgag acaaagtcta ttctgtggca gaggttgaa tgcagtggcg tgatcacagc	1740
ttactgcagc ttctacctcc tgagctcaag cgatccttcc accttggctt cctgagtagc	1800
tgggacctca ggtgtacatt accacgtca gctaattatt tattttattt ttatattttt	1860
gtgacggagt ttcgctcttg ttgcccgggc tggagtgaac tgggtgctatc tcagctcact	1920
gcaacctctg cctcctggat tccagtgatt ctctgtctc agcttcctga gtagctggga	1980
ttacaggtag atgcatcac gccagctaa tttttgtatt tttagtagag acgggggtttc	2040
atcatattgg tcaggctggc ctcgaaactc tgacctcagg tgatccacct gccttggcct	2100
cccaaagtgc tgggattaca ggcgtgaggc accacgcccg gcaatttttt ttttcttttt	2160
tttttttcag acagagtctt gctctgtcac ccaggctgga gtgcagtagc gtgatctcgg	2220
tttactgcaa cctccatctc ccgggttcaa gcgattctcc tttctcagcc tcccaagtag	2280
ctgggactac aggtgcacac caccacggcg ggctaatttt tgtattttta gtagacacca	2340
ggtttcacca tattggtcag actggctctc aactcctgac ctgaggtgat ccatctgcct	2400
cagcctcca aattgctggg attacaagcg tgagccacac acctggctta atttttttat	2460
ttttgatcga cacagggtct ccctatgttg tccaagctgg cagagatttt tgtttgtttg	2520
tttgagaggg aattttgctc ttgtagccca ggctggagta caatggtgca atcttggctc	2580
accacaactt ccgcctcccg ggtttaacag attctcctgc ctgagcctcc caagtagctg	2640
gaactacagg cacctaccac cacaccaggc taatttttgt gctttttagt agagatgagg	2700
tttcaccatg ttggccaggc tggctttaa ctcctggcct ccagtgatcc acccgcttg	2760

43

acctcccaaa gtgctgaaat tacaggcgtg agcaccgcgc ctggcctctc aacctacaat 2820
 ttcaacaccc aaggaaacag cccaccatga gtgagaacca gcagacacaa caaactatag 2880
 gattagctgc ctccaaactt cagggtgatag attatcaggc atgtacttga aactaaagga 2940
 cacaaaagaa gaatccgaaa tataaaataa aggattggac ttgtgtgaaa agaatccctt 3000
 agaaagggct actttcaggc tggccatggg ggctaattggc ctgtaatccc agcacttttg 3060
 aaggccgagg tgtgtggatc acctgaggtc aagagttcaa gaccagcctg gccaacatgg 3120
 tgaaaccccg tctctactga aaatacaaaa attagccagg tgggggtggca gatgcctgta 3180
 atcccagcta ctctggaggc tgaggcagga gaatcgcttg aactcaggag gcagagggtg 3240
 cagtgaactg agattgcgct atcgtgcccc agcctgggca ctagagttag atcaaaaaaa 3300
 aaaaaaaaaa aagaagaaga agaagaaagg gctactttca gactgccttg ccaaaaatca 3360
 taaccacaat gatgagcatg tattgagtca aaacagaatc aaaagagaag aaagtcaatt 3420
 tctgtgcaaa ctacttttat ttataaggaa agtttctcta ttttgtttat aaaca 3475

<210> 133
 <211> 4366
 <212> DNA
 <213> Homo sapiens

<400> 133
 caagaattca ggcattccgaa cccgctttcc ttccctccag taaaacaggc aacggagttt 60
 ccttctaagg atccagggtg cggcgcgccc caaattccgc cctgggacct ggcgccgag 120
 tccccccca atcctcccag ggacgcgggt gttgggcttt ttccaggcct ctgggtcccca 180
 ggagggtgaa actcacggat ccgggcagat cctggcacct gggggcttcc tccagctcgg 240
 gctccggctt ggggagcgga gaacggggcg gggcaggagc tgggaacagg ttagacgacg 300
 tgacttgggc tggaggagg cggttcccgg tggggagggg gagccaaggc cgcctcagc 360
 accttgggac ttgtagtccc ggagggacag gacgtagccc aagacgatcc catttgatt 420
 caccagagt ccatttcaca gacaggaagg gcgaggccca gaagccgaga gcgaccaggc 480
 caggagata cagaagagcc gagacgcctg cctcgtgtg gctggagact gactcctgag 540
 cccttgcccc accccttcag gcgcactatc ccctttcctg atcagtatcc cccagggtct 600
 ctgagcccga atctccccgt cgataaaaag cgcgggttgg atcttcaaag gatgtcccag 660
 caagagttca aaatcttagt ttggactaca acccccagca gcctccgcga ccgccctcgg 720
 gcgactcttt gcctcgggtc ctgtgggaat tgtagtcctg gagcccgag gggctgcacc 780
 ccggtgtctc tctcgccac gcgaaggaaa ccgtctggag atcctggata ggggaaacat 840
 ttccccctcc cttgaccct ccctccgctc tggaaagcct ctcccacctg gggagaaggg 900
 gtgccccaat tctggagtag gatcctaaat cttggcagag ggggcgggaa gtggcgctga 960
 cacactggcc aggaatgcag tcgggtcacc ctgtctagcc accgtctcgc ggctccaacc 1020
 gccgccaac gcggggcggc cccagtggga aggggaagtgg gtgcgtcccc caaatctgtg 1080

44

tccacgtgcc gctgtttaca cgctccctgg ggcagggagg agtcgccgat caggtccctt	1140
cctgaaagtc atcgaggttt cccacgcatg agactaaacc cccgagggca tctacaagtc	1200
ccatttgatc cacaaacgct acaccgtgcc cagcaccact ccacgcgtgt ggggctcctg	1260
gggccgaggg tccgccctcg agaaccacaa gctcctcccc ctatgtttcc cgctcccccg	1320
gagtcagaa gccccgcccc tggctggaac ttcacgccct ccggacggat tgccccattt	1380
tctccatttt cccgcttctc ccagtcaagt tctgaacttg tgaggcatct gggcctcccc	1440
agaagacatt taacacagaa agcacagccc tactaactag tattcttacc tgtctcttca	1500
agaatttcag accaatcgac cgtcctgtct ctttaaggct taggaagagc agtgtggctg	1560
cccctttaag gaggcgttgc aacaaaccat attggacaga cgatgggggc gacccatcgg	1620
gacccgacgg gcctctgact ccagcaatac agcgaatcag cggctttcgg gaatacattt	1680
ttcggaaaaa gacttcttcc tcggttttct gctctgcaca cgttgaaatt ttccccagtt	1740
tttctgcag atcgggagtc gagcaatgcc taccctcgcg ctcccgacc agttgggcgc	1800
tcccgatga tgccctaccc ctttgatcc acgtggtctg caacctggtg cgagcagccc	1860
gggctacagg gttgcctgag gtgtgggtcc caggatggag gagccccagg ccggcggtga	1920
gggtgcgggt tgacggggtg cggaggggtg gttggtggaa ggagaaaggg gcgtccgaga	1980
gggttcgggc ggaaaaggag gcgtacctgc aagcaggact tgcgaagagc gtgcattccc	2040
agtgggcgaa cgggaattcg aacggagaga gggttatctt gtgggggggt acccggtggag	2100
agcaaggcgc ccccgagggt tggatcgggt aaattgaggt cggccctggg gaacagggtg	2160
gcagaaagga gaaaccagggt tgaggggact ggagtgtca cgagggttaag accaatggac	2220
cgataggcgc gccctgcaag attggaccgg caaggagggtg tcagtcgacc ccatttcccc	2280
ttctgtgca gatgctgctc ggttctcttg tcccccaac tttaccgcga agccccagc	2340
ctcagagtcc cctcgtttct ccttggaggc gctgacgggt ccagatacgg agctgtggct	2400
tattcaggcc cctgcagact ttgccccaga atggtgagtg gtcttggtga cggaaaagag	2460
ggtcccgtc cagaccccaa gagcgggttc ttgaatttgt cacaggaaag aattagagggt	2520
gagtcacaga gcacagtga agaaacaagt ttattggaaa ctactccttt acagagtaga	2580
gtgtcctcag aaagcagggg gagaaacca cagccctttg ttagtatttc tacttataag	2640
aaactataag gaactatagt taaacttga gtgtgcagat aagctcacta aaggtagggg	2700
ctattggtgt tatccacgac cattaatcct gcaacctaa cttgctcatt tatgttatat	2760
ttaagtaatg ggggctgcat tcttaggaca tttggacatt ctgcaggctt ggtggaacat	2820
gttctgtatg gccataaata ttctgtaatt ataattggtg gtcagcctgg gatgtgggtta	2880
ttttcaggcc ataagcatga accttgtaag tgcctagcta ctactttaa gatggagtca	2940
ctctagtcat gttttattaa aaaccagagg ccagccaggc gcagtggctg gtgcctgtaa	3000
tcccatcctt tgggaggccg aggcgagcag atcacttgag gtcaggagtt caagaccagc	3060
ctggccaaca tagtgaaatt gtctctacta aaaatacaaa aattggctgg gcgtgggtggc	3120

45

aggtgcctgt aatcccagct acttgagagg ctgaggcagg agaatcgctt gaaccagga	3180
ggtggacatt gcagttagcc gagatcatgc cactgcactc cagcctaggc aacagagcaa	3240
gactctctca aaaaaaaaca aaaaaaaaaat caaaaaacct tccctctcct gttccactta	3300
agcctctgcc ctccctgttt ctctctgtag cttcaatggg cggcatgtgc ctctctctgg	3360
ctcccagatc gtcaagggca aattggcagg caagcggcac cgctatcgag tcctcagcag	3420
ctgtcccca gctggagaag cgaccctgct ggccccctca acggaggcag gaggtggact	3480
cacctgtgcc tcagccccc agggcacctt aaggatcctt gaggggtccc agcaatccct	3540
gtcagggagc cctctgcagc ccatcccagc aagtccccca ccacagatcc ctctggcct	3600
gaggcctcgg ttctgtgcct ttgggggcaa cccaccagtc acagggccta ggtcagcctt	3660
ggccccaac ctgctcacct caggaagaa gaaaaaggag atgcaggtga cagaggcccc	3720
agtcactcag gaggcagtga atgggcacgg ggccctggag gtggacatgg ctttggggtc	3780
gccagaaatg gatgtgcgga agaagaagaa gaaaaaaaaat cagcagctga aagaaccaga	3840
ggcagcaggg cctgtgggga cagagcccac agtggagaca ctggagcctc tgggagtgt	3900
gttcccgctc accaccaaga agaggaagaa gcccaaaggg aaagaaacct tcgagccaga	3960
agacaagaca gtgaagcagg aacagattaa cactgagcct ctagaagaca cagtcctgtc	4020
cccgaccaa aagagaaaga ggcaaaaggg gacggaaggg atggagccag aggagggggg	4080
gacagttgag tctcagccac aggtgaaggt ggagccactg gaggaagcca tccctctgcc	4140
ccctacgaag aagaggaaaa aagaaaaggg acagatggca atgatggagc cagggacgga	4200
ggcgatggag ccagtggagc cgagatgaa gcctctggag tccccagggg ggaccatggc	4260
gcctcaacag ccagaaggag cgaagcctca ggcccaggca gctctggcag ctcccaaaaa	4320
gaagacgaag aaagaaaaac agcaagatgc cacagtggag ccagag	4366

<210> 134
 <211> 20
 <212> DNA
 <213> Artificial

<220>
 <223> Primer

<400> 134
 aaaaaaatag ccgagcatgg 20

<210> 135
 <211> 22
 <212> DNA
 <213> Artificial

<220>
 <223> Primer

<400> 135
 tttgactgg gtaagaattt cc 22

<210> 136

46

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

<220>
<223> Primer

<400> 136
accatggcgc ctcaaca

17

<210> 137
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 137
gaattggctc agtcactgtg tga

23

<210> 138
<211> 27
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 138
cctggacaac atagggagac cctgtgt

27

<210> 139
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 139
caaacaaca aaaacctctg cca

23

<210> 140
<211> 17
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 140
accatggcgc ctcaaca

17

<210> 141
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 141

47

gaattggctc agtcactgtg tga

23

<210> 142

<211> 38

<212> DNA

<213> Homo sapiens

<400> 142

agaaagaaaa acagcaarat gccacagtgg agccagag

38