



(11) **EP 1 780 217 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
**02.05.2007 Bulletin 2007/18**

(51) Int Cl.:  
**C07K 14/47<sup>(2006.01)</sup> C12N 15/12<sup>(2006.01)</sup>**  
**C12Q 1/68<sup>(2006.01)</sup>**

(21) Application number: **06026286.2**

(22) Date of filing: **31.10.2000**

(84) Designated Contracting States:  
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE**

(30) Priority: **02.11.1999 EP 99121686**  
**31.05.2000 EP 00111696**

(62) Document number(s) of the earlier application(s) in  
accordance with Art. 76 EPC:  
**00972884.1 / 1 226 169**

(71) Applicant: **DRK-Blutspendedienst Baden-  
Württemberg-Hessen  
gemeinnützige GmbH  
68167 Mannheim (DE)**

(72) Inventors:  
• **Flegel, Willy A.  
64807 Dieburg (DE)**  
• **Wagner, Franz F.  
31832 Springe (DE)**

(74) Representative: **Vossius & Partner  
Siebertstrasse 4  
81675 München (DE)**

Remarks:

This application was filed on 19 - 12 - 2006 as a  
divisional application to the application mentioned  
under INID code 62.

(54) **Molecular structure of RHD negative locus**

(57) The present invention relates to a nucleic acid molecular structure representing the Rhesus genes locus comprising the *RHD*, *SMP1* and *RHCE* genes and/or the *Rhesus box(es)*, preferably the hybrid *Rhesus box*, the upstream *Rhesus box* and/or the downstream *Rhesus box*. Furthermore, the invention relates to a process for the specific detection of the common *RHD* negative haplotypes. The invention further relates to the detection of *RHD* positive haplotypes in D-negative individuals.

Various mutations in the *RHD* gene have been identified that allow for the development of diagnostic tools. The invention also relates to oligonucleotides, that specifically hybridize to the hybrid box, preferably the breakpoint or breakpoint region or to the upstream and downstream *Rhesus boxes*. Additionally, the invention relates to kits comprising or employing the above recited compounds of the invention.

**EP 1 780 217 A1**

## Description

**[0001]** The present invention relates to a nucleic acid molecular structure representing the Rhesus genes locus comprising the *RHD*, *SMP1* and *RHCE* genes and/or the *Rhesus box(es)*, preferably the hybrid *Rhesus box*, the upstream *Rhesus box* and/or the downstream *Rhesus box*. Furthermore, the invention relates to a process for the specific detection of the common *RHD* negative haplotypes. The invention further relates to the detection of *RHD* positive haplotypes in D-negative individuals. Various mutations in the *RHD* gene have been identified that allow for the development of diagnostic tools. The invention also relates to oligonucleotides, that specifically hybridize to the hybrid box, preferably the breakpoint or breakpoint region or to the upstream and downstream *Rhesus boxes*. Additionally, the invention relates to kits comprising or employing the above recited compounds of the invention.

**[0002]** Several documents are cited throughout the text of this specification; the disclosure content of each of the documents cited herein (including any manufacturer's specifications, instructions, etc.) is hereby incorporated by reference.

**[0003]** The Rhesus D antigen (ISBT 004.001; RH1) is the most important blood group antigen determined by a protein. Ariti-D remains the leading cause of hemolytic disease of the newborn (Filbey, *Acta Obstet Gynecol Scand*, 74:687, 1995; Bowman, J, *Semin Perinatol* 21:39, 1997). Depending on the population, 3% to 25% of whites lack the antigen D (Mourant, *The distribution of the human blood groups and other polymorphisms*, London, Oxford University Press, 1976). Anti-D immunizations can occur readily in D-negative recipients (Urbaniak, *Transfusion* 21:64, 1981).

**[0004]** The antigens of the RH blood group are carried by proteins coded by two genes, *RHD* and *RHCE*, that are located at chromosomal position 1p34.1 - 1p36 (Cherif-Zahar, *Hum. Genet.* 86: 398, 1991; MacGeoch, *Cytogenet. Cell Genet.* 59:261, 1992) probably within less than a 450,000 base pair (bp) distance (Carritt, *Hum. Mol. Genet.* 6:843, 1997). Both genes encompass ten exons and their structures are highly homologous. The relative orientation of the genes, their distance, and the possibility of interspersed other genes were unknown (Flegel, *Transfus. Med.* 8:281, 1998). Very recently, Okuda et al. (Okuda, *Biochem. Biophys. Res. Commun.* 263:378, 1999) reported a sequence of about 11,000 bp, which was thought to represent the DNA segment between *RHD* and *RHCE*.

**[0005]** In whites, the vast majority of D-negative haplotypes is due to a deletion of the *RHD* gene: This deletion spans the whole *RHD* gene, because *RHD*-specific sequences ranging from exon 1 to the 3' untranslated region are absent (Gassner, *Transfusion* 37:1020, 1997). The exact extent of the deletion was uncertain, leaving open the possibility that neighboring genes were also affected.

**[0006]** The identification of the *RHD* gene as the molecular basis of the D antigen allowed RhD phenotype prediction by DNA typing (Flegel, *Transfus. Med.* 8:281, 1998; Lo, *Lancet* 341:1147, 1993). However, since the structure of the prevalent D-negative haplotype is unknown, a specific detection of the *RHD* deletion remained impossible and the discrimination of *RHD*<sup>+</sup>/*RHD*<sup>+</sup> homozygous from *RHD*<sup>+</sup>/*RHD*<sup>-</sup> heterozygous individuals relied on indirect methods. This discrimination is of clinical interest in particular, because in D-negative mothers with an anti-D, the risk of an affected child is 100% with a *RHD*<sup>+</sup>/*RHD*<sup>+</sup> father, but only 50% with a *RHD*<sup>+</sup>/*RHD*<sup>-</sup> father.

**[0007]** Several indirect approaches have been applied to determine the zygosity: (i) a simple guess based on the phenotype is correct in about 95% of cases, (ii) determination of the D antigen density which can be confounded by factors such as the presence of the C antigen, and (iii) several methods involving the parallel quantitative amplification of *RHD*- and *RHCE*-specific sequences (Cossu, *Electrophoresis* 17:1911, 1996; Döscher, *Infusionsther. Transfusionsmed.* 26(suppl 1):31, 1999 (abstr.)). These elaborate techniques may not be practical in routine laboratories. In addition, several investigators identified polymorphisms in the *RHCE* gene or neighboring sequences genetically linked to the lack of the *RHD* gene (Carritt, *Hum. Mol. Genet.* 6:843, 1997; Huang, *Am. J. Hum. genet.* 58: 133, 1996; Fujiwara, *Hum. genet.* 104:301, 1999; Onda, *Gene* 159:225, 1995). This indirect approach relied on the linkage disequilibrium associating the *RHD* deletion with a polymorphism.

**[0008]** Furthermore, the utility of the *RHD* PCR is limited by the incomplete knowledge of presumably rare *RHD* positive alleles in RhD-negative. *RHD* positive alleles in RhD negative are caused by *RHD-CE-D* hybrid genes (Huang, *Blood* 88:2326-33, 1996; Faas, *Transfusion* 37:38-44, 1997; Faas, *Transfusion* 36:506-11, 1996), nonsense-mutations (Avent, *Blood* 89:2568-77, 1997), frameshifts (Andrews, *Blood* 92:1839-40, 1998; Cherif-Zahar *Br. J. Haematol.* 102:1263-70, 1998), or pseudogenes (Singleton, *Blood* 95:12-8, 2000). Such alleles are frequent in Africans (Faas, *Transfusion* 37: 38-44, 1997; Singleton, *Blood* 95:12-18, 2000) and Asians (Okuda, *J. Clin. Invest.* 100:373-9, 1997) but rare in whites. Nevertheless, recent analyses (Avent, *Blood* 89:2568-77, 1997; Flegel, *Transfus. Med.* 8:281-302, 1998) suggested that even for whites these alleles are likely the leading cause of incorrect Rh phenotype prediction. Several observations in whites (Avent, *Blood* 89:2568-77, 1997; Hyland, *Blood* 84:321-4, 1994) indicated that these alleles clustered in the Cde and cdE haplotypes.

**[0009]** The most direct approach for analyzing the *RHD* locus on the molecular level would be PCR amplification spanning the *RHD* deletion site. Such an assay has, so far, not been available because the structure of the *RHD* locus in RhD positives and RhD negatives was incompletely understood.

**[0010]** Accordingly, the technical problem underlying the present invention was to provide means and methods for a

reliable, nucleic acid based analysis of the Rhesus D locus. These means and methods should be, inter alia, suitable for the detection and/or discrimination of *RHD*<sup>+</sup>/*RHD*<sup>+</sup> and *RHD*<sup>+</sup>/*RHD*<sup>-</sup> individuals.

[0011] The solution to said technical problem is achieved by providing the embodiments characterized in the claims.

[0012] Thus, the invention relates to a nucleic acid molecular structure representing the *Rhesus* genes locus comprising the *RHD*, *SMP1*, and *RHCE* genes and/or the *Rhesus boxes*, preferably the hybrid *Rhesus box*, the upstream *Rhesus box* and/or the downstream *Rhesus box*, the sequences of which are shown in Figures 8 to 10.

[0013] In the context of the present invention, the term "nucleic acid molecular structure" is defined as a linear DNA-segment that comprises, in its broadest meaning, the combination of the above mentioned genes, namely the *RHD*, *SMP1* and *RHCE* genes, arranged in this 5' to 3' order and/or *Rhesus boxes* that co-determine said *Rhesus* gene locus. DNA sequences that give rise to the molecular structure of the invention include the following: The nucleotide sequence structure consists of a *Rhesus box* 5' flanking region, the hybrid *Rhesus box* or two *Rhesus boxes* with intervening *RHD* gene, and the *Rhesus box* 3' flanking region.

[0014] The following sequences represent preferred embodiments contained in the nucleic acid molecular structure of the invention.

[0015] The *Rhesus box* 5' flanking region is represented in the genomic clone HS465N24 (GenBank accession number AL031432.1), bases 1 to 120,156.

[0016] The hybrid *Rhesus box* is represented in GenBank accession number AL252313 bases 33 to 9,180.

[0017] The two *Rhesus boxes* with intervening *RHD* gene consists of the upstream *Rhesus box*, represented in GenBank accession number AL252311 bases 34 to 9,175, the *RHD* gene and the downstream *Rhesus box* represented in GenBank accession number AL252312 bases 23 to 9,177 (see Fig. 8 to 10).

[0018] The *Rhesus box* 3' flanking region consists of a small DNA segment between the downstream or hybrid *Rhesus box* and the *SMP1* gene, the *SMP1* gene and the *RHCE* gene.

[0019] The *RHD* gene consists of a *RHD* 5' region homologous to genomic clone HS469D22 (GenBank accession number AL031284.9) bases 56,012 to 51,472; also represented by a nucleotide segment dubbed "stuffer fragment" (GenBank accession number AB029152) bases 7,716 to 11,005; the *RHD* promoter (GenBank accession number AJ252314) bases 1 to 1,246 (see Figure 11) and the *RHD* gene defined by the *RHD* cDNA (GenBank accession number X63097) bases 1 to 1,371 and by its intron sequences.

[0020] The small DNA segment preferably comprises 15 nucleotides between the downstream or hybrid *Rhesus box* and the *SMP1* gene and is represented in AL252312 by bases 9,178 to 9,192.

[0021] The *SMP1* gene is defined by the *SMP1* cDNA represented in GenBank accession number AF0811282 and by its intron sequences.

[0022] The *RHCE* gene is defined by the *RHCE* cDNA represented in GenBank accession number X63095 and by its intron sequences and further represented in part by the genomic clone HS469D22 (GenBank accession number AL031432.1) bases 1 to 51,471 and the *RHCE* 5' flanking region represented by genomic clone HS469D22 bases 51,472 to 84,811.

[0023] Whereas the upstream *Rhesus box* is located 5' of the *RHD* gene, the downstream *Rhesus box* is located between the *RHD* and *SMP1* genes in this structure of the present invention. Alternatively, the term "nucleic acid molecular structure" relates to DNA segments solely comprising the referenced *Rhesus boxes*. This term, in a further alternative, relates to DNA segments comprising the *RHD*, *SMP1* and *RHCE* genes and two *Rhesus boxes*, namely the upstream *Rhesus box* and the downstream *RHD* box. Comprised by this term are also, in a further alternative, DNA segments that comprise the hybrid *Rhesus box*, the *SMP1* gene and the *RHCE* gene. In another alternative, the term relates to DNA segments comprising the *SMP1* gene and the hybrid *Rhesus box*. This term in a further alternative relates to DNA segments comprising the upstream *Rhesus box*, *RHD*, downstream *Rhesus box* and *SMP1*.

[0024] This term in another alternative relates to DNA segments comprising the downstream *Rhesus box* and *SMP1*. For a better understanding of the claimed subject-matter, it is referred to figures 1 and 7, *infra*.

[0025] In accordance with the present invention, the term "nucleic acid molecular structure" comprises also any feasible derivative of the above referenced nucleic acid structure to which a nucleic acid probe may hybridize. In other words, the structure of the invention may be prepared by synthetic or semisynthetic means and thus consist of or comprise peptide nucleic acid. Said term also bears the meaning of a nucleic acid molecule.

[0026] In accordance with the present invention, the term "*Rhesus box*" describes upstream and downstream DNA segments that flank the *RHD* gene on the 5' and 3' end. The three *Rhesus boxes* are defined by their nucleotide sequences. The hybrid *Rhesus box* is represented in one embodiment in GenBank accession number AL252313 bases 33 to 9,180. The two *Rhesus boxes* with intervening *RHD* gene consists of the upstream *Rhesus box*, represented in one embodiment in GenBank accession number AL252311 bases 34 to 9,175 and the downstream *Rhesus box* represented in one embodiment in GenBank accession number AL252312 bases 23 to 9,177. As exemplified in the appended examples the *Rhesus boxes* are preferably approximately 9000bp long, having 98,6% identity and identical orientation. According to the present invention the upstream and downstream *Rhesus boxes* are at least 95% homologous. The length of these *Rhesus boxes* may vary. It is expected that the length of these *Rhesus boxes* may vary, because, among

other structural features, multiple repetitive elements, some of them are organized in tandem arrays, are known to be prone to (array) elongation and deletion events. If such events occur the length of the *Rhesus boxes* may shrink to less than 1,000 nucleotides length or extend to more than 20,000 nucleotides length.

**[0027]** In accordance with the present invention the term "identity" refers to the determination of sequence identity using suitable alignment programs, such as BLAST.

**[0028]** As has been pointed out above, the diagnostic analysis of *RHD* negatives on the molecular level has so far been hampered by the fact, that the overall structure of the *RHD/RHCE* loci was unknown. It has now been surprisingly found, that the two genes, *RHD* and *RHCE*, have opposite orientation and face each other with their 3' ends. In accordance with the present invention it has further been found that the *RHD* gene is surrounded by two highly homologous *Rhesus boxes*. The physical distance between *RHD* and *RHCE* is about 30,000 bp and is filled with a *Rhesus box* and the *SMP1* gene. The breakpoints of the *RHD* deletion in the prevalent *RHD* negative haplotypes are located in the 1,463 bp identity region of the *Rhesus boxes*. Similar *RHD* deletion events may involve any other region within the highly homologous *Rhesus boxes*. Hence, a region of a breakpoint comprising an *RHD* deletion other than the common *RHD* deletion may be anticipated to occur anywhere within the *Rhesus boxes* as defined above.

**[0029]** The opposite orientation of the two *RH* genes explains the different character of hybrid genes in the MNS and RH blood group: The glycoprotein genes encoding the MNS antigens occur in the same orientation (Onda, Gene 159: 225, 1995), and many recombinations may be explained as unequal crossing over resulting in single hybrid genes (Blumenfeld, Hum. Mutat. 6:1999, 1995). Based on the surprising findings referred to above, the events on the molecular level that lead to *RHD* negatives can now be more fully understood. In the *RH* locus, the inversely oriented sequences are unlikely to trigger unequal crossing over, and if this event occurred, no functional hybrid gene would result. The conclusion that unequal crossing over at the *RH* gene locus is unlikely may explain that most *RH* hybrid genes are of *RHD-CE-D* or *RHCE-D-CE* type and involve stretches of homologous DNA positioned *in cis* as noted previously (Wagner, Blood 91:2157, 1998). Currently, the *RH* gene system is the only well investigated gene locus where the two genes have opposite orientation, rendering it a model system for the evolution of neighboring, oppositely oriented genes that are frequent throughout genomes.

**[0030]** Based on the structure of the *RH* gene locus (Fig. 1), a parsimonious model for the *RHD* gene deletion event is proposed (Fig. 7). Although the applicant does not wish to be bound to theory, the following is believed with regard to the generation of RhD negative. The *RHD* deletion may be explained by unequal crossing over triggered by the highly homologous *Rhesus boxes* embracing the *RHD* gene. The hybrid-type *Rhesus box* of *RHD*-negatives arises, when a crossover leading to a deletion event involving a breakpoint region within the identity region of the upstream and downstream *Rhesus boxes* takes place. Thus, the hybrid *RHD* box is characterized by a 5' portion derived from the upstream *RHD* box fused to a 3' portion from the downstream *RHD* box. In one preferred embodiment the breakpoint region is 903bp long. The sequence of this preferred hybrid *Rhesus box* is depicted in figure 5. In the specific embodiments described in the examples, said 903 bp breakpoint region in the *Rhesus boxes* is located in a 1,463 bp stretch of 99.9% homology resembling a THE-1B human transposable element and a L2 repetitive DNA element (Fig. 4). Interestingly, the >60,000 bp DNA segment that is deleted in the *RHD* negative haplotype consisted only of and contained all sequences that are duplicated in the *RHD* positive haplotype.

**[0031]** The findings of the present invention referred to herein above allow for the establishment of a number of easy to do or refined methods for the analysis of the genotype of an individual with regard to the *RH* gene locus. Examples of such methods are provided herein below.

**[0032]** While the molecular mechanism resulting in the prevalent *RHD* negative haplotype is now apparent, it is less clear how the much older duplication event gave rise to the structure of the *RH* genes in *RHD* positives. The duplication of the *Rhesus box* and the *RH* genes probably occurred as a single event, because the overall homology of the two *Rhesus boxes* is very similar to that of the *RH* genes. Without being bound by theory, it is tempting to speculate that the *RHD* duplication originate in causal connection with the insertion of the near full-length THE-1B transposon-like human element in duplicate. However, the open reading frame of the THE-1 B element probably was non-functional at the time of the duplication.

**[0033]** In a preferred embodiment of the present invention, said nucleic acid molecular structure is representative of the common *RHD* negative haplotypes.

**[0034]** According to the present invention, the term "is representative of" relates to a nucleic acid molecular structure comprising all sequential and structural features to relate said structure to a group of molecular structures sharing said features. In the above preferred embodiment, said features give rise to the common *RHD* negative haplotype. In the present context this means preferably the deletion of the *RHD* gene encompassing the whole *RHD* gene and its 5' region, which are located between the upstream *Rhesus box* and the downstream *Rhesus box*.

**[0035]** In the present context this could also mean, for example, that all structures sharing a nonsense mutation, missense mutation, splice site mutation, partial deletion, partial insertion, partial inversion or a combination thereof within the *RHD* gene, which terminates or obliterates the expression of a protein product of the *RHD* gene, are representative of the *RHD* negative haplotype.

**[0036]** The term "haplotype" relates to a series of linked alleles within a defined region on a single maternal or paternal chromosome.

**[0037]** The term "common *RHD* negative haplotype" refers to any RhD antigen negative haplotype that comprises a hybrid *Rhesus box*. Preferably the DNA segment encompassing the whole *RHD* gene and its 5' region, which are located between the upstream *Rhesus box* and the downstream *Rhesus box*, is deleted.

**[0038]** In another embodiment, the invention relates to a nucleic acid molecular structure, dubbed *Rhesus box*, which is flanking the breakpoint region of the *RHD* deletion in the common *RHD* negative haplotypes.

**[0039]** In accordance with the present invention the term "breakpoint region of the *RHD* deletion" describes a distinct DNA segment that is involved in an *RHD* deletion. As has been pointed out above, said deletion may be the result of an unequal crossing over event involving both the upstream and downstream *Rhesus boxes*, deleting interspersed sequences and finally giving rise to a nucleic acid molecular structure (the referenced *Rhesus boxes* for a better delimitation from the upstream and downstream *Rhesus box* also referred to as hybrid *Rhesus box*) wherein the 5' portion of the upstream *RHD* box is in close spatial proximity to the 3' portion of the downstream *RHD* box. As mentioned above and depicted in figures 7 and 4 this region can preferably be 903bp long and be located in a 1,463bp stretch within the *Rhesus boxes*, having 99,9 % homology in this segment. In another preferred alternative said region is located downstream from said 903 bp fragment but is still contained within the 1463 bp stretch. Preferably, said fragment is 556 to 560bp long. The actual breakpoint may vary such that the contribution of the upstream *Rhesus box* and the downstream *Rhesus box* are different in different individuals. However, in accordance with the present invention, the breakpoint in any case occurs in the upstream and downstream *Rhesus boxes*.

**[0040]** The hybrid *Rhesus box* is particularly useful for the analysis of RHD-negative haplotypes. For example, oligonucleotides may be employed that hybridize to nucleic acid sequences comprising the breakpoint which arose as a result of the *RHD* deletion. It is to be understood that such oligonucleotides need to hybridize to a significant portion preferably encompassing 20 nucleotides, that is located 5' and 3' of the region of the actual breakpoint in order to be indicative of a deletion event. For example, when such an oligonucleotide is hybridized under stringent conditions such as 0.2 x SSC, 0.1 SDS at 65°C and the probe would be 943 nucleotides long, then the hybridizing region should include portions that hybridize 3' as well as 5' of the breakpoint.

**[0041]** For example, a *Rhesus box* or a part thereof encompassing the region of the breakpoint is amplified. Thereafter the amplification product is assayed in a sequence specific way by hybridization to an oligonucleotide of about six or more nucleotides length.

**[0042]** Preferably, a stretch of DNA representative of a *Rhesus box* or part thereof encompassing the region of the breakpoint is amplified using two primers. One primer may be located in the *Rhesus box* 5' of the identity region and is specific for both the upstream *Rhesus box* and the hybrid *Rhesus box*. The other primer may be located in the *Rhesus box* 3' of the identity region and is specific for both the downstream *Rhesus box* and the hybrid *Rhesus box*. In this application, the presence of an amplification product of the expected size is indicative of the presence of a hybrid *Rhesus box* and hence, of the *RHD* deletion.

**[0043]** Another possible combination of primers is the following: One primer may be located in the *Rhesus box* 5' of the identity region and is specific for both the upstream *Rhesus box* and the hybrid *Rhesus box*. The other primer may be located in the *Rhesus box* 3' of the identity region. In this application, the presence of a hybrid *Rhesus box* is determined by examining the specificity of the parts of the amplification product pertaining to a DNA stretch of the *Rhesus box* 3' of the identity region. This may for example be effected by hybridization with an oligonucleotide that hybridizes to the hybrid *Rhesus box* and to the downstream *Rhesus box* but not to the upstream *Rhesus box*, or by digestion with a restriction enzyme that cuts the hybrid *Rhesus box* and the downstream *Rhesus box* but does not cut the upstream *Rhesus box*, or by digestion with a restriction enzyme that does not cut the hybrid *Rhesus box* and the downstream *Rhesus box* but cuts the upstream *Rhesus box*, or by nucleotide sequencing.

**[0044]** Another possible combination of primers is the following: One primer may be located in the *Rhesus box* 5' of the identity region. The other primer may be located in the *Rhesus box* 3' of the identity region and is specific for both the downstream *Rhesus box* and the hybrid *Rhesus box*. In this application, the presence of a hybrid *Rhesus box* is determined by examining the specificity of the parts of the amplification product pertaining to a DNA stretch of the *Rhesus box* 5' of the identity region. This may for example be effected by hybridization with a nucleotide that hybridizes to the hybrid *Rhesus box* and to the upstream *Rhesus box* but not to the downstream *Rhesus box*, or by digestion with a restriction enzyme that cuts the hybrid *Rhesus box* and the upstream *Rhesus box* but does not cut the downstream *Rhesus box*, or by digestion with a restriction enzyme that does not cut the hybrid *Rhesus box* and the upstream *Rhesus box* but cuts the downstream *Rhesus box*, or by nucleotide sequencing.

**[0045]** The hybrid *Rhesus box* may also serve as a diagnostic tool for the presence of the *RHD* deletion when analyzed by an anti-DNA antibody specific for one or more embodiments of the hybrid box, a fragment or derivative thereof such as an scFvFab or F(ab')<sub>2</sub> fragment or an aptamer etc. Thus, antibodies, fragments or derivatives thereof or such aptamers can be generated by the person skilled in the art according to conventional technology (see, for example, Harlow and Lane, "Antibodies, A Laboratory Manual", CSH Press 1988, Cold Spring Harbor).

**[0046]** In a preferred embodiment, the invention relates to a nucleic acid molecular structure representative of an *RHD* negative haplotype comprising an *RHD* gene deletion involving the upstream *Rhesus box*, the downstream *Rhesus box*, or both.

**[0047]** The invention further relates to a nucleic acid molecular structure that is flanking the *Rhesus box* in the common *RHD* negative haplotypes. These structures or sequences can be used to derive primers for amplification reactions such as long range PCR for the molecular analysis of the *RHD* locus.

**[0048]** For example, a stretch of DNA representative of a *Rhesus box* and parts of their flanking regions or parts thereof encompassing the region of the breakpoint is amplified using two primers. One primer may be located in the 5' flanking region of the *Rhesus box*. Alternatively, this primer may be located in the *Rhesus box* 5' of the identity region and is specific for both the upstream *Rhesus box* and the hybrid *Rhesus box*. The other primer may be located in the *Rhesus box* 3' flanking region. Alternatively, this primer may be located in the *Rhesus box* 3' of the identity region and is specific for both the downstream *Rhesus box* and the hybrid *Rhesus box*. In this application the presence of an amplification product of the expected size is indicative of the presence of a hybrid *Rhesus box* and hence, of the *RHD* deletion.

**[0049]** Another possible combination of primers is the following: One primer may be located in the 5' flanking region of the *Rhesus box*. The other primer may be located in the *Rhesus box* 3' of the identity region. In this application, the presence of a hybrid *Rhesus box* is determined by examining the specificity of the parts of the amplification product pertaining to a DNA stretch of the *Rhesus box* 3' of the identity region. This may for example be effected by hybridization with an oligonucleotide that hybridizes to the hybrid *Rhesus box* and to the downstream *Rhesus box* but not to the upstream *Rhesus box*, or by digestion with an restriction enzyme that cuts the hybrid *Rhesus box* and the downstream *Rhesus box* but does not cut the upstream *Rhesus box*, or by digestion with a restriction enzyme that does not cut the hybrid *Rhesus box* and the downstream *Rhesus box* but cuts the upstream *Rhesus box*, or by nucleotide sequencing.

**[0050]** Another possible combination of primers is the following: One primer may be located in the *Rhesus box* 5' of the identity region. The other primer may be located in the 3' flanking region of the *Rhesus box*. In this application, the presence of a hybrid *Rhesus box* is determined by examining the specificity of the parts of the amplification product pertaining to a DNA stretch of the *Rhesus box* 5' of the identity region. This may for example be effected by hybridization with an oligonucleotide that hybridizes to the hybrid *Rhesus box* and to the upstream *Rhesus box* but not to the downstream *Rhesus box*, or by digestion with an restriction enzyme that cuts the hybrid *Rhesus box* and the upstream *Rhesus box* but does not cut the downstream *Rhesus box*, or by digestion with a restriction enzyme that does not cut the hybrid *Rhesus box* and the upstream *Rhesus box* but cuts the downstream *Rhesus box*, or by nucleotide sequencing.

**[0051]** In a preferred embodiment the nucleic acid molecular structure is representative of *RHD* positive haplotypes.

**[0052]** The term "*RHD* positive haplotype" refers to any haplotype that comprises DNA sequences specific for the *RHD* gene.

**[0053]** In a preferred embodiment the invention relates to a nucleic acid molecular structure representative of the common *RHD* positive haplotype.

**[0054]** In another preferred embodiment the nucleic acid molecular structure is derived from a sample comprising an *RHD* positive haplotype that is serologically classified RhD negative.

**[0055]** In the context of the invention, the term "serologically classified RhD negative" describes a sample that has been tested for the presence of RhD antigen using, e.g., routine serological assays wherein the result of such assays was negative.

**[0056]** In a particularly preferred embodiment the sample that is classified RhD negative is obtained from a Caucasian population.

**[0057]** In another more preferred embodiment the nucleic acid molecular structure comprises a partial *RHD* deletion.

**[0058]** One alternative for explaining that an allele routinely diagnosed *RHD* positive gives rise to a RhD-negative phenotype is the deletion of a part of the *RHD* gene or the substitution of a part of the *RHD* gene by the corresponding DNA segments derived from the *RHCE* gene not detected by standard diagnostic methods such as PCR.

**[0059]** In a preferred embodiment said deletions or substitutions comprise *RHD* exons 3 to 7 or 4 to 7 giving rise to a CcddEe phenotype, or 1 to 9.

**[0060]** In a most preferred embodiment said substitution comprises a *RHD-CE(3-7)-D* hybrid gene, a *RHD-CE(4-7)-D* hybrid gene giving rise to a CcddEe phenotype or a *RHCE(1-9)-D(10)* hybrid gene, all of which correlate with a RhD negative phenotype.

**[0061]** In another most preferred embodiment the nucleic acid molecular structure of the present invention comprising an *RHD-CE-D* hybrid allele, which is representative of a *Cde<sup>s</sup>* haplotype but also occurs in other Rhesus haplotypes, carrying a 5' breakpoint region located in intron 3, the sequence of which breakpoint region is shown in Figure 12, and/or a 5' breakpoint region located in intron 7, the sequence of which breakpoint region is shown in Figure 13, or both breakpoint regions.

**[0062]** *Cde<sup>s</sup>*, also known as r<sup>s</sup>, is a *RH* haplotype resembling *Cde* that was initially characterized as expressing antigen e<sup>s</sup> instead of antigen e, expressing antigen c, and expressing reduced and altered antigen C (Issitt, P.D. Applied Blood Group Serology, Miami: Montgomery Scientific Publications, 1985, page 239). The molecular structure underlying this

haplotype has recently been elucidated (Blunt, T., Daniels, G., and Carritt, B. Serotype switching in a partially deleted RHD gene. *Vox Sang.* 67:397-401, 1994; Faas, B.H.W., Becker, E.A.M., Wildoer, P., Ligthart, P.C., Overbeeke, M.A.M., Zondervan, H.A., von dem Borne, A.E.G.K., and van der Schoot, C.E. Molecular background of VS and weak C expression in blacks. *Transfusion* 37:38-44, 1997; Daniels, G.L., Faas, B.H., Green, C.A., Smart, E., Maaskant-van Wijk, P.A., Avent, N.D., Zondervan, H.A., von dem Borne, A.E., and van der Schoot, C.E. The VS and V blood group polymorphisms in Africans: a serologic and molecular analysis. *Transfusion* 38:951-958, 1998.): The *Cde<sup>s</sup>* haplotype contains a *RHD-CE-D* hybrid gene encoding for an antigen C immunoreactivity, in which exons 4 to 7 derived from *RHCE* and exon 3 has a *RHD* like structure but possesses a *RHCE* specific Thr at codon 152.

**[0063]** Several additional *RHD* positive alleles occurring in RhD negative individuals have previously been partly or fully characterized (Table 10). Three of these ten published *RHD* alleles represented *RHD-CE-D* hybrid alleles in which the *RHCE* specific stretch encompassed at least exons 4 to 7. For each of these three hybrid *RHD* alleles, alleles were found whose patterns would be compatible (Table 10). Out of the seven RhD negative patterns observed in the present study, six were compatible with such type of hybrid *RHD* allele. Seven out of ten published *RHD* alleles represented deletions, nonsense mutations or a pseudogene. None of these alleles occurred in this study, which may indicate that they are rare in whites. In another embodiment, the invention provides for the detection or determination of the *RHD* allele previously described as *RHD* exon 9 negative (Gassner, *Transfusion* 37:1020ff, 1997), preferably representing an hybrid *RHD-CE(9)-D* hybrid allele, by its lack of *RHD* specific sequences in parts of intron 7 and intron 8. The specific steps carried out in this method may be any of the steps referred to in the further methods of the invention described in the specification, alone or in any combination.

**[0064]** In a particularly preferred embodiment the nucleic acid molecular structure of the present invention wherein the *RHD-CE-D* hybrid of the present invention encodes a polypeptide having antigen C reactivity.

**[0065]** Antigen C is a blood group antigen belonging to the RH blood group system known in the art and designated as 004.002 according to the ISBT nomenclature. A description is contained in many textbooks on immunohematology, e.g. Reid, M.E. and Lomas-Francis, C. *The Blood Group Antigen Facts Book*, San Diego: Academic Press, 1997.

**[0066]** Furthermore, in a preferred embodiment of the invention the nucleic acid molecular structure of the invention or a nucleic acid molecule being derived from the *RHD* gene comprises a single nucleotide substitution within the coding region of the *RHD* gene or within a 5' or 3' splice site.

**[0067]** The term "a nucleic acid molecule derived from the *RHD* gene" is intended to mean that this nucleic acid molecule originates from the *RHD* gene but carries a mutation, deletion, insertion, substitution or duplication within the coding region, any of the splice sites or a non-coding region. Preferably, said nucleic acid molecule gives rise to an aberrant polypeptide.

**[0068]** In a further preferred embodiment said nucleotide substitution gives rise to a stop-codon at codon 16.

**[0069]** In a more preferred embodiment said substitution gene gives rise to an *RHD(W16X)* mutation.

**[0070]** In an additional more preferred embodiment said substitution is a G→A substitution at nucleotide position 48.

**[0071]** In a further preferred embodiment of the invention said nucleotide substitution gives rise to a stop codon at codon 330.

**[0072]** In a more preferred embodiment of the invention said substitution gives rise to a *RHD(Y330X)* mutation.

**[0073]** In an even more preferred embodiment of the invention said substitution is a C → G substitution at nucleotide position 985.

**[0074]** In another preferred embodiment of the invention said substitution gives rise to a missense mutation at codon 212.

**[0075]** In another preferred embodiment of the invention said substitution gives rise to a *RHD(G212V)* missense mutation.

**[0076]** In a more preferred embodiment of the invention said substitution is a G→T substitution at nucleotide position 635.

**[0077]** In a different preferred embodiment of the invention said substitution gives rise to a mutation within a 4 nucleotide sequence, a 6-nucleotide sequence or an 8-nucleotide sequence comprising the consensus splice site at the exon 8/intron 8 boundary.

**[0078]** In another more preferred embodiment of the invention said substitution give rise to a *RHD(G1153(+1)A)* mutation.

**[0079]** In an additional more preferred embodiment of the invention said substitution is a substitution at the 5' splice site intron 8 from AGgt to AGat.

**[0080]** In a further more preferred embodiment the nucleic acid molecular structure of the invention or a nucleic acid molecule being derived from the *RHD* gene correlates with a RhD-negative phenotype.

**[0081]** In another preferred embodiment of the invention said substitution gives rise to a mutation within a 4-nucleotide, a 6-nucleotide sequence or an 8-nucleotide sequence sequence comprising the consensus splice site of the exon 3/intron3 boundary.

**[0082]** In a further preferred embodiment of the invention said substitution gives rise to a *RHD(G486(+1)A)* mutation.

**[0083]** In an additional more preferred embodiment of the invention said substitution is a substitution at the 5' splice

site intron 3 from ACgt to ACat.

**[0084]** In a further preferred embodiment of the invention said substitution gives rise to a mutation within a 4-nucleotide sequence, a 6-nucleotide sequence or an 8-nucleotide sequence comprising the consensus splice site of exon 9/intron 9 boundary.

**[0085]** In another preferred embodiment said substitution gives rise to a *RHD(K409K)* mutation.

**[0086]** In an additional more preferred embodiment of the invention said substitution is a substitution at the 5' splice site intron 9 from AGgt to AAgt.

**[0087]** In a more preferred embodiment of the invention the nucleic acid molecular structure of the invention or a nucleic acid molecule being derived from the *RHD* gene correlates with a  $D_{el}$ -phenotype.

**[0088]** In summary and referring to the above, *RHD* positive alleles can harbour single nucleotide substitutions leading to termination or reduction of the D-antigen expression. Using the improved detection methods disclosed in the present invention four *RHD* positive alleles in RhD negatives were found that had not been described previously. Two alleles, *RHD(W16X)* and *RHD(Y330X)* harbored stop codons preventing the expression of the full RhD protein. In three alleles, splice site mutations were found that may prevent correct splicing and RhD expression.

**[0089]** These alleles typed *RHD* positive in all *RHD* PCR methods tested, and a correct antigen D prediction necessitates a specific detection of these alleles or of polymorphisms linked to these alleles.

**[0090]** Previously, the discrimination of *RHD* homozygotes from *RHD* heterozygotes was difficult. The prevalent *RHD* negative allele could not be detected specifically (Flegel, *Transfus. Med.* 8:281, 1998; Cossu, *Electrophoresis* 17:1911, 1996). The above defined mutation found in accordance with the present invention provides the basis for the detection of the prevalent *RHD* negative haplotypes, and hence true *RHD* genotyping is now feasible.

**[0091]** The invention also relates to a process to specifically detect a *RHD* negative haplotype in a sample by utilizing any structural feature or nucleotide sequence or both of the above-described nucleic acid molecular structure or combinations thereof with techniques known in the art, preferably amplification reactions, such as polymerase chain reaction (PCR), more preferably by PCR-RFLP, PCR-SSP or long-range PCR.

**[0092]** The described PCR-RFLP and long-range PCR methods utilize either *Rhesus box* sequences or *Rhesus box* flanking sequences. By utilizing the same DNA stretches or combinations thereof, other methods, like PCR-SSO or biochips, can be developed or applied.

**[0093]** In one preferred embodiment of the present invention said process to specifically detect a common *RHD* negative haplotype comprises the following steps:

- (a) isolating the DNA from a blood sample or blood donor;
- (b) hybridizing at least two oppositely oriented primers under stringent conditions to the DNA so as to carry out a PCR;
- (c) amplifying the target sequence;
- (d) separating the amplification products on a gel; and
- (e) analyzing the amplicons.

**[0094]** Said sample may or may be derived from blood, serum, sputum, feces or other body fluid. The sample to be analyzed may be treated as to extract, inter alia, nucleic acids. The isolation of DNA from preferably EDTA- or citrate agglutinated blood samples can be carried out by a modified salting out methods, following the standard techniques as described in Gassner, *Transfusion* 37: 1020, 1997. The primers are preferably oligonucleotides that either occur naturally or in a purified restriction digest or are produced synthetically. The primers are preferably single stranded for a maximum of efficiency in the method of the present invention, and are preferably oligodeoxyribonucleotides. Purification of said primers is generally envisaged, prior to their use in the methods of the present invention, said purification comprising High Performance Liquid Chromatography (HPLC) or Polyacrylamide gel electrophoresis (PAGE), all technologies that are well known to the skilled artisan. Amplification methods such as PCR or LCR are well known in the art and described, for example in Flegel, *Transfusion Medicine* 8 (1998), 281-302; Maaskant, *Transfusion* 38 (1998), 1015-1021 and Legler, *Transfusion* (1996), 426-31.

**[0095]** According to the present invention a preferred method to detect the *RHD* deletion is performing PCR-RFLP using the expand high fidelity PCR-system and non-specific primers binding 5' of the end of the *Rhesus box* identity region as well as primers specific for the downstream *Rhesus box* and binding 3' of the end of the *Rhesus box* identity region. The PCR conditions involve preferably annealing at 65°C, extension for 10 min at 68°C. Thereafter, PCR amplicons are digested with PstI for 3h at 37°C and fragments resolved using 1% agarose gel. Additional preferred methods are further described in examples 10 and 11.

**[0096]** Another embodiment of the invention relates to a process to specifically detect a common *RHD* negative haplotype comprising the detection of the hybrid *Rhesus box*.

**[0097]** The detection of the hybrid *Rhesus box* provides the practitioner with an unambiguous result as regards the nature of the corresponding *RHD* allele. If the hybrid *Rhesus box* is detected, then the *RHD* gene is deleted. Detection of the hybrid *Rhesus box* is preferentially effected by using an oligonucleotide that specifically hybridizes to a region

comprising the breakpoint. The oligonucleotide used for hybridization must directly hybridize to that breakpoint and, in addition, hybridize to at least 943 nucleotides 5' and 3' of the breakpoint. Hybridization occurs preferably under stringent conditions such as 0.2 X SSC, 0.1% SDS at 65°C. The actual breakpoint within the hybrid *Rhesus box* may vary due to the exact nature of the putative crossover event. Accordingly, the hybrid *Rhesus box* may also be detected using a number of overlapping or non-overlapping oligonucleotides used for hybridization. The hybrid *Rhesus box* may also be detected using other protocols such as restriction analysis (preferably in combination with Southern blot analysis), or PCR technology, as described herein above.

**[0098]** Furthermore, another embodiment of the invention relates to a process to specifically detect a common *RHD* negative haplotype comprising assessing the nucleic acid molecular structure comprising the hybrid *Rhesus box* and the flanking regions thereof.

**[0099]** In accordance with the present invention, assessment of the molecular nucleic acid structure comprises analysis steps such as gel electrophoresis using either agarose gels or polyacrylamide gels, treatment with restriction-enzymes, blotting techniques, such as Southern or Northern blotting or related techniques, such as fluorescence-guided detection of hybridization and other techniques known in the art.

**[0100]** The present invention also relates to a process to specifically detect a *RHD* negative haplotype in a sample comprising the step of detecting any of the breakpoint regions mentioned in the present invention.

**[0101]** In a preferred embodiment the invention relates to the above-mentioned process wherein said detection or assessment comprises the determination of the length of a nucleic acid molecule comprising the hybrid *Rhesus box* or parts thereof.

**[0102]** Again, this preferred embodiment of the method of the invention utilizes standard separation techniques, such as gel electrophoresis or chromatography or standard techniques of nucleotide sequencing as known to a skilled artisan. Preferably the present invention utilizes a commercially available sequencing kit and an automatic sequencing machine from Applied Biosystems (ABI 373A or ABI 377), as further described in Example 5, for this purpose.

**[0103]** Another preferred embodiment of the invention relates to the above-mentioned process wherein said detection or assessment is effected by using PCR-RFLP, PCR-SSP or long-range PCR or a probe specifically hybridizing to the hybrid *Rhesus box*, preferably to the breakpoint or breakpoint region depicted in Figure 4 or 5 or Figure 12 or 13, or hybridizing to the upstream or downstream *Rhesus box*, preferably by Southern blot analysis, gel electrophoresis, biochip-analysis, molecular weight determination or fluorescence.

**[0104]** According to the present invention the term "hybridizing to" relates to stringent or non-stringent conditions. The setting of conditions is well within the skill of the artisan and to be determined according to protocols described, for example, in Sambrook, loc. cit. or Hames and Higgins, "Nucleic acid hybridization, a practical approach", IRC Press, Oxford (1985). The detection of specifically hybridizing sequences will usually require stringent hybridizing and washing conditions such as 0.2 x SSC, 0.1% SDS at 65°C. Non-stringent hybridization conditions for the detection of homologous and not exactly complementary sequences may be set at 6x SSC, 1% SDS at 50°C or 65°C. As is well known, the length of the probe and the composition of the nucleic acid to be determined constitute further parameters of the hybridization conditions.

**[0105]** Furthermore, the invention relates to a vector comprising the nucleic acid molecular structure or the nucleic acid molecule of the invention.

**[0106]** The vector may be used for propagation and/or expression or may be designed for gene transfer or targeting purposes. Methods of producing such vectors are well known in the art. The same holds true for cloning the nucleic acids of the mutation into said vectors, as well as the propagation of vectors in suitable hosts, etc.

**[0107]** The vector may particularly be a plasmid, a cosmid, a virus or a bacteriophage used conventionally in genetic engineering that comprise the nucleic acid molecule of the invention. Expression vectors derived from viruses such as retroviruses, vaccinia virus, adeno-associated virus, herpes viruses, or bovine papilloma virus, may be used for delivery of the nucleic acid molecules or vector of the invention into targeted cell populations. Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors; see, for example, the techniques described in Sambrook, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, the polynucleotides and vectors of the invention can be reconstituted into liposomes for delivery to target cells. The vectors containing the nucleic acid molecules of the invention can be transferred into the host cell by well-known methods, which vary depending on the type of cellular host. For example, calcium chloride transfection is commonly utilized for prokaryotic cells, whereas calcium phosphate treatment or electroporation may be used for other cellular hosts; see Sambrook, supra.

**[0108]** Such vectors may comprise further genes such as marker genes which allow for the selection of said vector in a suitable host cell and under suitable conditions. Preferably, the nucleic acid molecule of the invention is operatively linked to expression control sequences allowing expression in prokaryotic or eukaryotic cells. Expression of said polynucleotide comprises transcription of the polynucleotide into a translatable mRNA. Regulatory elements ensuring expression in eukaryotic cells, preferably mammalian cells, are well known to those skilled in the art. They usually comprise regulatory sequences ensuring initiation of transcription and optionally poly-A signals ensuring termination of transcription

and stabilization of the transcript. Additional regulatory elements may include transcriptional as well as translational enhancers, and/or naturally-associated or heterologous promoter regions. Possible regulatory elements permitting expression in prokaryotic host cells comprise, e.g., the PL, lac, trp or tac promoter in *E. coli*, and examples for regulatory elements permitting expression in eukaryotic host cells are the AOX1 or GAL1 promoter in yeast or the CMV-, SV40-, RSV-promoter (Rous sarcoma virus), CMV-enhancer, SV40-enhancer or a globin intron in mammalian and other animal cells. Beside elements which are responsible for the initiation of transcription such regulatory elements may also comprise transcription termination signals, such as the SV40-poly-A site or the tk-poly-A site, downstream of the nucleic acid molecule. Furthermore, depending on the expression system used leader sequences capable of directing the polypeptide to a cellular compartment or secreting it into the medium may be added to the coding sequence of the polynucleotide of the invention and are well known in the art. The leader sequence(s) is (are) assembled in appropriate phase with translation, initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein, or a portion thereof; into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an C- or N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product. In this context, suitable expression vectors are known in the art such as Okayama-Berg cDNA expression vector pcDV1 (Pharmacia), pCDM8, pRc/CMV, pcDNA1, pcDNA3 (In-vitrogene), or pSPORT1 (GIBCO BRL).

**[0109]** Preferably, the expression control sequences will be eukaryotic promoter systems in vectors capable of transforming or transfecting eukaryotic host cells, but control sequences for prokaryotic hosts may also be used.

**[0110]** As mentioned above, the vector of the present invention may also be a gene transfer or targeting vector. Gene therapy, which is based on introducing therapeutic genes into cells by ex-vivo or in-vivo techniques is one of the most important applications of gene transfer. Suitable vectors and methods for in-vitro or in-vivo gene therapy are described in the literature and are known to the person skilled in the art; see, e.g., Giordano, *Nature Medicine* 2 (1996), 534-539; Schaper, *Circ. Res.* 79 (1996), 911-919; Anderson, *Science* 256 (1992), 808-813; Isner, *Lancet* 348 (1996), 370-374; Muhlhauser, *Circ. Res.* 77 (1995), 1077-1086; Wang, *Nature Medicine* 2 (1996), 714-716; WO94/29469; WO 97/00957 or Schaper, *Current Opinion in Biotechnology* 7 (1996), 635-640, and references cited therein. The polynucleotides and vectors of the invention may be designed for direct introduction or for introduction via liposomes, or viral vectors (e.g. adenoviral, retroviral) into the cell. Preferably, said cell is a germ line cell, embryonic cell, or egg cell or derived therefrom, most preferably said cell is a stem cell.

**[0111]** Additionally, the invention relates to a non-human host transformed with the vector of the invention

**[0112]** Appropriate hosts comprise transgenic animals, cells such as bacteria, yeast cells, animal, preferably mammalian cells, fungal cells or insect cells. Transformation protocols including transfection, microinjection, electroporation, etc., are also well known in the art.

**[0113]** Further, the invention relates to a method of producing a protein product of the *RHD* gene comprising culturing the host of the invention under suitable conditions and isolating the protein product of the *RHD* gene.

**[0114]** It is preferred that the protein product of the *RHD* gene is exported into the culture medium where it can be collected according to conventions/methods. The term "culturing" as used in accordance with the present invention also comprises the raising of transgenic animals. Using appropriate vectors constructions and optionally appropriate feeds, the antigen may, e.g., be isolated from milk of, e.g. transgenic cows.

**[0115]** The invention additionally relates to a protein product of the *RHD* gene encoded by the nucleic acid molecule of the invention or produced by the method of the invention.

**[0116]** Preferably, the protein is in the same way post translationally modified and has the same chemical structure as naturally occurring antigen. Accordingly, said protein, when produced by the method of the invention, is preferably produced in human cells.

**[0117]** Furthermore, the invention relates to an oligonucleotide hybridizing under stringent conditions to a portion of the nucleic acid molecular structure or the nucleic acid molecules of the invention, wherein said portion comprises said (missense) mutation or said stop codon or to the complementary portion thereof or hybridizing to a breakpoint of the gene conversion identified here in the above.

**[0118]** In this embodiment of the invention, it is understood that the oligonucleotides hybridizes directly to the mutated sequence or to the breakpoint. The setting of stringent hybridization conditions is well described, for example, in Sambrook et al, "Molecular Cloning, A Laboratory Handbook" CSH Press, Cold Spring Harbor 1989 or Hames and Higgins, "Nucleic acid hybridization, a practical approach", IRL Press, Oxford (1985). Thus, the detection of the specifically hybridizing sequences will usually require hybridization and washing conditions such as 0.2 x SSC, 0.1% SDS at 65°. As is well known, the length of the probe and the composition of the nucleic acid to be determined constitute further parameters of the stringent hybridization conditions. Preferably, the oligonucleotide is a deoxynucleotide. It is further preferred that the oligonucleotide comprises 12 to 50 nucleotides and more preferably 15 to 24 nucleotides. The hybridization to the breakpoint may be under stringent or non-stringent conditions. An example of non-stringent hybridization conditions is hybridization and washing at 50°C in 4 x SSC, 0,1% SDS.

**[0119]** Further, the invention relates to an antibody or an aptamer specifically binding to the protein product of the

*RHD* gene of the invention.

**[0120]** The antibody may be tested and used in any serologic technique well known in the art, like agglutination techniques in tubes, gels, solid phase and capture techniques with or without secondary antibodies, or in flow cytometry with or without immunofluorescence enhancement.

**[0121]** The antibody of the invention may be a monoclonal antibody or an antibody derived from or comprised in a polyclonal antiserum. The term "antibody", as used in accordance with the present invention, further comprises fragments of said antibody such as Fab, F(ab')<sub>2</sub>, Fv or scFv fragments; see, for example, Harlow and Lane, "Antibodies, A Laboratory Manual" CSH Press 1988, Cold Spring Harbor, N.Y. The antibody or the fragment thereof may be of natural origin or may be (semi) synthetically produced. Such synthetic products also comprise non-proteinaceous or semi-proteinaceous material that has the same or essentially the same binding specificity as the antibody of the invention. Such products may, for example, be obtained by peptidomimetics.

**[0122]** The term "aptamer" is well known in the art and defined, e.g., in Osborne et al., Curr. Opin. Chem. Biol. 1 (1997), 5-9 or in Stall and Szoka, Pharm. Res. 12 (1995), 465-483.

**[0123]** Furthermore, the invention relates to a method for testing for the presence of a nucleic acid molecule encoding a mutant Rhesus D antigen or of a nucleic acid molecule carrying a deletion of the *RHD* gene as characterized by the nucleic acid molecular structure or the nucleic acid molecule of the invention in a sample comprising hybridizing the oligonucleotide of the invention under stringent conditions to nucleic acid molecules comprised in the sample obtained from a human and detecting said hybridization.

**[0124]** Preferably, the method of the invention further comprises digesting the product of said hybridization with a restriction endonuclease or subjecting the product of said hybridization to digestion with a restriction endonuclease and analyzing the product of said digestion.

**[0125]** This preferred embodiment of the invention allows by convenient means, the differentiation between an effective hybridization and a non-effective hybridization. For example, if the wild type *RHD* gene comprises an endonuclease restriction site, the hybridized product will be cleavable by an appropriate restriction enzyme whereas a mutated sequence will yield no double-stranded product or will not comprise the recognizable restriction site and, accordingly, will not be cleaved. Alternatively, the hybridizing oligonucleotide may only hybridize to the mutated sequence. In this case, only a hybrid comprising the mutated sequence, but not the wild type sequence, will be cleaved by the appropriate restriction enzyme. The analysis of the digestion product can be effected by conventional means, such as by gel electrophoresis which may be optionally combined by the staining of the nucleic acid with, for example, ethidium bromide. Combinations with further techniques such as Southern blotting are also envisaged.

**[0126]** Detection of said hybridization may be effected, for example, by an anti-DNA double-strand antibody or by employing a labeled oligonucleotide. Conveniently, the method of the invention is employed together with blotting techniques such as Southern or Northern blotting and related techniques. Labeling may be effected, for example, by standard protocols and includes labeling with radioactive markers, fluorescent, phosphorescent, chemiluminescent, enzymatic labels, etc.

**[0127]** The invention also relates to a method to simultaneously detect the presence of *RHD* $\Psi$  and any of the *RHD* molecular structures of the invention comprising hybridizing the oligonucleotide of the invention and at least an other oligonucleotide hybridizing to a *RHD* $\Psi$  structure under stringent conditions to nucleic acid molecules comprised in the sample obtained from a human and detecting said hybridization.

**[0128]** The present invention further relates to a method for testing simultaneously for the presence of *RHD* $\Psi$  and any of the *RHD* molecular structures of the present invention in a sample comprising determining the nucleic acid sequence of at least a portion of the nucleic acid molecular structure or nucleic acid molecule of the present invention, said portion encoding said (missense) mutation, said stop codon or a breakpoint of said hybrid gene and determining of at least a portion of a *RHD* $\Psi$  structure.

**[0129]** The invention additionally relates to a method for testing for the presence of a nucleic acid molecule encoding a mutant Rhesus D antigen or of a nucleic acid molecule carrying a deletion of the *RHD* gene as characterized by the nucleic acid molecular structure or the nucleic acid molecule of the invention in a sample comprising determining the nucleic acid sequence of at least a portion of the nucleic acid molecular structure or the nucleic acid molecule of the invention, said portion encoding said (missense) mutation, said stop codon or a breakpoint of said hybrid gene.

**[0130]** Preferably, the method of the invention further comprises, prior to determining said nucleic acid sequence, amplification of at least said portion of said nucleic acid molecular structure.

**[0131]** Moreover, the invention relates to a method for testing simultaneously for the presence of *RHD* $\Psi$  and any of the *RHD* molecular structures of the present invention in a sample comprising carrying out an amplification reaction using a set of primers that amplifies at least a portion of said sequence wherein at least one of the primers employed in said amplification reaction is the oligonucleotide of the present invention and at least a primer amplifying (e.g. by specifically hybridizing to) a *RHD* $\Psi$  structure and analysing the amplified product(s).

**[0132]** *RHD* $\Psi$  is a *RHD* allele detected in D negatives that has been characterized by Singleton et al. (Singleton, B.K., Green, C.A., Avent, N.D., Martin, P.G., Smart, E., Daka, A., Narter-Olaga, E.G., Hawthorne, L.M., and Daniels, G. The

presence of an RHD pseudogene containing a 37 base pair duplication and a nonsense mutation in africans with the Rh D- negative blood group phenotype. Blood 95(1):12-18, 2000).

**[0133]** Furthermore, the invention relates to a method for testing for the presence of a nucleic acid molecule encoding a mutant Rhesus D antigen or of a nucleic acid molecule carrying a deletion of the *RHD* gene as characterized by the nucleic acid molecular structure or the nucleic acid molecule of the invention in a sample comprising carrying out an amplification reaction using a set of primers that amplifies at least a portion of said sequence wherein at least one of the primers employed in said amplification reaction is the oligonucleotide of the invention.

**[0134]** Moreover, in a further embodiment the method of the invention wherein at least one of the primers employed in said amplification reaction is the oligonucleotide of the invention.

**[0135]** Preferably, amplification is effected by polymerase chain reaction (PCR). Other amplification methods such as ligase chain reaction may also be employed.

**[0136]** In a preferred embodiment of the method of the invention said PCR is PCR-RFLP, PCR-SSP or long-range PCR.

**[0137]** Additionally, in another preferred embodiment of the invention the molecular weight of the amplification product is analyzed. Said analysis of the molecular weight utilizes standard techniques, such as agarose gel electrophoresis, SDS-PAGE, mass spectrometry such as MALDI-TOF for this purpose, which are well known to the person skilled in the art.

**[0138]** In one preferred embodiment of the method of the invention, said method detects *RHD* positive alleles comprising the following steps:

- (a) isolating DNA from a blood sample or blood donor;
- (b) hybridizing at least two oppositely oriented primers under stringent conditions to the DNA so as to carry out a PCR;
- (c) amplifying the target sequence;
- (d) separating the amplification products on a gel; and
- (e) analyzing the amplicons.

**[0139]** With regard to specific conditions to be applied in the various steps, it is referred to the corresponding description herein above.

**[0140]** In a preferred embodiment the *RHD* positive alleles are derived from a serologically RhD negative population. In another preferred embodiment the RhD-negative sample is selected from a Caucasian population.

**[0141]** The method of the invention will result in an amplification of only the target sequence, if said target sequence carries the or at least one mutation. This is because the oligonucleotide will, under preferably stringent hybridization conditions, not hybridize to the wild type sequence (with the consequence that no amplification product is obtained) but only to the mutated sequence. Naturally, primer oligonucleotides hybridizing to one or more as one, such as two mutated sequences may be employed in the method of the invention. The latter embodiment may be favorable in cases where combinations of mutations are tested for. It is important to note that not all or none of said mutations are necessarily missense mutations. This may be true for cases where other types of mutations occur in combination with the above missense mutations or with the above gene conversion.

**[0142]** Preferably, in the method of the invention said amplification or amplification reaction is or is effected by the polymerase chain reaction (PCR). Other amplification methods such as ligase chain reaction may also be employed.

**[0143]** Further, the invention relates to a method for testing for the presence of a protein product of the *RHD* gene of the invention in a sample comprising assaying a sample obtained from a human for specific binding to the antibody or aptamer or phage of the invention.

**[0144]** Testing for binding may, again, involve the employment of standard techniques such as ELISAs; see, for example, Harlow and Lane, "Antibodies, A Laboratory Manual" CSH Press 1988, Cold Spring Harbor.

**[0145]** In another preferred embodiment the invention relates to a method for testing for the presence of a protein product of the *RHD* gene encoding the nucleic acid molecular structure or the nucleic acid molecule of the invention, comprising utilizing direct agglutination methods, indirect antiglobulin tests, monoclonal anti-D antibodies and adsorption/elution techniques.

**[0146]** Thus, the embodiment may comprise direct agglutination with two monoclonal anti-D antibodies, alternatively indirect antiglobulin tests using a gel matrix comprising an oligoclonal anti-D antibody, in a further alternative using monoclonal anti-Rhesus antibodies in another alternative adsorption of polyclonal anti-D antibodies to red cells and elution using a chloroform technique. Further description of the methods is given in example 18.

**[0147]** Preferably, in the method of the invention said sample is blood, serum, plasma, fetal tissue, saliva, urine, mucosal tissue, mucus, vaginal tissue, fetal tissue obtained from the vagina, skin, hair, hair follicle or another human tissue.

**[0148]** Furthermore, the method of the invention preferably comprises the step of enrichment of fetal cells. This enrichment may be achieved by using appropriate antibodies, lectins or other reagents specifically binding fetal cells or by any technique attempting the differential separation of maternal and fetal cells, like by density gradients. Also preferably, in said method fetal DNA or mRNA from maternal tissue like peripheral blood, serum or plasma may be extracted, advantageously according to conventional procedures.

**[0149]** In an additional preferred embodiment of the method of the invention, said nucleic acid molecule or proteinaceous material from said sample is fixed to a solid support.

**[0150]** Preferably, said solid support is a chip.

5 **[0151]** The advantages of chips are well known in the art and need not be discussed herein in detail. These include the small size as well as an easy access of computer based analysis of analytes.

**[0152]** Furthermore, the present invention relates to the use of the nucleic acid molecular structure or the nucleic acid molecule of the invention for the analysis of a negative or a positive Rhesus D phenotype.

**[0153]** The analysis can be effected, for example, on the basis of the methods described herein above.

10 **[0154]** The invention also relates to the use of the nucleic acid molecular structure or the nucleic acid molecule of the invention, the vector of the invention or the protein product of the *RHD* gene of the invention for the assessment of the affinity, avidity and/or reactivity of monoclonal anti-D or anti-C antibodies or of polyclonal anti-D or anti-C antisera or of anti-globulin or of anti-human-globulin antisera or of preparations thereof.

**[0155]** Anti-C is a monoclonal antibody or polyclonal antiserum binding to antigen C.

15 **[0156]** The invention also relates to the use of cells, preferably red blood cells, from probands carrying the nucleic acid molecular structure or the nucleic acid molecule of the invention for the assessment of the affinity, avidity and/or reactivity of monoclonal anti-D or anti-C antibodies or of polyclonal anti-D or anti-C antisera or of anti-globulin or of anti-human-globulin antisera or of preparations thereof.

20 **[0157]** Said preparations can be provided according to techniques well known in the art. Said preparations may comprise stabilisators such as albumins, further sodium azide, salt ions, buffers etc. The formulation of the preparation may have an influence on the binding characteristics of the antibodies, as is well known in the art.

25 **[0158]** For example, in a first step, the Rhesus D gene of a carrier or of a blood donor and its allelic status is analyzed and it is determined whether said gene comprises a mutation that was found in accordance with the present invention. In a second step, said mutation is correlated to a certain RhD antigen density on the surface of red blood cells. Conveniently, said correlation can be established by data provided in the present invention (such as mutations per se) and techniques that are well known in the art (see, e.g. Jones et al. 1996, Flegel and Wagner, 1996). In a third step, the features of an antibody or an antiserum such as reactivity, sensitivity, affinity, avidity, and/or specificity are determined with suitable blood group serological techniques preferably using red blood cells that were molecularly and with respect to the RhD antigen surface density characterized as described in step 2. Such data can be used, for example, in quality controls, standardization, etc.

30 **[0159]** The invention will be most useful for the characterization, standardization and quality control of monoclonal and polyclonal antisera, preferably anti-D monoclonals or antisera. Further, for example, anti-globulin and anti-human-globulin antisera can be characterized on the basis of the teachings of the present invention. An appropriately characterized anti-D monoclonal antibody can be conveniently used in RhD diagnostics. For example, a suitably characterized monoclonal antibody will be useful in determining the D antigen density on the surface of blood cells. Cut-off values for monoclonal antibodies useful in diagnosis can thus be established. This is important for the quality control of antibodies used in RhD diagnosis.

35 **[0160]** Thus, the invention also relates to a method for the characterization of monoclonal antibodies or polyclonal antisera or of a preparation thereof, said method comprising

- 40 (a) testing the nucleic acid of sample of a proband for the presence of a breakpoint or mutation as defined in accordance with the invention;
- (b) correlating, on the basis of the mutation or deletion status and the allelic status of the *RHD* gene, the nucleic acid with the density of the protein product of the *RHD* gene on the surface of red blood cells of said proband;
- 45 (c) reacting said monoclonal antibodies or polyclonal antisera or said preparation thereof with a cell carrying the protein product of the *RHD* gene on its surface;
- (d) characterizing said monoclonal antibodies or polyclonal antisera or said preparation thereof on the basis of the results obtained in step (c).

50 **[0161]** As regards the term "allelic status", this term describes the possibilities that the *RHD* alleles in a proband are present in a homozygous, heterozygous or hemizygous state. Also comprised by this term is the possibility that the two alleles carry two different mutations (including the conversion) defined herein above.

**[0162]** In a preferred embodiment of the method of the invention, said characterization comprises the determination of reactivity, sensitivity, avidity, affinity, specificity and/or other characteristics of antibodies and antisera.

55 **[0163]** Furthermore preferred is a method wherein said cell carrying the protein product of the *RHD* gene on its surface is a red blood cell.

**[0164]** The invention also relates to a method for determining whether a patient in need of a blood transfusion is to be transfused with RhD negative blood from a donor comprising the step of testing a sample from said patient for the presence of one or more nucleic acid molecular structures or nucleic acid molecules of the invention, wherein a positive

testing for two different of said nucleic acid molecular structures or nucleic acid molecules is indicative of the need for a transfusion with Rh negative blood. Alternatively, a positive testing indicating the concomitant presence of two identical copies of one of said nucleic acid molecular structures or nucleic acid molecules is indicative of the need for a transfusion with Rh negative blood.

5 **[0165]** Alternatively, a negative testing for the presence of the nucleic acid molecular structure or nucleic acid molecule representative of the common *RHD* negative haplotype with or without a negative testing for one or more nucleic acid molecular structures or nucleic acid molecules representative of the other *RHD* negative nucleic acid molecular structures or nucleic acid molecules of this invention permits the transfusion of blood that is typed as RhD positive. The invention has important implications for devising a transfusion therapy in humans. For example, it can now be conveniently tested  
10 whether the patient actually needs a transfusion with a RhD negative blood or whether such precautions need not be taken.

**[0166]** The invention also relates to a method for determining whether blood of a donor is suitable for transfusion to a patient in need thereof who should not be exposed to antigen C comprising the step of testing a sample from said donor for the presence of the nucleic acid molecular structure of the present invention wherein a positive testing for the nucleic acid molecular structure of the present invention precludes the transfusion of the donor's blood.

15 **[0167]** Furthermore, the invention relates to a method for determining whether blood of a donor may be used for transfusion to a patient in need thereof comprising the step of testing a sample from said donor for the presence of one or more of said nucleic acid molecular structures or nucleic acid molecules of the invention, wherein a negative testing for the nucleic acid molecular structures representative of the common *RHD* negative haplotype with or without a negative testing for one or more nucleic acid molecular structures or nucleic acid molecules representative of the other *RHD*  
20 negative haplotypes of this invention excludes the transfusion of the donor's blood to a patient that is typed as RhD negative.

**[0168]** The invention also relates to a method for determining whether the blood of a donor may be transfused to a patient typed as RhD negative comprising the step of testing a sample from said donor for the presence of one or more nucleic acid molecular structures or nucleic acid molecules of the invention, wherein a positive testing for two different  
25 of said nucleic acid molecular structures or nucleic acid molecules is indicative of the possibility to transfuse the donor's blood to a patient typed as RhD negative.

**[0169]** Alternatively, a positive testing indicating the concomitant presence of two identical copies of one of said nucleic acid molecular structures or nucleic acid molecules is indicative of the possibility to transfuse this donor's blood to a patient that is typed as RhD negative.

30 **[0170]** The samples referred to in the above recited methods may be samples that are referred to throughout the specification, such as blood, serum, etc.

**[0171]** As regards the guidelines for transfusing a patient on the basis of any of the above recited methods, the utmost care must be taken that suboptimal transfusion policy is avoided. The risk factor is always to be considered by the physician in charge. In all cases, the potential risk for the patient is to be minimized.

35 **[0172]** The invention also relates to a method for assessing of the risk of a RhD negative mother of conceiving or carrying an RhD positive fetus or of the risk of a mother having an anti-D titer of conceiving or carrying a fetus at risk to develop hemolytic disease of the newborn comprising assessing a sample obtained from the father of the fetus for the presence of one or more of said nucleic acid molecular structures or nucleic acid molecules of the invention, wherein a negative testing for nucleic acid molecular structures or nucleic acid molecules representative of the common *RHD*  
40 negative haplotype with or without a negative testing for one or more nucleic acid molecular structures or nucleic acid molecules representative of the other *RHD* negative haplotypes of this invention is indicative for a high risk of conceiving an RhD positive fetus.

**[0173]** In a preferred embodiment of the method of the present invention said nucleic acid molecular structure carries mutations or deletions.

45 **[0174]** The invention also relates to a method for determining whether the father is RhD negative comprising the step of testing a sample from the father for the presence of one or more nucleic acid molecular structures or nucleic acid molecules of the invention, wherein a positive testing for two different of said nucleic acid molecular structures or nucleic acid molecules is indicative of the father being RhD negative.

**[0175]** Alternatively, a positive testing indicating the concomitant presence of two identical copies of one of said nucleic acid molecular structures or nucleic acid molecules representative of *RHD* negative haplotypes is indicative of the father  
50 being RhD negative.

**[0176]** Furthermore, the invention relates to a method for assessing the possibility or likelihood of a man being the father of a child by assaying a sample obtained from said man for the presence of one or more of said nucleic acid molecular structures or nucleic acid molecules of the invention, wherein the test results are used to determine the homozygosity for, the heterozygosity for or the absence of any nucleic acid molecular structures or nucleic acid molecules  
55 representative of the *RHD* negative haplotype of the present invention used to infer the possibility or likelihood of said man being the father of the child.

**[0177]** The preparation may be a diagnostic or pharmaceutical preparation.

**[0178]** The pharmaceutical composition of the present invention may further comprise a pharmaceutically acceptable

carrier and/or diluent. Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions etc. Compositions comprising such carriers can be formulated by well known conventional methods. These pharmaceutical compositions can be administered to the subject at a suitable dose.

**[0179]** Preparations for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like. Furthermore, the pharmaceutical composition of the invention may comprise further agents such as interleukins or interferons depending on the intended use of the pharmaceutical composition.

**[0180]** The invention also relates to a method of treating a pregnant woman being Rhesus D negative wherein the fetus does not carry two nucleic acid molecular structures or nucleic acid molecules of the invention or is not homozygous for any nucleic acid molecular structure or nucleic acid molecule of the invention, comprising administering anti-D to said woman.

**[0181]** Pregnant women may be currently treated with an anti-D prophylaxis, when a Rhesus negative mother carries a RhD positive fetus. The invention allows the discrimination of an anti-D prophylaxis requirement depending on the status of the mother's and/or the fetus' possessing a RhD protein of the invention. One or more of the RhD proteins of the invention may be prone to immunization of their carriers and, hence, would be indicative for the therapy of the mother. Similarly, one or more RhD proteins of the invention, when carried by the fetus, may be known to be of low immunogenicity to the mother and, hence, would be indicative for the omission of anti-D prophylaxis in difference to current clinical therapy.

**[0182]** The administration can be effected by standard routes and doses which can be defined by the attending physician; Mollison, 1993. Preferably, a monoclonal anti-D or combinations/mixtures of monoclonal anti-Ds is/are administered in doses of 50  $\mu$ g to or exceeding 500  $\mu$ g anti-D antibody/antisera for intravenous or intramuscular administration (Bowman, 1998). For the quality control of these anti-D antibodies/antisera, the results and methods provided by the present invention may be advantageously employed.

**[0183]** The invention also relates to the use of a phage, aptamer, monoclonal antibody or a polyclonal antisera or a preparation thereof as characterized in the present invention for determination of the protein product of the *RHD* gene.

**[0184]** In a preferred embodiment of said use, said determination of the protein product of the *RHD* gene is effected in connection with blood group typing.

**[0185]** Furthermore, the invention relates to a preparation comprising the antibody or aptamer or phage of the invention.

**[0186]** The present invention also relates to a method of identifying an antibody  $V_H$  or  $V_L$  chain or a combination thereof or an aptamer specifically binding to a protein product of the *RHD* gene of the invention comprising

(a) contacting the protein product of the *RHD* gene of the invention with a phage library displaying  $V_H$  or  $V_L$  chains or combinations thereof on the surface of the phage or with aptamers;

(b) identifying phage or aptamers that bind to said protein product of the *RHD* gene; and optionally

(c) repeating steps (a) and (b) one or more times.

**[0187]** The preparation of phage library and the screening/identification of desired antibody (chains) per se is well known in the art and reviewed, for example, in Winter et al., Annu. Rev. Immunol. 12 (1994), 433-455 and references cited therein. Also, aptamers can be prepared and cloned in phage according to conventional protocols. Whereas single  $V_H$  or  $V_L$  chains may be identified by the method of the invention as binding to the protein product of the *RHD* gene of the invention, it is preferred to identify  $V_H$ - $V_L$  combinations expressed by the phage because this situation resembles the situation of natural antibody binding. By repeating steps (a) and (b) one or more times, better binding specificities may be identified. Protocols for the optimization of binding properties such as affinities, including elution steps for removing bound phage, are well established in the art. For example, once a  $V_H$  chain with a convenient binding capacity has been found,  $V_L$  chains may be identified that significantly improve the binding capacity of the antibody, e.g. by replacing the  $V_L$  chain that was associated with the  $V_H$  chain in the first selection step with a more suitable  $V_L$  chain.

**[0188]** The invention also relates to a method of identifying a monoclonal antibody specifically binding to a protein product of the *RHD* gene of the invention comprising

(a) contacting the protein product of the *RHD* gene of the invention with one or more monoclonal antibodies;

(b) identifying monoclonal antibodies that bind to said protein product of the *RHD* gene; and optionally

(c) repeating steps (a) and (b) one or more times.

**[0189]** The invention also relates to a method of identifying an antibody V<sub>H</sub> or V<sub>L</sub> chain or a combination thereof or an aptamer specifically binding to a protein product of the *RHD* gene of the invention comprising

(aa) contacting said protein product of the *RHD* gene and

(ab) a normal D polypeptide

wherein the normal D polypeptide is present in a molar mass that is higher, equal or less than the protein product of the *RHD* gene of (aa) with a phage library displaying V<sub>H</sub> or V<sub>L</sub> chains or combinations thereof on the surface of the phage or with aptamers;

(b) identifying phage or aptamers that bind to said protein product of the *RHD* gene of (a); and optionally

(c) repeating steps (a) and (b) one or more times.

**[0190]** Particularly preferred in step (ab) is that the molar mass of the normal D polypeptide is higher than that of the protein product of the *RHD* gene of (aa).

**[0191]** In the case that only one round of selection is employed for the identification (i.e. when step (c) does not apply), it is preferred that the number of protein product of the *RHD* gene of (aa) is in molar excess over the number of phage particles. The preferred embodiments of the method of identifying an antibody V<sub>H</sub> or V<sub>L</sub> chain or of a combination thereof or of an aptamer described hereinbefore equally apply to this embodiment of the invention.

**[0192]** The invention also relates to a method of identifying a monoclonal antibody specifically binding to a protein product of the *RHD* gene of the invention comprising

(aa) contacting the protein product of the *RHD* gene and

(ab) a normal D polypeptide

wherein the normal D polypeptide is present in a molar mass that is higher, equal or less than the protein product of the *RHD* gene of (aa) with one or more monoclonal antibodies;

(b) identifying monoclonal antibodies that bind to said protein product of the *RHD* gene of (aa); and optionally

(c) repeating steps (a) and (b) one or more times.

**[0193]** Preferably, the protein product of the *RHD* gene is exposed on the surface of a cell. An appropriate surface is the surface of an erythrocyte. However, other host cells may be transfected with a vector suitable for expression of the protein product of the *RHD* gene of the invention and express the same on their surface. Antibodies may also bind to recombinant proteins or parts of proteins of D antigen and purified proteins.

**[0194]** It is further preferred that the polypeptide or host cell is affixed to a solid support. Suitable examples for solid supports are microtiter plates or beads.

**[0195]** In an additionally preferred antibody, subsequent to step (b) or (c), the following step is carried out:

(d) identifying the amino acid sequence of the V<sub>H</sub> or V<sub>L</sub> chains and/or identifying the nucleic acid sequences encoding said amino acid sequence.

**[0196]** The identification of the amino acid/nucleic acid sequences can be effected according to conventional protocols; see, e.g., Sambrook et al., loc. cit.

**[0197]** Hence and in summary, the present invention provides means and methods for the detection of *RHD* haplotypes, comprising common *RHD* negative haplotypes, as described above, as well as presumably rare *RHD* positive alleles in serologically RhD negative populations. Latter alleles, harbouring *RHD* sequences and therefore determined as *RHD*-positive, can comprise either *RHD/RHCE* hybrid genes, stop codons, splice site mutations or gene deletions, that terminate or reduce the RhD antigen expression. Carrying out the improved detection methods of the invention, it was surprisingly found, that several samples, determined as RhD negative in routine serology, could be identified having *RHD* positive alleles. Furthermore, some of those samples were even RhD antigen positive when performing a detection assay based on adsorption and elution, indicating that the molecular basis for the *RHD* positive alleles in RhD negatives is more heterogeneous than anticipated. Advantageously, the disclosure content of the present invention now provides new and practicable nucleic acid amplification techniques to determine whether *RHD* specific sequences cause RhD positive or RhD negative phenotypes.

**[0198]** In a particularly preferred embodiment the method of present invention, wherein, in the case that only one round of selection is employed for the identification, the number of protein molecules of the *RHD* gene of (a) is in molar excess over the number of phage particles.

**[0199]** Moreover, the present invention relates to the use of cells, preferably red blood cells comprising the protein product of the *RHD* gene of the present invention or produced by the method of the present invention, from probands for the assessment of the affinity, avidity and/or reactivity of monoclonal anti-D or anti-C antibodies of the present invention or of polyclonal anti-D or anti-C antisera or of anti-globulin or of anti-human-globulin antisera or of preparations

thereof.

**[0200]** Furthermore, the invention relates to the use of *SMP1* polymorphisms to determine specific *RH* (*RHD-RHCE*)-haplotypes genetically linked to said polymorphisms.

**[0201]** Basis for this embodiment of the present invention is provided by the unique structure of the *RH* locus comprising three genes: *RHD*, *RHCE*, and *SMP1*. The nucleotide sequence of the latter gene has been deposited in the Genbank as putative member of an 18 kDa small membrane protein family. Its function is as yet unknown. It shows homology to an open reading frame on chromosome 21 (Reboul, Genome res. 9:242, 1999). Its position between both *RH* genes implies that any polymorphism of the *SMP1* gene would be tightly linked to a specific *RH* haplotype, and it might be anticipated that functionally relevant mutations of the *SMP1* gene may cause selection pressure for or against specific *RH* haplotypes. Such factors might explain some previously unresolved issues of *RH* haplotype distribution, like the high frequency of *RH* negatives in Europe. Screening for polymorphisms in *SMP1* is therefore of high interest for the understanding of the *RH* locus as well as for diagnostic applications thereof.

**[0202]** According to the present invention the term "polymorphism" relates to the existence in a population of more than one genetic structure or a gene of a haplotype or of a DNA segment. Nevertheless, sometimes such a genetic polymorphism does not always result in a differing phenotype, but may only be detected at the genetic level.

**[0203]** In another preferred embodiment the invention relates to a method to detect specific *RH* (*RHD-RHCE*)-haplotypes comprising the determination of *SMP1*-polymorphisms in the *SPM1* gene by utilizing any structural feature or nucleotide sequence or both of the *RH* locus or combinations thereof with techniques known in the art, preferably by PCR-RFLP, PCR-SSP or long-range PCR.

**[0204]** Furthermore, the invention relates to a kit comprising

- (a) the oligonucleotide of the invention; and/or
- (b) the antibody of the invention;
- (c) the aptamer of the invention; and/or
- (d) the phage of the invention;
- (e) a pair of primers useful for carrying out the amplification reaction of the invention.

**[0205]** Parts of the kit can be packaged individually in vials or in combination in containers or multicontainer units. The kit of the present invention may be advantageously used for carrying out the method of the invention and could be, inter alia, employed in a variety of applications referred to above. The manufacture of the kits follows preferably standard procedures which are known to people skilled in the art.

**[0206]** The present invention also relates to a process to determine the presence of an antigen C encoded by a *RHD* gene comprising the step of detecting any of the breakpoint regions mentioned in the present invention.

**[0207]** Finally, the invention relates to a process to determine the presence of an antigen C comprising the steps of the processes of the present invention.

**[0208]** The disclosure content of the documents as cited in this specification is herewith incorporated by reference.

**[0209]** The figures show

**Figure 1** Schematic structure of the *RH* gene locus. The positions and orientations of the genes and the *Rhesus boxes* are indicated by open arrows and triangles, respectively (Panel A). The exons are shown as vertical bars and their exon number is indicated. The two *RH* genes have opposite orientation, face each other with their 3' ends, and are separated by about 30,000 bp. A third gene, *SMP1*, has the same orientation as *RHD* and is positioned in between *RHD* and *RHCE*. The *RHD* gene is flanked on both sides by the two highly homologous *Rhesus boxes* (b). All exons are shorter than 200 bp with the exception of the *RHD* and *SMP1* 3' terminal exons. Data used to establish this structure (Panel B) include the extension of genomic sequences represented in the cDNAs (horizontal arrows), identities and homologies to genomic clones (bar a: identity with dJ465N24; b: homology of *RHD* to dJ469D22; c: homology of *RHD* 3' part to dJ465N24; d: identity with dJ469D22). The positions of three bridging PCR reactions are indicated. The correct position of a nucleotide stretch previously reported by Okuda et al. (Okuda, Biochem. Biophys. Res. Commun. 263:378, 1999) as "spacer" sequence between *RHD* and *RHCE* is indicated by the bar labeled s.

**Figure 2** Chromosomal organization of the DNA regions located 5' to the *RHD* and *RHCE* genes. The proposed structure of the *RHCE* and *RHD* 5' flanking regions is depicted (Panel A). A total of 4,941 bp immediately 5' of the ATG start codons are homologous between the *RHCE* and *RHD* genes (vertically hatched bars). No homology is present further beyond this homology region (diagonally hatched bars). Two genomic clones, dJ469D22 and dJ465N24, were utilized for primer design. DJ469D22 comprises the full length of the depicted *RHCE* region, whereas dJ465N24 extends only 466 bp into the homology region. The positions

## EP 1 780 217 A1

of several PCR primers are indicated (a, rey14a; b, rend32; c, rey15a; d, re014; e, re011d). This proposed structure is supported by several PCR reactions (panel B). Forward priming was done with primer a (*RHCE* specific, lane 1 - 3), primer b (*RHD* specific, lane 4 - 6), and primer c (*RHCE* and *RHD* homology region, lane 7 - 9). Amplicons were lacking for primer a with *RHD* specific reverse primer e (lane 2) and for primer b with *RHD* negative DNA (lane 6). The other seven PCR reactions yielded amplicons of the predicted sizes in accordance with the genomic structure shown in panel A.

**Figure 3** Chromosomal organization of the *SMP1* gene. The *SMP1* gene has seven exons. The positions and approximate sizes of the introns are shown. The start of the published cDNA (GenBank accession number AF081282) is separated by 15 nucleotides from the downstream *Rhesus box*. Exon 1 contains only 5' untranslated sequence, the *SMP1* start codon is located in exon 2. Exon 7 contains 16 codons and 1,656 bp 3' untranslated sequences and is contiguous with the 3' untranslated sequence of *RHCE* exon 10.

**Figure 4** Chromosomal organization of the *Rhesus boxes*. The physical extension of the upstream *Rhesus box* (5' to *RHD*) is 9,145 bp (black bar). About 63% of the boxes' nucleotide sequence consists of repetitive DNA; the types of the repeat families are indicated. The overall homology between the upstream and downstream *Rhesus box* is 98.6%, but within an 1,463 bp identity region (horizontal arrows), there is only a single 4 bp insertion (double vertical line). A CpG-island (double-headed arrow) is located at the 3' end and is in the downstream *Rhesus box* (3' to *RHD*) adjacent to the *SMP1* promoter.

**Figure 5** *RHD* gene deletion in the Rh negative haplotypes. Three 3,100 bp segments of the *Rhesus boxes* are shown. The upper line indicates the nucleotide sequence of the upstream *Rhesus box* in D-positives, the lower line the nucleotide sequence of the downstream *Rhesus box* in D-positives. The middle line gives the nucleotide sequence of the single *Rhesus box* carried by Rh negatives. Asterisks denote identical nucleotides. The *RHD* deletion occurred in a 903 bp segment of absolute identity that was part of a 1,463 bp identity region. The positions of primers rez7 and mb31 is shown (m indicates mismatch). *Pst*I restriction sites are indicated by carets (^). The three *Rhesus boxes* are deposited at EMBL under accession numbers AJ252311 (upstream *Rhesus box*), AJ252312 (downstream *Rhesus box*), and AJ252313 (hybrid *Rhesus box*).

**Figure 6** Two technical procedures for specific detection of the *RHD* deletion in the common *RHD* negative haplotypes. A long-range PCR amplification with primers located in non-*Rhesus box* sequences (Panel A) and PCR-RFLP with primers located in the *Rhesus boxes* are shown (Panel B). The deduced genotypes are indicated. The primers of the long-range PCR were located 5' of the upstream *Rhesus box* (primer rez4) and in *SMP1* exon 1 (primer sr9). *RHD* negative haplotypes were detected specifically (Panel A, lane 1 - 6). DNA homozygous for the *RHD* gene was negative, because the PCR cannot amplify the 70,000 bp DNA stretch of the *RHD* gene. For the PCR-RFLP method, the PCR amplicons (primer rez7 and rnb31) were digested with *Pst*I. In D-negatives, there are three *Pst*I sites in the amplicon (see Fig. 5) resulting in fragments of 1,888 bp, 564 bp, 397 bp, and 179 bp (lane 1 to 3). The downstream *Rhesus box* of D-positives lacks one *Pst*I-site resulting in fragments of 1,888 bp, 744 bp, and 397 bp (lane 7 to 9). *RHD*<sup>+</sup>/*RHD*<sup>-</sup> heterozygotes show both fragments of 744 and 564 bp (lane 4 to 6). The 564 bp fragment appears weaker because heterodimers are not cut by *Pst*I. Primer mb31 does not amplify the upstream *Rhesus box* of D-positives.

**Figure 7** Model of the proposed mechanism causing the prevalent *RHD* negative haplotypes in whites. The physical structure of the *RHD* and *RHCE* gene locus is depicted (panel A). An unequal crossing over between the upstream and downstream *Rhesus boxes* can be triggered by their high homology (panel B). The breakpoint region in the *Rhesus boxes* was found to be of 100% homology for 903 bp (see Fig. 5). Resolving the crossed over chromosome yields the *RH* gene structure of the extant *RHD* negative haplotype (panel C).

**Figure 8** DNA sequence of the hybrid *Rhesus box* of *RHD* negatives.

**Figure 9** DNA sequence of the upstream *Rhesus box* of D-positives.

**Figure 10** DNA sequence of the downstream *Rhesus box* of D-positives.

**Figure 11** DNA sequence of the *RHD* promoter. The last three nucleotides represent codon 1 of the *RHD* gene.

**Figure 12.** *Cde<sup>s</sup>* breakpoint region in *RHD* intron 3. The nucleotide sequence of a part of the intron 3 of *Cde<sup>s</sup>*, *RHD* and *RHCE* 2,938 to 3,636 bp 3' of the exon 3/intron 3 junction is shown. The human DNA sequence from clone RP3-469D22 on chromosome 1p35.1-36.13 containing the 5' part of the gene for *RHCE* (GenBank accession number AL031284) was taken as reference; numbers indicate the position in this sequence relative to the first base of intron 3 in the *RHCE* gene. The corresponding *RHD* gene sequence derives from GenBank accession number AL139426. Nucleotides indicating *RHD* or *RHCE* origin of the *Cde<sup>s</sup>* sequences are highlighted. A 154 bp DNA stretch comprising the breakpoint region of *Cde<sup>s</sup>* is indicated by asterisks.

**Figure 13.** *Cde<sup>s</sup>* breakpoint region in *RHD* intron 7. The nucleotide sequence of a part of the intron 7 of *Cde<sup>s</sup>*, *RHD* and *RHCE* about 2,726 to 3,719 3' of the exon 7/intron 7 junction is shown. The human DNA sequence from clone RP3-469D22 on chromosome 1p35.1-36.13 containing the 5' part of the gene for *RHCE* (GenBank accession number AL031284) was taken as reference; numbers indicate the position in this sequence relative to the first base of intron 7 in the *RHCE* gene. The corresponding *RHD* gene sequence derives from GenBank accession number AL139426. Nucleotides indicating *RHD* or *RHCE* origin of the *Cde<sup>s</sup>* sequences are highlighted. A 666 bp DNA stretch comprising the breakpoint region of *Cde<sup>s</sup>* is indicated by asterisks.

**Figure 14.** Specific detection of *Cde<sup>s</sup>* by PCR-SSP. A PCR-SSP detecting the 3' breakpoint region of *Cde<sup>s</sup>* in intron 7 is shown. Both, a *RHD* negative sample (lane 1, ccddee) and a normal *RHD* positive sample (lane 2, ccD.EE) yield the 434 bp control product only, which is derived from the HGH gene. In contrast, a *Cde<sup>s</sup>* sample (CcddEe, lane 3) yields the 338 bp specific product, which is derived from the breakpoint region in intron 7, and in addition the 434 bp control fragment. This reaction is specific for *Cde<sup>s</sup>*; the two partial D phenotypes D<sup>IVa</sup> (lane 4) and D<sup>III</sup> type IV (lane 5) do not yield a specific product. The reaction also detects *Cde<sup>s</sup>* specifically, if *Cde<sup>s</sup>* occurs *in trans* to other *RHD* alleles, like in a *RHD<sup>ψ</sup>/Cde<sup>s</sup>* sample (lane 6).

**Figure 15.** *RHD* PCR-SSP for routine DNA typing. The PCR is performed as a modular system consisting of two multiplex reactions, an intron 4/exon 7 multiplex PCR-SSP (Panel A) and an intron 7 PCR enhanced by specific detection of *RHD(W16X)* and *RHD<sup>ψ</sup>* (Panel B). Results are shown for a normal D positive sample (lane 1), a normal D negative sample (lane 2), several rare D negative samples (lanes 3 to 6) and major D positive *RHD* variants (lanes 7 and 8). Standard D positive and D negative samples and D categories IV and VI are recognized in reaction A. *RHD-CE(8-9)-D* is detected in reaction B by the absence of the intron 7 band. The presence of *RHD(W16X)* and *RHD<sup>ψ</sup>* is also detected in reaction B. Band size is Panel A, control, 434 bp (HGH gene); intron 4, 226 bp; exon 7, 123 bp; Panel B, control, 659 bp (chromosome 1 genomic sequence about 90,000 bp 5' of *Rhesus box*); intron 7, 390 bp; *RHD(W16X)*, 248 bp; *RHD<sup>ψ</sup>*, 154 bp. The internal control amplicons, which were devised to be larger than the specific amplicons, may be suppressed because of competition, if a specific product is amplified.

**[0210]** The examples illustrate the invention:

#### Example 1: Blood samples and DNA isolation

**[0211]** EDTA- or citrate-anticoagulated blood samples were collected from white blood donors and characterized as D negative in routine typing including an antiglobulin test with anti-D (Wissenschaftlicher Beirat der Bundesärztekammer; Paul-Ehrlich-Institut. Richtlinien zur Blutgruppenbestimmung und Bluttransfusion (Hämotherapie). Köln: Deutscher Ärzte-Verlag; 1996; Wagner, Infusionsther Transfusionsmed 22:285-90, 1995). If necessary, samples were collected at random for specific CcEe phenotypes. A total of 314 ccddee, 433 Ccddee, 271 ccddEe, 19 CcddEe, 24 CCddee, 1 CcddEE and 6 ccddEE samples were tested. DNA was isolated by a modified salting-out procedure as described in Gassner et al., Transfusion 37; 1020, 1997.

#### Example 2: Molecular work-up

**[0212]** All samples were tested by PCR-SSP for the presence of four different *RHD* specific polymorphisms located in the *RHD* promoter, intron 4, exon 7 and the 3' untranslated region of exon 10. 48 samples with at least one positive PCR reaction were detected (Table 5). Those samples were further investigated for the presence of *RHD* specific polymorphisms in exon 3, exon 4, exon 5, exon 6, exon 7, intron 7 and exon 9. Twenty-six samples showed one of eight distinct PCR patterns involving a mixture of positive and negative reactions (Table 6). Twenty-two samples were positive for all *RHD* specific polymorphisms investigated and were assigned to eight *RHD* alleles by *RHD* specific sequencing

## EP 1 780 217 A1

of the ten *RHD* exons from genomic DNA (Table 7). For each PCR pattern and each *RHD* allele, one sample was serologically investigated. The phenotypes were determined to represent weak D, partial D, and D<sub>el</sub>, or confirmed as serologically D negative by adsorption/elution (Table 6 and 7).

### 5 Example 3: DNA database searches and analysis

[0213] The GenBank (<http://www.ncbi.nlm.nih.gov/BLAST/>) and the chromosome 1 database of the Sanger Center ([http://www.sanger.ac.uk/cgi-bin/nph-Blast\\_Server.html](http://www.sanger.ac.uk/cgi-bin/nph-Blast_Server.html)) were searched with cDNA sequences representative of *RHD* (RhXIII, accession number X63097) and *RHCE* (RhVI, X63095) using the BLAST program. The 84,810 bp genomic clone dJ469D22 (GenBank accession number AL031284), the 129,747 bp genomic clone dJ465N24 (GenBank accession number AL031432) and the 2,234 bp *SMP1* cDNA (GenBank accession number AF081282) were identified. dJ469D22 represented a major fragment of the *RHCE* gene, starting 33,340 bp 5' of the *RHCE* start codon and ending 1,142 bp 3' of exon 9. In dJ465N24, an internal stretch of 1,418 bp located between position 120,158 and 121,568 was 96% homologous to the 3' end of the *RHD* cDNA. The 3' end of the *SMP1* cDNA was complementary to the 3' end of the *RHCE* cDNA with an overlap of 58 bp.

### Example 4: PCR

[0214] If not mentioned otherwise, PCR reactions were done with 60°C annealing, 10 min extension at 68°C and denaturation at 92°C using the expand long template or the expand high fidelity PCR systems (Boehringer Mannheim, Mannheim, Germany) and the listed primers (Table 1). Three PCR reactions were used to bridge gaps in the 3' flanking regions of the *RH* genes. PCR 1 was done using primers rea7 and rend31 (PCR 2, rend32, sf1c; PCR 3, rea7, sf3). The structure of the 5' flanking regions was confirmed with PCR amplifications involving sense primers rend32, rey14a, rey15a and antisense primers re011d and re014. Intron 9 size was estimated to be about 9,000 bp based on PCR amplifications using rb10b and rr4 for *RHD* (re96 and rh7 for *RHCE*).

### Example 5: Nucleotide sequencing

[0215] Nucleotide sequencing was performed with a DNA sequencing unit (Prism BigDye terminator cycle-sequencing ready reaction kit; ABI 373A, Applied Biosystems, Weiterstadt, Germany).

### Example 6: Characterizing the *RH* gene locus

[0216] A physical structure of the *RH* genes' locus was derived (Fig. 1). This structure was deduced from the following considerations: (i) **3' flanking regions**. The 3' flanking region of *RHD* was highly homologous to the 3' part of dJ465N24 (Fig. 1 B, region c). This homology continued beyond the end of the *RHD* cDNA and extended for at least 8,000 bp as proven by the fact that it was possible to obtain PCR amplicons (Fig. 1B, PCR 1). Sequences homologous to the 3' part of dJ465N24 were neighboring to the 5' region of the *SMP1* gene (Fig. 1B; PCR 2). The 3' end of the *SMP1* gene occurred immediately adjacent to the *RHCE* gene as indicated by the complementarity of the 3' ends of the respective cDNAs and confirmed by PCR (Fig. 1 B, PCR 3). Further details of the *RHD* 3' flanking region (*Rhesus box*) and the *SMP1* gene are described below. (ii) **5' flanking regions**. dJ469D22 comprised 33,340 bp 5' flanking region of *RHCE*. For *RHD*, a 466 bp homology between the 3' end of dJ465N24 and dJ469D22 indicated that dJ465N24 might represent the 5' flanking sequence of *RHD*. This assumption was proven by PCR (Fig. 2). (iii) **Analysis of YAC 38A-A10**. DNA from the YAC 38A-A10 (UK HGMP resource centre, Cambridge, UK) was isolated after a single growth phase by standard methods ([http://hdktab.wustl.edu/lab\\_manual/yeast](http://hdktab.wustl.edu/lab_manual/yeast)). It was confirmed that this YAC contained *RH* DNA. Furthermore, shotgun cloning experiments indicated that some of its insert probably derived from the X chromosome (data not shown). This YAC had been known to contain *RHCE* exons 2 to 10 and *RHD* exons 1 to 10 (Carritt, Hum. Mol. Genet. 6:843, 1997) and was thus expected to contain the DNA segments interspersed between *RHD* and *RHCE*. The presence of DNA segments representative of different parts of the *RH* locus in this YAC was observed (Table 2). The results were concordant with the proposed structure of the *RH* locus shown in Fig. 1, Panel A.

### Example 7: Identification of *RHD* specific sequences in the *RHD* promoter

[0217] About 2,000 bp *RHD* promoter sequence was established by chromosomal walking (GenomeWalker kit, Clontech, Heidelberg, Germany). D-positive and D-negative samples were amplified using primers re04 and re11d (Table 1) and *RHD*- and *RHCE*-specific sequences established for 1,200 bp 5' of the start codon by sequencing with internal primers. A short deletion in the *RHD* gene was identified and used to develop the *RHD*-specific primer re011d. The 1,200 bp sequence including the *RHD* promoter has been deposited at EMBL under accession no. AJ252314.

**Example 8: Characterization of *Rhesus boxes***

[0218] Two DNA segments of about 9,000 bp, located 5' and 3' of the *RHD* gene, were highly homologous, had identical orientation, and were designated "*Rhesus boxes*" (Fig. 4). The *Rhesus boxes* were amplified and sequenced using internal primers in two overlapping fragments using PCR primer pairs rez4/rend31 and rend32/re011d (upstream *Rhesus box*), rea7/rend31 and rend32/sr9 (downstream *Rhesus box*), and rez4/rend31 and rend32/sr9 (hybrid *Rhesus box* of *RHD*-negative). The upstream *Rhesus box* (5' of *RHD*) was about 9,142 bp long and ended about 4,900 bp 5' of the *RHD* start codon. The downstream *Rhesus box* (3' of *RHD*) was 9,145 bp long and started 104 bp after the *RHD* stop codon. The *Rhesus boxes* exactly embraced the part of *RHD* with homology to *RHCE*. The central portion of both *Rhesus boxes* contained an almost complete remnant of a transposon-like human element (THE-1B). The single open reading frame usually found in the THE-1B element was, however, abolished due to several nucleotide aberrations occurring in both *Rhesus boxes* in parallel, including a nonsense mutation in codon 4. While there was overall 98.6% homology between both *Rhesus boxes*, a 1,463 bp "identity region" located between positions 5,701 and 7,163 was completely identical with the single exception of a 4 bp T insertion in a poly T tract.

**Example 9: Evaluation of the genomic structure of *SMP1***

[0219] The genomic structure of the *SMP1* gene was evaluated by PCR using internal primers and nucleotide sequencing (Fig. 3). The sizes of the *SMP1* introns were estimated by PCR amplicons obtained with primers rend32, sr9, sf1c, sf1, sm19, sr45, sr47, sr47c, sr5, sr5c, sr55, sr55c, sr3, sr3kp, rea7. The positions of the intron/exon junctions and the absence of additional introns were determined by nucleotide sequencing. Six introns could be identified. Exon 1 contained 5' untranslated sequences only and was separated from the *Rhesus box* by 15 bp. The long 3' untranslated sequence of exon 7 overlapped with *RHCE* exon 10. The total gene size was estimated to be 20,000 bp resulting, in conjunction with the downstream *Rhesus box*, in a distance between *RHD* and *RHCE* of about 30,000 bp (Fig. 1).

**Example 10: Localization of the *RHD* gene deletion in the *RHD* negative haplotypes**

[0220] It was reasoned that the homology of the two *Rhesus boxes* may have been instrumental for the mechanism of the *RHD* deletion in the common *RHD* negative haplotypes. The nucleotide sequence of the *Rhesus box* in *RHD* negative DNA was determined (Fig. 5). The single *Rhesus box* detected in *RHD* negatives had a hybrid structure. The 5' end of this *Rhesus box* represented a upstream *Rhesus box*, the 3' end a downstream *Rhesus box*. It was determined that the 903 bp breakpoint region of the *RHD* deletion was located in the identity region of the *Rhesus boxes* (Fig. 4, arrow pointing to left).

**Example 11: Specific detection of the *RHD* deletion by PCR**

[0221] Two PCR based methods were developed for specific detection of the *RHD* gene deletion occurring in the prevalent *RHD* negative haplotypes (Fig. 6). Long-range PCR-SSP was performed using the expand long template PCR system with buffer 3 and primers rez4 (5' of upstream *Rhesus box*) and sr9 (*SMP1* exon 1). Annealing was at 60°C and extension 20 min at 68°C. PCR amplicons were resolved using a 1 % agarose gel. PCR-RFLP was performed using the expand high fidelity PCR system and primers rez7 (non-specific, 5' of *Rhesus box* identity region) and mb31 (specific for downstream *Rhesus box*, 3' of *Rhesus box* identity region). Annealing was at 65°C and extension 10 min at 68°C. PCR amplicons were digested with *Pst*I for 3 hrs at 37°C and fragments resolved using a 1 % agarose gel. These techniques allowed the ready and direct detection of the common *RHD* negative haplotypes, even if they are *in trans* to *RHD* positive haplotypes. PCR-RFLP was further applied to a larger number of samples (Table 3). As expected, all 33 samples with known genotype were correctly typed. In 68 additional samples representative of the most common phenotypes, the results were consistent with the known haplotype frequencies in the population.

**Example 12: *RHD* PCR-SSP**

[0222] The PCR-SSP reactions (Table 4) were adapted and extended from a previously described *RHD* exon specific PCR-SSP method (Gassner, Transfusion 37:1020-6, 1997) and were triggered to work under identical thermocycling conditions. Concentrations of specific primers were 0.2 µM for all reactions with the exception of exon 6 (0.1 µM), intron 7 (0.4 µM) and exon 9 (0.4 µM). For most samples intron 4/exon 7 was tested as multiplex reaction containing 0.2 µM of exon 7 (primer set ga71/ga72) and 0.1 µM of intron 4 primers. Each reaction contained a set of HGH primers (Gassner, Transfusion 37:1020-6, 1997) as an internal control in concentrations of 0.05 µM for promoter, intron 4, and exon 7 with ga71/ga72; 0.075 µM for exon 10; 0.1 µM for intron 7; 0.15 µM for exon 3, exon 4, exon 7 with rb26/re71, and exon 9; 0.2 µM for exon 5 and exon 6. Mg<sup>2+</sup> concentration was 0.4 µM for intron 7 and for all other reactions 0.15 µM. For exon

6, 20 % solution Q (Qiagen, Hilden, Germany) was added.

#### Example 13. Improved *RHD* PCR-SSP for routine DNA typing.

5 **[0223]** Based on the alleles detected in this study and described previously, we devised an improved *RHD* PCR-SSP for routine DNA typing that included the specific detection of *RHD*<sup>Ψ</sup> and alleles detected in this study, like *RHD*(W16X) in a single PCR tube. Reaction A contained primers ga71 and ga72 at 0.3 μM, rb12 and re41 at 0.1 μM, and HGH primers at 0.1 μM. Mg<sup>2+</sup> was at 0.175 μM. Reaction B contained primers RhPsiF and RhPsiB at 0.5 μM, re11d and RhX1f1 at 0.3 μM, re721 and rb9 at 0.2 μM and as control primers rend9b1 and rend 9b2 at 0.2 μM. Primer sequences were ga71, gttgtaaccgagtgctgggattc; ga72, tgccggctccgacggatc; rb12, tcctgaacctgctctggaagtgc; re41, cgataccagttt-  
10 gtctgccatgc; RhPsiF, agacagactaccacatgaacttac; RhPsiB, tctgatcttatctccgttcctc; re11d, agaagatgggggaatcttttct; RhX1f1, cgctgcctgccctctga; re721, ctggaggctctgagaggttgag; rb9, aagctgagttcccaatgctgagg; rend9b1, cactgcactt-  
ggcaccattgag; rend9b2, ttccgaaggctgctttccc.

15 **[0224]** The PCR reactions could be performed in two tubes (Fig. 15), tested five polymorphisms and were expected to have a false-positive rate of less than 1:10,000 (Table 11).

#### Example 14: PCR reactions for *Cde*<sup>s</sup>

20 **[0225]** A hybrid exon 3 with a N152T substitution occurring in the *Cde*<sup>s</sup> haplotype was detected by a PCR-SSP reaction using specific primers Rh152Tb and ga31 at 0.3 μM. The L245V substitution observed in *Cde*<sup>s</sup> was detected with specific primers Rh223Vf and Rh245Vb at 0.2 μM. HGH primer concentrations were 0.1 μM. The other PCR conditions were identical as described in the previous paragraph. Primers sequences were Rh152Tb, gatattactgatgaccatctcatgg; Rh223Vf, ttgtgatgttctggccaagtg; and Rh245Vb, gctgtcaccactctgactgctac. The *Cde*<sup>s</sup> haplotype, that is frequent in Africans (Faas, Transfusion 37:38-44, 1997; Singleton, Blood 95:12-8, 2000), possesses a hybrid exon 3 harboring the *RHCE* specific N152T substitution (Faas, Transfusion 37:38-44, 1997). This hybrid exon is expected to be typed as *RHD* positive by the *RHD* exon 3 specific PCR that detected an A at position 383 (codon 128) and was used in the population survey. Since pattern 4 and pattern 8 were compatible with the known data about the *Cde*<sup>s</sup> haplotype, the presence of a hybrid exon 3 was evaluated in the two samples by sequencing the 3' part of exon 3 and by a PCR-SSP specific for an exon 3 hybrid indicative of *Cde*<sup>s</sup>. The pattern 4 sample possessed a normal *RHD* exon 3, while the pattern 8 sample had a  
25 hybrid exon 3 as predicted for a *Cde*<sup>s</sup> haplotype. Also, the T at position 410 (A137V substitution) typical for the *Cde*<sup>s</sup> haplotype (Daniels, Transfusion 38:951-8, 1998) and also present in *D category III type IV* was detected. The identity of pattern 8 and *Cde*<sup>s</sup> was further corroborated by a PCR-SSP detecting G at position 733 (L245V substitution).

#### Example 15: 5' breakpoint region of *Cde*<sup>s</sup> in intron 3.

35 **[0226]** Based on its cDNA, *Cde*<sup>s</sup> had been characterized as an *RHD-CE(3-7)-D* hybrid gene, in which the 5' part of exon 3 derived from *RHD* and the 3' part of exon 3 including codon 152 derived from *RHCE*. We noted that a similar hybrid exon 3 with a N152T substitution was found in *D category III type IV* (Wagner, F.F., Frohmajer, A., Ladewig, B., Eicher, N.I., Lonicer, C.B., Müller, T.H., Siegel, M.H., and Flegel, W.A. Weak D alleles express distinct phenotypes. Blood 95:2699-2708, 2000) and in *D category IVa* (Rouillac, C., Colin, Y., Hughes-Jones, N.C., Beolet, M., D'Ambrosio, A.-M., Cartron, J.P., and Le Van Kim, C. Transcript analysis of *D category* phenotypes predicts hybrid Rh D-CE-D proteins associated with alteration of D epitopes. Blood 85:2937-2944, 1995), two aberrant *RHD* alleles in which exons 4 to 7 derived from *RHD*. We reasoned that the N152T substitution might have antedated the substitution of *RHD* exons 4 o 7 in *Cde*<sup>s</sup>. In this case, the 5' breakpoint region was expected to be located in intron 3 rather than exon 3 as predicted from the cDNA. We hence evaluated the presence of *RHD* specific polymorphisms in *Cde*<sup>s</sup> intron 3.

40 **[0227]** To evaluate the presence of the *EcoRV*-site at nucleotide position 752 (*RHD* specific) and 2872 (*RHCE* specific) and of the *PvuII*-site at nucleotide position 1777 (*RHCE* specific), the 5' part of intron 3 of *RHD* and *RHCE* was amplified using primers rb3 and rb33 and digested with *EcoRV* or *PvuII*. To evaluate the presence of the *SacI*-site at nucleotide position 7797 (*RHCE* specific) and of the *Alw44I*-site at nucleotide position 8550 (*RHD* specific), the 3' part of intron 3 of *RHD* and *RHCE* was amplified using primers rb34 and rb5 and digested with *SacI* or *Alw44I*. Primer sequences were rb3, aaggtcaacttgccgagttggtg; rb33, gtgagactgagttctgtattctg; rb34, ccagaatacagaactcagttctcac; rb5, ggcagacaaact-  
45 gggtatcgttgc.

50 **[0228]** The PCR-RFLP analysis of these intron 3 polymorphisms indicated that *RHD* specific sequences were present at least up to intron 3 position 2872. To further determine the 5' breakpoint region of *Cde*<sup>s</sup>, we sequenced a DNA stretch encompassing the breakpoint region. DNA was amplified using primers rb3 and re37 and sequenced using primers rb33, rb34 and Cdesf1. Primer sequences were re37, gggttaaagtcacatacacagatg; Cdesf1, atacagaactcagttctcaacttag. We determined that the breakpoint region was located in intron 3 as shown in Figure 12.

**Example 16: 3' breakpoint region of *Cde<sup>s</sup>* in intron 7.**

[0229] To determine the 3' breakpoint region of *Cde<sup>s</sup>* in intron 7, we sequenced parts of intron 7. DNA was amplified using sense primers rb8, re77 and rex1 and antisense primers rb51 and re711b. Primers rb43, rex19c, cdes7b2, and cdes7f2 were used for nucleotide sequencing. Primer sequences were rb8, gtgttgtaaccgagtctgggg; re77, tctccacagctc-  
 catcatggg; rex1, ggctgtaaaaatggctgaagcag; rb51, gcatgacgtgttctgcctcttg; re711b, ctatcagcattctgatctcaacg; rb43, gaat-  
 agcagagaaaaacctcagactgcc; rex19c, gctccattcttgacaatacaggc; cdes7b2, gcttatactataaagtgggtttttgg; cdes7f2, gtttgaatc-  
 ccaagagccactcat. We established the breakpoint region as shown in Fig. 13. The structure of the 3' breakpoint region  
 was intriguing, because there were multiple switches between *RHCE* and *RHD* specific sequences. Those features are  
 unusual for a breakpoint region and may be used for specific diagnosis of *Cde<sup>s</sup>*. They may indicate that the parental  
 alleles differed from the standard *RHCE* and *RHD* sequences or that after the major gene conversion, additional small  
 gene conversions were introduced.

**Example 17: A PCR-SSP to specifically detect *Cde<sup>s</sup>*.**

[0230] Usually, the presence of *Cde<sup>s</sup>* is identified by the *RHD-CE-D* hybrid pattern in an *RHD* exon specific PCR. Such an approach does not allow the specific detection of the D negative *Cde<sup>s</sup>* haplotype, if an *RHD* positive haplotype occurs *in trans*. Since *Cde<sup>s</sup>* does not contain a hybrid *Rhesus box*, a *RHO/Cde<sup>s</sup>* heterozygous person is likely mistyped as *RHD<sup>+</sup>/RHD<sup>+</sup>* homozygous. There are several distinct features of *Cde<sup>s</sup>* in the promoter, intron 2, exon 2, and exon 3 that might be used for a specific detection. These features are, however, shared by the D positive alleles D category IVa and partially by D category III type IV, which would hence confound such methods of detection.

[0231] Based on the *Cde<sup>s</sup>* specific DNA sequence in intron 7, we developed a PCR-SSP that specifically detected *Cde<sup>s</sup>*. The 3' breakpoint region of *Cde<sup>s</sup>* in intron 7 was detected by PCR-SSP using specific primers Cdes7f2 and Cdes7b2 at 0.4  $\mu$ M and HGH control primers at 0.15  $\mu$ M. The other PCR conditions were identical as described in example 12. Primer sequences were Cdes7f2, gtttgaatccaagagccactcat; Cdes7b2, gcttatactataaagtgggtttttgg. We obtained a specific product with the index *Cde<sup>s</sup>* sample (Figure 14), two additional *Cde<sup>s</sup>* samples and a *RHD<sup>+</sup>/Cde<sup>s</sup>* heterozygous sample (Figure 14). Normal *RHD* positive and *RHD* negative samples as well as samples of D category III type IV and of D category IVa did not result in a specific PCR product (Figure 14). We concluded that our PCR-SSP method allowed a specific detection of *Cde<sup>s</sup>*, even if it occurred *in trans* to another *RHD* positive allele. Furthermore, the detection method was not confounded by D category III type IV or D category IVa that shared the N152T substitution with *Cde<sup>s</sup>*. It should be noted that the latter haplotypes are frequent in populations comprising African ethnic background, in which *Cde<sup>s</sup>* is prevalent. The method described by us in this example allowed the specific detection of *Cde<sup>s</sup>*, is not confounded by the other haplotypes and hence represents a considerable improvement to the prior art. Our characterization of the 5' breakpoint region (example 15) will likewise allow the specific detection of *Cde<sup>s</sup>* by any suitable method known in the art, like PCR-SSP, PCR-LP, PCR-RFLP, PCR-SSO, Southern blotting etc.

[0232] The specific detection of *Cde<sup>s</sup>* is also important for the correct prediction of the antigen C. The *RHD* gene of *Cde<sup>s</sup>* encodes for an antigen C that is often missed in DNA based methods for the prediction of antigen C.

**Example 18: Immunohematology**

[0233] One sample of each *RHD* positive allele was evaluated by direct agglutination with two monoclonal anti-D (Seraclon anti-D, clone BS226; Biotest, Dreieich, Germany, and Frekaklon anti-D, clone MS201; Gull, Bad Homburg, Germany). Indirect antiglobulin test was done in a gel matrix test (LISS-Coombs 37 °C, DiaMed-ID Micro Typing System, DiaMed, Cressier sur Morat, Switzerland) using an oligoclonal anti-D (Seraclon anti-D blend, clones H41 11B7, BS221 and BS232; Biotest). Samples reactive in gel matrix technique were further investigated using the monoclonal anti-D HM10, HM16, P3x61, P3x35, P3x212 11F1, P3x212 23B10, P3x241, P3x249, P3x290 (Diagast, Loos, France) and H41 11 B7 (Biotest). The presence of a D<sub>el</sub> phenotype was determined by adsorption of 500  $\mu$ l of a polyclonal anti-D (Human incomplete anti-D; Lorne Laboratories, Reading, UK) to 500  $\mu$ l red cells for 1 h at 37 °C and elution using a chloroform technique (Flegel, Transfusion 40:428-434, 2000). The analysis of samples routinely grouped as D negative revealed 16 D<sub>el</sub> samples and 3 D positive samples with weak or partial D. These samples clustered among samples previously believed to be D negative with a C or E (Table 9). Nineteen of twenty-seven discrepancies between routine serology and a PCR testing intron 4 and exon 7 represented D positive samples missed by serology, only eight were due to false-positive PCR.

**Example 19: Haplotype frequencies**

[0234] For alleles observed more than once, their haplotype association was trivial. Alleles that were observed only once were assumed to be associated with the Cde or cdE haplotype rather than the cde haplotype, because no *RHD*

positive allele was detected in any ccddee sample. An allele occurring in a single CcddEe sample was formally counted half for Cde and half for cdE. CCddee samples were assumed to harbour one aberrant and one normal Cde allele. The frequency of a given aberrant *RHD* allele in its haplotype was calculated as the number of observed samples divided by the number of the corresponding haplotypes under observation (500 Cde, 302 cdE). The population frequency of an *RHD* allele was calculated from the frequency of this allele in its haplotype and the known frequency of the haplotype in the local population (Wagner, Infusionsther. Transfusionsmed. 22:285-90, 1995). The haplotype frequencies were calculated for each PCR pattern and for each *RHD* allele (Table 8). In accordance with a previous study in England by Avent et al. (Avent, Blood 89:2568-77, 1997), 4.9% of Cde haplotypes and 1.5% of cdE haplotypes were *RHD* positive in our population. As no *RHD* positive allele was detected among 314 ccddee samples, the frequency in the cde haplotype was less than 0.5 % (upper limit of one-sided 95% confidence interval, Poisson distribution). The three frequencies differed statistically significantly from each other ( $p < 0.05$ ; two sided Fisher's exact test for each pairwise comparison corrected according to Bonferoni-Holm). The population frequency of any D negative *RHD* positive haplotype was estimated to be 1:1,606.  $D_{el}$  alleles could only be observed in the presumed Cde haplotypes. About 3% of samples carrying antigen C that were typed D-negative in the blood bank routine represented  $D_{el}$ . The population frequency of  $D_{el}$  alleles was 1:3,030.

Table 1. Primers

	Primer	Nucleotide sequence	Localization	Position
20	rb10b	ggctaaatatttgatgaccaagtt	<i>RHD</i> cDNA	1,194 to 1,217
	re011d	gcagccaactcccctgtg	<i>RHD</i> promoter	-883 to -901
	re014	gctctaccttggtcacctcc	dJ469D22	52,189 to 52,209
	re04	aggtcacatccattatcccactg	dJ469D22	53,968 to 53,945
	re11d	agaagatgggggaatcttttct	dJ469D22	51,193 to 51,216
25	re96	ttgtgactgggctagaagaaggtg	dJ469D22	242 to 216
	rea7	tgtgcctgcattgtacgtgag	<i>RHD</i> cDNA	1,311 to 1,333
	rend31	ttctgtctgggtggggagg	dJ465N24	128,649 to 128,629
	rend32	ggaggggttaatatgggtggc	dJ465N24	127,355 to 127,375
30	rend8b1	ttgtcctgggtgcctgtggtc	dJ465N24	69,296 to 69,274
	rend8b2	caaatcctgtgactggtctcgg	dJ465N24	68,451 to 68,473
	rend9a1	aacggctccatcaccctaag	dJ465N24	50,008 to 49,987
	rend9a2	cccactcctagataccaaccaag	dJ465N24	49,059 to 49,083
	rey14a	ctttatgcactgcctcgtgaatc	dJ469D22	56,792 to 56,769
35	rey14b	ttgactggtgtggtgctgttg	dJ469D22	55,863 to 55,884
	rey15a	gcagaaagggaggtgatgctg	dJ469D22	55,416 to 55,395
	rey7	ctgacaaagttgagagcccactg	dJ469D22	62,324 to 62,346
	rey8	ttaagcctacatccacatgctgag	dJ469D22	62,854 to 62,831
40	rez2	ccttggtctgcagaatttca	<i>RHD</i> cDNA	2738 to 2717
	rez4	gtttggcatcataggagattggc	dJ465N24	120,101 to 120,124
	rez7	cctgtccccatgattcagttacc	dJ465N24	124,831 to 124,854
	rh7	acgtacaaatgcaggcaac	<i>RHD</i> cDNA	1,330 to 1,312
	rnb31	ccttttttgtttttggcgggtgc	downstream <i>Rhesus</i> box	6,710 to 6,684
45	rr4	agcttactggatgaccacca	<i>RHD</i> cDNA	1,541 to 1,522
	sf1	gactgggggaaaagcgcaatac	<i>SMP1</i> cDNA	142 to 164
	sf1c	gtattgcgctttccccccagtc	<i>SMP1</i> cDNA	164 to 142
	sf3	tgactgtctctatcccacatg	<i>SMP1</i> cDNA	1,696 to 1,717
	sm19	gggctgaagcaagtaaatggaag	<i>SMP1</i> intron 1	- 58 to -35
50	sr1	gctatcaatatttctgttacagacac	<i>SMP1</i> cDNA	2,172 to 2,144
	sr3	gttactgccataagtcttctcagtc	<i>SMP1</i> cDNA	575 to 551
	sr3kp	tggccgactgaagacttatgg	<i>SMP1</i> cDNA	546 to 567
	sr45	cagctgcatctatgataatccacc	<i>SMP1</i> cDNA	224 to 243
55	sr47	atggacaagtcagggtgatag	<i>SMP1</i> cDNA	315 to 344
	sr47c	atcacctcggactgtccattc	<i>SMP1</i> cDNA	342 to 321
	sr5	gcaatcagagatccaaaggccaac	<i>SMP1</i> cDNA	428 to 405

EP 1 780 217 A1

(continued)

Primer	Nucleotide sequence	Localization	Position
sr5c	gttggccttggatctctgattgc	<i>SMP1</i> cDNA	405 to 428
sr55	gacatagataccctggaattgctgt	<i>SMP1</i> cDNA	472 to 497
sr55c	acagcaattccagggtatactatgctc	<i>SMP1</i> cDNA	497 to 472
sr9	ctccccgatttagccaagaa	<i>SMP1</i> cDNA	27 to 6

For the *RHD* promoter and the *RHD* cDNA, the positions refer to the distance from the A of the start codon. For introns, they refer to the distance from the intron/exon junction. For all other sequences including the *SMP1* cDNA, they refer to the distance from the start of the published sequences. The mismatches in primers rey14b, rnb31, and sf3 were inadvertently introduced. Primers re11d, re014 and re04 do not exactly match dJ469D22, because they were designed from our raw sequences covering the 5' flanking region of *RHD*.

Table 2. Presence of *RHD* flanking sequences in the YAC 38A-A10

Primer	Amplicons obtained with				
sense	antisense Predicted position	Amplicon size	RHD <sup>+</sup>	RHD <sup>-</sup>	
rend9a1	rend9a2 <i>RHD</i> 5' flanking region yes	about 85,000 bp from ATG	948 bp	yes	yes
rend8b1	rend8b2 <i>RHD</i> 5' flanking region yes	about 50,000 bp from ATG	845 bp	yes	yes
rea7	rez2 <i>RHD</i> 3' flanking region yes	about 1,500 bp from STOP	1,412 bp	yes	no
rend32	sr9 <i>RHCE</i> 3' flanking region yes	about 20,000 bp from STOP	1,989 bp	yes	yes
sr1	sf3 <i>RHCE</i> 3' flanking region yes	about 1,000 bp from STOP	477 bp	yes	yes
rey14b	rey14a <i>RHCE</i> 5' flanking region no	about 5,300 bp from ATG	929 bp	yes	yes
rey7	rey8 <i>RHCE</i> 5' flanking region no	about 10,000 bp from ATG	530 bp	yes	yes

Table 3. PCR-RFLP for the specific detection of the *RHD* deletion

Phenotype	Known genotype	Samples tested (n)	Number of samples with <i>RHD</i> genotype						p¶
			determined			expected			
			+/+	+/-	-/-	+/+	+/-	-/-	
<i>Known genotype</i>									
ccddee	cde/cde	14	0	0	14	0	0	14	N.A.
CCddee	Cde/Cde <sup>†</sup>	5	0	0	5	0	0	5	N.A.
ccddEE	cdE/cdE <sup>†</sup>	1	0	0	1	0	0	1	N.A.
D variants	D/cde <sup>‡</sup>	9	0	9	0	0	9	0	N.A.
ccDEe	cDe/cDE <sup>§</sup>	4	4	0	0	4	0	0	N.A.
<i>Common phenotypes</i>									
CcDee		10	1	9	0	0.5	9.5	0	>0.4
ccDEe		10	0	10	0	0.3	9.7	0	>0.5
ccDee		10	1	9	0	0.5	9.5	0	>0.4
CCDee		10	9	1	0	9.5	0.5	0	>0.4

EP 1 780 217 A1

(continued)

Common phenotypes

5	CcDEe	12	11	1	0	11	1	0	>0.5
	ccDEE	10	10	0	0	9.2	0.8	0	>0.4
	CCDEe	6	5	1	0	5.8	0.2	0	>0.1

\* Expected number of *RHD*<sup>+</sup>/*RHD*<sup>+</sup> and *RHD*<sup>+</sup>/*RHD*<sup>-</sup> samples based on known genotypes or the haplotype frequencies in the local population<sup>41</sup>

† *RHD*-negative in PCR.

‡ *RHD*<sup>+</sup>/*RHD*<sup>-</sup>, because a weak or partial D phenotype would be masked in a *RHD*<sup>+</sup>/*RHD*<sup>+</sup> genotype. These samples were weak D type 1 (n=2), type 2 (n=2), type 3 (n=2), type 4 (n=2) and D<sup>VII</sup> (n=1).

§ Presence of two *RHD* genes differing in their polymorphic *Hae*III-site in intron 3<sup>42</sup> demonstrated by PCR-RFLP.

N.A. - not applicable. Probabilities were calculated based on confidence limits of binomial distribution.

Table 4. *RHD* PCR-SSP

Region	Name Reference	DNA sequence	Position	Polymorphisms detected	Amplicon size
Promoter	re012	tccactttccacctccctgc	Promoter -1,137 to -1,119	7 bp deletion at -1125	255 43
Promoter	re011d	gcagccaactcccctgtg	Promoter -883 to -901	4 bp deletion at -896	44
Exon 3	ga31 (D-3-383)	ttgtcgggtgctgatctcagtgga	Exon 3 361 to 383	383 A	154 21
Intron 3	rb21	aggtcctcctccagcac	Intron 3 28 to 11		42
Exon 4	ga41 (D-4-527)	acatgatgcacatctactcgtgtcgc	Exon 4 503 to 527		123 21
Exon 4	ga42 (D-4-602)	cagacaaactgggtatcgttgctg	Exon 4 625 to 602	602 C	21
Intron 4	re41 this study	cgatacccagttgtctgccaatgc	Exon 4 608 to 631		226
Intron 4	rb12	tcctgaacctgctctgtgaagtgc	Intron 4 198 to 175	Intron 4 deletion in <i>RHD</i>	40
Intron 4	rb24	agaccttggagcaggagtg	Intron 4 -53 to -34		228 40
Intron 5/Exon 5	ga51 (D-5-787)	ctgctcaccttgctgatcttccc	Intron 5/Exon 5	8 to 787	787 G 21
Exon 6	ga62 (D-6-826)	ttatgtgcacagtcggtgttgg	Exon 6 804 to 826		133 21
Exon 6	ga61 (D-6-916)	caggctactggctcccccgac	Exon 6 936 to 916	916 G	21
Exon 7†	ga71 (D-7-967)	gttgaaccgagtgctggggattc	Exon 7 944 to 967		123 21
Exon 7	ga72 (D-7-1048)	tgccggctccgacggtatc	Exon 7 1,066 to 1,048	1,048 G	21
Intron 6	rb26	aggggtgggtaggaatatg	Intron 6 -62 to -43		130 42
Exon 7	re71	accagcaagctgaagttgtagcc	Exon 7 1,008 to 985	985/986 GG	42

EP 1 780 217 A1

(continued)

Region	Name Reference	DNA sequence	Position	Polymorphisms detected	Amplicon size			
5	Intron 7	rb52	ccaggttgtaagcattgctgtacc	Intron 7	6,666 to 6,690 C	42		
		rb51 this study	gcatgacgtgtctgcctcttg	Intron 7	6,734 to 6,713 C	169		
10	Exon 9	re83	gagattaaaaatcctgtgtccta	Intron 8	-56 to -34	119	42	
		re94 this study	cttggtcatcaaaaatttagcct	Exon 9	1,216 to 1,193 A			
15	Exon 10 (3' UTR)	rea7	tgttgctgcattgtacgtgag	3'UTR	1,310 to 1,333	<i>RHD/Rhesus box</i>	23	44
		rr4	agcttactggatgaccacca	3'UTR	1,541 to 1,522	junction		42

\* Primer names in brackets are as described by Gassner et al. 21.

† Primer set ga71/ga72 was used for the screening, primer set rb26-re71 for *RHD* exon specific PCR-SSP.

**Table 5.** Population survey of known D negative blood donors screened by *RHD*PCR-SSP

Documented phenotype	Samples (n)	
	screened	PCR-SSP positive*
ccddee	314	0
Ccddee	433	34
ccddEe	271	5
CCddee	24	4
CcddEe	19	4
ccddEE	6	1
CcddEE	1	0
Total	1,068	48

Positive for at least one of four *RHD* specific polymorphisms tested (promoter, intron 4, exon 7 or 3' UTR).

**Table 6.** PCR patterns compatible with *RHD-RHCE-RHD* hybrid genes or partial *RHD* deletions in 25 D negative samples

PCR pattern	<i>RHD</i> specific PCR-SSP* P E3E414 E5E6E717 E9E10	Samples Possible cause†	Phenotype (n)	Documented		Haplotype	Reference‡
				Documented	Confirmed		
Pattern 1 Africans 24	+ - - - - - - +	<i>RHD-CE</i> (3-9)-D	11	Ccddee§	D negative	Cde	Whites 1,25,

EP 1 780 217 A1

(continued)

5	PCR pattern	<i>RHD</i> specific PCR-SSP*	Samples Possible cause†	Phenotype (n)	Phenotype		Haplotype	Reference‡
		P E3E414 E5E6E717 E9E10			Documented	Confirmed		
10	Pattern 2	+ - - - - - + +	<i>RHD-CE</i> (3-7)- <i>D</i>	4	Ccddee	D negative	Cde	this study
	Pattern 3	+ + - - - - + +	<i>RHD-CE</i> (4-7)- <i>D</i>	3	ccddEe	D negative	cdE	Whites 16
	Pattern 4	+ + - - - + + +	<i>RHD-CE</i> (4-7)- <i>D</i>	1	CcddEe	D negative	n.k. ¶	this study
15	Pattern 5	+ + - - + + + + +	<i>RHD-CE</i> (4-5)- <i>D</i>	2	ccddEe <sup>α</sup>	partial D/D <sub>el</sub> <sup>α</sup>	cDE	Whites 3,22,30,40
	Pattern 6	+ + + + + + + - - +	<i>RHD-CE</i> (8-9)- <i>D</i>	3	CCddee	D negative	Cde	Whites 21
20	Pattern 7	- - - - - +	<i>RHCE</i> (1-9)- <i>D</i> (10)	1	ccddEe	D negative	cdE	this study
25	Pattern 8	- + - - - + + +	<i>RHD</i> (1-3)- <i>CE</i> (4-7)- <i>D</i>	1	CcddEe	D negative	Cde <sup>β</sup>	Africans 5,15

\* P - Promoter, E3 - Exon 3, E4 - Exon 4, 14 - Intron 4, E5 - Exon 5, E6 - Exon 6; E7 - Exon 7; 17 - Intron 7; E9 - Exon 9; E10 - Exon 10 (3' UTR)

† Assuming the presence of a single *RHD-CE-D* hybrid allele.

‡ Previously described alleles that fit PCR pattern and haplotype.

§ 11 samples: 9 Ccddee, 1 CCddee, 1 CcddEe

¶ n.k. - not known.

α 2 samples, 1 labeled CcddEe with D<sub>el</sub> phenotype, 1 labeled ccddEe with partial D D<sup>VI</sup> phenotype.

β Probably identical to Cde<sup>8</sup> (see below).

**Table 7.** *RHD* alleles with single nucleotide substitutions in 22 D negative samples

Allele	Substitution Reference	Effect(s)	Samples (n)	Phenotype		Haplotype	
				Documented	Confirmed		
40 <i>RHD</i> (W16X) study	G->A at 48	Stop codon at codon 16	2	Ccddee	D negative	Cde	this
<i>RHD</i> (G486 (+1)A) study	g->a at 486+1	5' splice site intron 3 ACgt->ACat	3	Ccddee	D <sub>el</sub>	CDe	this
45 <i>RHD</i> (G212V) study	G->T at 635	3' splice site intron 4 agGC->agTC	1	Ccddee	D negative	Cde	this
50 <i>RHD</i> (C285Y) study <sup>1</sup>	G->A at 854	Missense mutation G212V Missense mutation C285Y	1	ccddEe	partial D	cDE	this
55 <i>RHD</i> (M295I)	G->T at 885	Missense mutation M295I	7	Ccddee	D <sub>el</sub>	CDe <sup>†</sup>	42

EP 1 780 217 A1

(continued)

	Allele	Substitution Reference	Effect(s)	Samples (n)	Phenotype		Haplotype	
					Documented	Confirmed		
5	RHD(Y330X) study	C->G at 985	Stop codon at codon 330	1	Ccddee	D negative	Cde	this
	RHD(G1153(+1)A) study	g->a at 1153+1	5' splice site intron 8 AGgt->AGat	1	Ccddee	D negative	Cde	this
10	RHD(G385A)	G->C at 1154	3' splice site intron 8 agGT->agCT Missense mutation G385A	1	CcddEe	weak D	cDE	42
15	RHD(K409K) study	G->A at 1227	5' splice site intron 9 AGgt->AAgt	5	Ccddee	D <sub>el</sub>	CDe	this

\* A detailed serologic analysis of this sample representing the partial D DIM has been published previously<sup>43</sup>.  
 † The same allele occurring in a cDe haplotype has been described as weak D type 11.

Table 8. Estimated frequencies in population

	PCR pattern/Allele	Frequency	
		among Cde/ cdE	in population
	Pattern 1	1:45	1:4,132
	Pattern 2	1:125	1:11,364
30	Pattern 3	1:101	1:17,976
	Pattern 4	1:500*	1:45,455*
	Pattern 6	1:167	1:15,152
	Pattern 7	1:302	1:53,929
35	Pattern 8	1:500	1:45,455
	RHD(W16X)	1:250	1:22,727
	RHD(G212V)	1:500	1:45,455
	RHD(Y330X)	1:500	1:45,455
	RHD(G1153(+1)A)	1:500	1:45,455
40	Any D negative	1:20 / 1:67†	1:1,607
	RHD(G486(+1)A)	1:167	1:15,152
	RHD(M295I)	1:71	1:6,493
	RHD(K409K)	1:100	1:9,091
45	Any D <sub>el</sub>	1:33‡	1:3,030

\* Assuming a Cde haplotype; a cdE haplotype would result in a frequency of 1:302 among cdE and 1:53,929 in the population. For statistics and sum frequencies, the haplotype was formally counted as 0.5 Cde and 0.5 cdE.

† 1:20 among Cde, 1:67 among cdE.

‡ 1:33 relative to the Cde haplotype.

EP 1 780 217 A1

**Table 9** False negative rate in routine typing for antigen D

Documented Phenotype	Samples (n)	Confirmed phenotype (n)		False negatives	
		D <sub>el</sub>	partial or weak D	n	Rate
ccddee	314	0	0	0	0%†
Ccddee	433	15	0	15	3.5%
ccddEe	271	0	2	2	0.7%
other	50	1	1	2	4%
D negative	N.A. ‡	0.15%	0.02%		0.17%

\* CCddee, CcddEe, ccddEE, and CCddEe  
† Upper limit of 95% confidence interval was 0.95% (Poisson distribution)  
‡ N.A. - not applicable. The frequencies are estimates based on the phenotype frequencies in the population <sup>41</sup>.

**Table 10** Previously described D negative, *RHD* positive alleles

Allele	Haplotype	Population	Possible match
<i>RHD</i> (Q41X) <sup>2</sup>	Cde	Whites	not detected
<i>RHD-CE</i> (2-9)-D <sup>1,24,25</sup>	Cde	Whites <sup>1,25</sup> , Blacks <sup>24</sup>	Pattern 1
<i>RHD-CE</i> (3:455-7)-D <sup>5,15</sup>	Cde <sup>s</sup>	Blacks	Pattern 8
<i>RHD</i> (488del4) <sup>1</sup>	Cde	Whites	not detected
<i>RHD-CE</i> (4-7)-D <sup>16</sup>	cdE	Whites	Pattern 3 or 4
<i>RHD</i> Ψ <sup>38</sup>	cde	Blacks	not detected
<i>RHD</i> (600del) <sup>10</sup>	Cde	somatic mutation*	not detected
<i>RHD</i> (exon 5 variant) <sup>8</sup>	cde	not communicated	not detected
<i>RHD</i> (G314V) <sup>34</sup>	Cde	Japanese	not detected
<i>RHD</i> (exon 9 variant)	Cde	Whites	Pattern 6

\* Allele acquired by somatic mutation in a woman with chronic myelogenous leucemia and restricted to the myeloid lineage

**Table 11** Population rates of false positives and positive predictive value of different *RHD* PCR strategies

PCR strategy	Rate of false positives	Positive predictive value of positive result	Number of polymorphism tested
Exon 10 only	1:1,275	0.999216	1
Intron 4/Exon 7	1:4,081	0.999755	2
Intron 4/Exon 7/ <i>RHD</i> Ψ	1:4,700	0.999787	3
Intron 4/Exon 7/W16X	1:5,212	0.999808	3
Intron 4/Exon 7/Intron 7	1:6,051	0.999835	3
Exons 3, 4, 5, 6, 7, 9	1:6,051	0.999835	6
Exons 2, 3, 4, 5, 6, 7, 9, 10	1:6,051	0.999835	8
Intron 4/Exon 7/W16X/ <i>RHD</i> Ψ	1:6,267	0.999840	4
All Exons/ <i>RHD</i> Ψ	1:7,520	0.999867	9
Intron 4/Exon 7/Intron 7/W16X	1:8,921	0.999888	4
Intron 4/Exon 7/Intron 7/W16X/ <i>RHD</i> Ψ	1:12,533	0.999920	5

## References

[0235]

- 5 1. Andrews KT, Wolter LC, Saul A, Hyland CA: The RhD- trait in a white patient with the RhCCee phenotype attributed to a four-nucleotide deletion in the RHD gene. *Blood* 92;1839, 1998
2. Avent ND, Martin PG, Armstrong-Fisher SS, Liu W, Finning KM, Maddocks D, Urbaniak SJ: Evidence of genetic diversity underlying Rh D negative, weak D (Du) and partial D phenotypes as determined by multiplex PCR analysis of the RHD gene. *Blood* 89:2568, 1997
- 10 3. Avent ND, Liu W, Jones JW, Scott ML, Voak D, Pisacka M, Watt J, Fletcher A. Molecular analysis of Rh transcripts and polypeptides from individuals expressing the DVI variant phenotype: an RHD gene deletion event does not generate all DVIccEe phenotypes. *Blood* 1997;89:1779-86.
- 15 4. Blumenfeld OO, Huang CH: Molecular genetics of the glycoprotein gene family, the antigens for MNSs blood groups: multiple gene rearrangements and modulation of splice site usage result in extensive diversification. *Hum Mutat* 6:199, 1995
- 20 5. Blunt T, Daniels G, Carritt B: Serotype switching in a partially deleted RHD gene. *Vox Sang* 67:397, 1994
6. Bowman J: The management of hemolytic disease in the fetus and newborn. *Semin Perinatol* 21:39,1997
7. Carritt B, Kemp TJ, Poulter M: Evolution of the human RH (rhesus) blood group genes: a 50 year old prediction (partially) fulfilled. *Hum Mol Genet* 6:843, 1997
- 25 8. Carritt B, Steers FJ, Avent ND. Prenatal determination of fetal RhD type. *Lancet* 1994;344:205-6.
9. Cherif-Zahar B, Mattei MG, Le Van Kim C, Bailly P, Cartron JP, Colin Y: Localization of the human Rh blood group gene structure to chromosome region 1p34.3-1p36.1 by in situ hybridization. *Hum Genet* 86:398, 1991
- 30 10. Cherif-Zahar B, Bony V, Steffensen R, Gane P, Raynal V, Goosens D, Laursen JS, Varming K, Jersild C, Cartron JP. Shift from Rh-positive to Rh-negative phenotype caused by a somatic mutation within the RHD gene in a patient with chronic myelocytic leukaemia. *Br J Haematol* 1998;102:1263-70.
- 35 11. Colin Y, Cherif-Zahar B, Le Van Kim C, Raynal V, van Huffel V, Cartron J-P. Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by southern analysis. *Blood* 1991;78:2747-52.
12. Cossu G, Angius A, Gelfi C, Righetti PG: Rh D/d genotyping by quantitative polymerase chain reaction and capillary zone electrophoresis. *Electrophoresis* 17:1911, 1996
- 40 13. Daniels GL, Faas BH, Green CA, Smart E, Maaskant-van Wijk PA, Avent ND, Zondervan HA, von dem Borne AE, van der Schoot CE. The VS and V blood group polymorphisms in Africans: a serologic and molecular analysis. *Transfusion* 1998;38:951-8.
- 45 14. Döscher A, Schunter F, Müller TH: Quantitative PCR to assess RHD zygosity. *Infusionsther Transfusionsmed* 26(suppl 1):31, 1999 (abstr.)
15. Faas BHW, Becker EAM, Wildoer P, Ligthart PC, Overbeeke MAM, Zondervan HA, von dem Borne AEGK, van der Schoot CE: Molecular background of VS and weak C expression in blacks. *Transfusion* 37:38, 1997
- 50 16. Faas BHW, Beckers EAM, Simsek S, Overbeeke MAM, Pepper R, van Rhenen DJ, von dem Borne AEGK, van der Schoot CE: Involvement of Ser103 of the Rh polypeptides in G epitope formation. *Transfusion* 36:506, 1996
- 55 17. Filbey D, Hanson U, Wesstrom G: The prevalence of red cell antibodies in pregnancy correlated to the outcome of the newborn: a 12 year study in central Sweden. *Acta Obstet Gynecol Scand* 74:687, 1995
18. Flegel WA, Khull SR, Wagner FF. Primary anti-D immunization by weak D type 2 RBCs. *Transfusion* 2000; 40:

428 - 434

- 5 19. Flegel WA, Wagner FF, Müller TH, Gassner C: Rh phenotype prediction by DNA typing and its application to practice. *Transfus Med* 8:281, 1998
20. Fujiwara H, Okuda H, Omi T, Iwamoto S, Tanaka Y, Takahashi J, Tani Y, Minakami H, Araki S, Sato I, Kajii E: The STR polymorphisms in intron 8 may provide information about the molecular evolution of RH haplotypes. *Hum Genet* 104:301, 1999
- 10 21. Gassner C, Schmarda A, Kilga-Nogler S, Jenny-Feldkircher B, Rainer E, Müller TH, Wagner FF, Flegel WA, Schönitzer D: RhesusD/CE typing by polymerase chain reaction using sequence-specific primers. *Transfusion* 37:1020, 1997
- 15 22. Huang CH. Human DVI category erythrocytes: correlation of the phenotype with a novel hybrid RhD-CE-D gene but not an internally deleted RhD gene. *Blood* 1997;89:1834-9.
23. Huang CH, Reid ME, Chen Y, Coghlan G, Okubo Y: Molecular definition of red cell Rh haplotypes by tightly linked SphI RFLPs. *Am J Hum Genet* 58:133, 1996
- 20 24. Huang CH: Alteration of RH gene structure and expression in human dCCee and DCW- red blood cells: phenotypic homozygosity versus genotypic heterozygosity. *Blood* 88:2326, 1996
- 25 25. Hyland CA, Wolter LC, Saul A. Three unrelated Rh D gene polymorphisms identified among blood donors with Rhesus CCee (r'r) phenotypes. *Blood* 1994; 84:321-4.
26. Kemp TJ, Poulter M, Carritt B: Microsatellite variation within the human RHCE gene. *Vox Sang* 77:159, 1999
- 30 27. Lo Y-MD, Bowell PJ, Selinger M, Mackenzie IZ, Chamberlain P, Gillmer MDG, Littlewood TJ, Fleming KA, Wainscoat JS: Prenatal determination of fetal RhD status by analysis of peripheral blood of rhesus negative mothers. *Lancet* 341:1147, 1993
28. Maaskant-van Wijk PA, Hemker MB, Douglas-Berger L, et al. Ethnic variability of the Rhesus system. *Transfusion* 1999;39:(10S)103S-4S (Abstract).
- 35 29. Maaskant-van Wijk PA, Faas BHW, de Ruijter JAM, Overbeeke MAM, von dem Borne AEGK, van Rhenen DJ, van der Schoot CE. Genotyping of RHD by multiplex polymerase chain reaction analysis of six RHD-specific exons. *Transfusion* 1998;38:1015-21.
- 40 30. Maaskant-van Wijk PA, Beckers EAM, van Rhenen DJ, Mouro I, Colin Y, Cartron J-P, Faas BHW, van der Schoot CE, Apoil PA, Blancher A, et al. Evidence that the RHD(VI) deletion genotype does not exist. *Blood* 1997; 90:1709-11.
- 45 31. MacGeoch C, Mitchell CJ, Carritt B, Avent ND, Ridgwell K, Tanner MJ, Spurr NK: Assignment of the chromosomal locus of the human 30-kDal Rh (rhesus) blood group-antigen-related protein (Rh30A) to chromosome region 1p36.13----p34. *Cytogenet Cell Genet* 59:261, 1992
32. Mourant AE, Kopec AC, Domaniewska-Sobczak K: The distribution of the human blood groups and other polymorphisms. London, Oxford University Press, 1976
- 50 33. Okuda H, Suganuma H, Tsudo N, Omi T, Iwamoto S, Kajii E: Sequence analysis of the spacer region between the RHD and RHCE genes. *Biochem Biophys Res Commun* 263:378, 1999
34. Okuda H, Kawano M, Iwamoto S, Tanaka M, Seno T, Okubo Y, Kajii E: The RHD gene is highly detectable in RhD-negative Japanese donors. *J Clin Invest* 100:373, 1997
- 55 35. Onda M, Fukuda M: Detailed physical mapping of the genes encoding glycoporphins A, B and E, as revealed by P1 plasmids containing human genomic DNA. *Gene* 159:225, 1995

## EP 1 780 217 A1

36. Reboul J, Gardiner K, Monneron D, Uze G, Lutfalla G: Comparative genomic analysis of the interferon/interleukin-10 receptor gene cluster. *Genome Res* 9:242, 1999

5 37. Salvignol I, Calvas P, Socha WW, Colin Y, Le Van Kim C, Bailly P, Ruffie J, Cartron JP, Blancher A: Structural analysis of the RH-like blood group gene products in nonhuman primates. *Immunogenetics* 41:271, 1995

10 38. Singleton BK, Green CA, Avent ND, Martin PG, Smart E, Daka A, Narter-Olaga EG, Hawthorne LM, Daniels G: The presence of an RHD pseudogene containing a 37 base pair duplication and a nonsense mutation in africans with the Rh D- negative blood group phenotype. *Blood* 95:12, 2000

39. Urbaniak SJ, Robertson AE: A successful program of immunizing Rh-negative male volunteers for anti-D production using frozen/thawed blood. *Transfusion* 21:64, 1981

15 40. Wagner FF, Gassner C, Müller TH, Schönitzer D, Schunter F, Flegel WA: Three molecular structures cause Rhesus D category VI phenotypes with distinct immunohematologic features. *Blood* 91:2157, 1998

20 41. Wagner FF, Kasulke D, Kerowgan M, Flegel WA: Frequencies of the blood groups ABO, Rhesus, D category VI, Kell, and of clinically relevant high-frequency antigens in South-Western Germany. *Infusionsther Transfusionsmed* 22:285, 1995

42. Wagner FF, Gassner C, Müller TH, Schönitzer D, Schunter F, Flegel WA: Molecular basis of weak D phenotypes. *Blood* 93:385, 1999

25 43. Wagner FF, Frohmajer A, Ladewig B, Eicher NI, Lonicer CB, Müller TH, Siegel MH, Flegel WA. Weak D alleles express distinct phenotypes. *Blood* 2000; 95: 2699-2708

44. Wagner FF, Flegel WA. RHD gene deletion occurred in the Rhesus box. *Blood* 2000;

30 45. Wissenschaftlicher Beirat der Bundesärztekammer; Paul-Ehrlich-Institut. Richtlinien zur Blutgruppenbestimmung und Bluttransfusion (Hämotherapie). Köln: Deutscher Ärzte-Verlag; 1996.

35 46. Legler TJ, Kohler M, Mayr WR, Panzer S, Ohto H, Fischer GF; Genotyping of the human platelet antigen systems 1 through 5 by multiplex polymerase chain reaction and ligation-based typing. *Transfusion* 1996 May; 36(5): 426-31

EP 1 780 217 A1

SEQUENCE LISTING

<110> DRK Blutspendedienst Baden-Wuerttemberg gGmbH  
 5 <120> Molecular structure of RHD negative  
 <130> D2600 EP/1  
 <141> 2000-10-31  
 10 <160> 66  
 <170> PatentIn Ver. 2.1  
 <210> 1  
 15 <211> 100  
 <212> DNA  
 <213> Homo sapiens  
 <400> 1  
 20 gcatgcgca ctgagccggg tggatggtag tgctgcatcc ggggtgtctgg aggctgtggc 60  
 cgttttgttt tcttggctaa aatcggggga gtgagggcggg 100  
 <210> 2  
 <211> 26  
 <212> DNA  
 25 <213> Homo sapiens  
 <400> 2  
 actgcacgac ggggctggac tgacgt 26  
 30 <210> 3  
 <211> 45  
 <212> DNA  
 <213> Homo sapiens  
 35 <400> 3  
 agctgaaaaa aatgtctgga tttctagagg gcttgagatg ctcag 45  
 40 <210> 4  
 <211> 26  
 <212> DNA  
 <213> Homo sapiens  
 <400> 4  
 45 gcttccattg ctgctggtgt actagt 26  
 <210> 5  
 <211> 44  
 <212> DNA  
 <213> Homo sapiens  
 50 <400> 5  
 agttttttac aggctggtgg attatcatag atgcagctgt tatt 44  
 55 <210> 6  
 <211> 26  
 <212> DNA

EP 1 780 217 A1

<213> Homo sapiens  
 <400> 6  
 5     tatagcaacc atagccttcc taatgt                     26  
  
 <210> 7  
 <211> 44  
 <212> DNA  
 10     <213> Homo sapiens  
  
 <400> 7  
 aggattaatg cagtatcgaa tggacaagtc cgaggtgata gtta                     44  
  
 <210> 8  
 <211> 26  
 <212> DNA  
 15     <213> Homo sapiens  
  
 <400> 8  
 20     tgaaggttgt ctgggtcaaa acaggt                     26  
  
 <210> 9  
 <211> 44  
 <212> DNA  
 25     <213> Homo sapiens  
  
 <400> 9  
 aggtgctcgc atttggcttt tcgttggttt catgttggcc tttg                     44  
  
 <210> 10  
 <211> 26  
 <212> DNA  
 30     <213> Homo sapiens  
  
 <400> 10  
 35     tttttgagg ttatgttgct aaaggt                     26  
  
 <210> 11  
 <211> 38  
 <212> DNA  
 40     <213> Homo sapiens  
  
 <400> 11  
 agaaaaagac atagtatacc ctggaattgc tgtatntt                     38  
  
 <210> 12  
 <211> 26  
 <212> DNA  
 45     <213> Homo sapiens  
  
 <400> 12  
 50     tccagaatgc cttcatcttt tttggt                     26  
  
 <210> 13  
 <211> 53  
 55     <212> DNA

EP 1 780 217 A1

<213> Homo sapiens

<400> 13

aggagggctg gtttttaagt ttggccgcac tgaagactta tggcagtga cac 53

5

<210> 14

<211> 249

<212> DNA

<213> Homo sapiens

10

<400> 14

agcatcatcc taatgaaact aacatthtat tttaaactta ttaaattgac tcttaacta 60  
 agtttttagt ctttaatttt ttaatatcaa atctgtctct gaccttggtt cattatacat 120  
 aaggagcttt gctgtcatga gcgtttctca cgtacaaatg caggcaacag tgagaggaag 180  
 ttgtcttggt tttgaacagg ccttgthttt cttggatgct tttgcttaa atccaacagc 240  
 caaatgagg 249

15

<210> 15

<211> 3093

<212> DNA

<213> Homo sapiens

20

<400> 15

attcactatc acaagaacag cacgggtaag acctgtcccc atgattcagt tacctccac 60  
 tgggtccctc ccacaacgca tgggaattca ggatgagatt tgggtgggga cacaaccaa 120  
 ccctatcatt ccacccatgg cccctcccaa atttcatgtc ctcacattc aaaaccaatc 180  
 acaccatccc aacagtcctt caaagtctta atgatttca gcattaactc aaaagtccac 240  
 agtctaattg ctcactctgag acaaggcaag tcctttccat ttatgagcct ataaaatcca 300  
 agcaagtta gttacttctt agatacaatg ggggtacagg cattgggtaa atacagccat 360  
 tccaaatggg ataaattggt caaaacaaag aggtacaggg cccatgagag tccaaaatcc 420  
 agtggggcag tcaaatctta aagctccaaa atgatctctt ttgactccac atctcacatc 480  
 caggtcacgc agatggaagg ggtgggttcc catgggtctg ggcagctctg ccctgtacc 540  
 tttgcagggt acagcctccc tctcagctgc tttcatgggc tggcattgag tgtctgcagc 600  
 ttttccaggt acacggtgca agctgtcggg ggatctacca ttctggggtc tggaggacct 660  
 cttctcacag ctcacttagg tggtgcccca gtagggactg tgtgtggggt ctctgacccc 720  
 acatttccct tctgcaactg cctggcagag gatctccatg agggcctgc tctgacaga 780  
 aacttctgac tgggcatcca ggcatttccg cacatcctct ttaatctagg cgaaggtttc 840  
 caaaccccaa ttcttgactt ctgtgcactc gcagtctcaa caccacatgg aagctgtcaa 900  
 ggcttggggc ttgcaactcc cgaagctaca gcccaagctc taccttgctt cccgtcagtc 960  
 atggttggga gtggctggga tgcagggcac caagtcctta ggctgcacac agcatgagga 1020  
 cccggggcct ggccaacaaa accatthttt cctgatacct ctggacctgt gatgggaggg 1080  
 gttgccataa agacctctga catgcctctg agacattttc cccattgtct tgggaattag 1140  
 catttggtct ctgttactca tgcaaaatttc tgcagccagc ttgaatttct cctcagaaa 1200  
 tgggaatttt tcttttctat cacattgtca ggtcgcfaat tttccgaact tttatgctct 1260  
 gcttccctta taaaactgaa tgtctttaa acgacccaag tcacctctg aatgctttgc 1320  
 tgcttagaaa tttctctctc cagatactct aatcatctc tctgaagttc aaagttctac 1380  
 aaatatctcg tgcaggggca aaatgccgcc agtatctttg ctaaaacata acaagagtcc 1440  
 cctttgctcc agttcccaac aagttcctca ttccgtctg agaccacctc agcctatgga 1500  
 ctttattgct cacagtgcta tcagcatttt gggcaaagcc attcaacaag tctctagga 1560  
 gttccaaact tcccacatt tgctgtctt cttctgagcc ctccaaactg ttccaaacc 1620  
 tgctgttac ccagttccaa agtcacatac ccatttttga gtatctacgg cagcaccaca 1680  
 ctctactggg accaatttag ccaactgaag agttggagaa cagaagtaat agactctggt 1740  
 ttacattgta aaagcttctc tgtggctgct gtgtgaagaa aatatatgag aatgaagccc 1800  
 caagatgaag cagggacaca gttgcagtggt tttagagtaag aatgctgct ggctggcact 1860  
 gaagtagatg cctggagggt tgtgtgtgca catgcatgtg tatgtgttt acgatagta 1920  
 gcccaacaga tactgtaatc cacactgtt ttttttttt gagacagagt ctcactgtt 1980  
 gcctagacta gaatgcagtg gcacaactct ggctcactac aacctccacc tcccaggttc 2040  
 aaacaatcct tgtgcttcag cctcccaggt agttgggatt acaggtgtgt gccaccgtgc 2100  
 ccagctatat tttttgtatt ttttagcagag atgggatttt gccacattgg ccaggctggt 2160  
 cttgaaactc tggcctcaag caatcctccc accttagcct cccaaagtgc tgagccaaca 2220  
 cacctggccc caactgattt ttaatcatga aatgacacat acatttaaaa aaccaatac 2280

55

EP 1 780 217 A1

	ctataatatt	cctggctagt	actcttcaca	tctatatcat	caaaaacaaa	gaaagtatgt	2340
	gaaactgaca	cagccaaggg	gagactaagg	agacataaca	attaactgta	atgtgggtatt	2400
	ctggagggga	tcttgggaaca	gaaaaagaca	ttaggcaaaa	aactaaagaa	atctgaataa	2460
5	aatgtggatg	tcagttaata	ataatgtatc	atatttagtcc	agtaattgta	acaaatatac	2520
	cacaataatg	aaagccatta	attatagggga	aaatggaggg	gttaatatgg	gtggctggct	2580
	tttgctattt	ctagcagctc	cattttatct	gcaaaagaca	aacattcatt	aagtcccaaa	2640
	aaggtaaaaga	atgacaaaatt	aagcatgtat	cttattagta	agagtaatat	aaagatgctc	2700
	actcctattt	ataaatattt	gacaatcatg	ttaaggccac	aaaagagaaa	aaagggtagg	2760
	ggcaaaaaac	gcaaagagaa	aggagttagt	atcttttctc	ccgactcat	tagctattaa	2820
10	aagaggatgt	ttgtttaaag	ctgctcagag	ctggtaaact	aatgttaagt	cactaacggg	2880
	aattttaaag	gtttcattaa	gaactgcctg	cactagattc	ctccaccctg	agacattaaa	2940
	caatcacgat	aaacctcctg	agtggtaaga	acttgtccat	ttaaaaacag	gctatagatt	3000
	gtatcatgca	gttttatcta	ctaactggct	aatatcccg	caaaaacaaa	aaaccccaaa	3060
	gggatgaaag	tttcatccat	caaaggaaac	aac			3093
15		<210> 16					
		<211> 3099					
		<212> DNA					
		<213> Homo sapiens					
20		<400> 16					
	attcactatc	acaagaacag	cacgggtaag	acctgtcccc	atgattcagt	tacctccac	60
	tgggtccctc	ccacaacgca	tgggaattca	ggatgagatt	tgggtgggga	cacaacaaa	120
	ccctatcatt	ccacccatgg	cccctcccaa	atltcatgtc	ctcacatttc	aaaaccaatc	180
	acaccatccc	aacagtcctt	caaagtctta	aatgatttca	gcattaactc	aaaagtccac	240
25	agtctaattg	ctcatctgag	acaaggcaag	tcctttccat	ttatgagcct	ataaaatcca	300
	aagcaagtta	gttacttcct	agatacaatg	gggtacaggg	cattgggtaa	atacagccat	360
	tccaaatggg	ataaattggg	caaaacaaag	aggctacaggg	ccatgagag	tccaaaatcc	420
	agtggggcag	tcaaattctta	aagctccaaa	atgatctcct	ttgactccac	atctcacatc	480
	caggtcacgc	agatggaagg	ggtgggttcc	catggctctg	ggcagctctg	cccctgtacc	540
	tttgcagggt	acagcctccc	tctcagctgc	tttcatgggc	tggcattgag	tgtgtgcagc	600
	ttttccaggt	acacgggtgca	agctgtcggt	ggatctacca	ttctgggggt	tggaggacct	660
30	cttctcacag	ctccactagg	tgggtgcccc	gtagggactg	tgtgtggggg	ctctgacccc	720
	acatttccct	tctgcactgc	cctggcagag	gatctccatg	agggcctctg	tctctcagca	780
	aacttctgac	tgggcatcca	ggcatttccg	cacatcctct	ttaatctagg	cgaaggtttc	840
	caaaccceca	ttcttgactt	ctgtgcactc	gcagtctcaa	caccacatgg	aagctgtcaa	900
	ggcttggggc	ttgcactccc	cgaagctaca	gcccaagctc	taccttgctc	cccgtcagtc	960
35	atggttggga	gtggctggga	tgcagggcac	caagtcocct	ggctgcacac	agcatgagga	1020
	ccccgggcct	ggccaacaaa	accatthttt	cctgatacct	ctggacctgt	gatgggaggg	1080
	gttgccataa	agacctctga	catgccctgg	agacatttct	cccattgtct	tgggaattag	1140
	catttggctc	ctgttactca	tgcaaaatttc	tgcagccagc	ttgaatttct	cctcagaaaa	1200
	tgggaatttt	tcttttctat	cacattgtca	ggctgcaaat	tttccgaact	tttatgtctc	1260
	gcttccctta	taaaactgaa	tgtctttaac	agcaccocaa	tcacctcttg	aatgctttgc	1320
40	tgcttagaaa	tttctcctgc	cagatactct	aatcatctc	tctgaagttc	aaagttctac	1380
	aaatatctcg	tgcaggggca	aaatgccgcc	agtatctttg	ctaaaacata	acaagagtcc	1440
	cctttgctcc	agttcccaac	aagttcctca	tttccgtctg	agaccacctc	agcctatgga	1500
	ctttattgtc	cacagtgtca	tcagcatttt	gggcaaagcc	attcaacaag	tctctaggaa	1560
	gttccaaact	ttcccacatt	tgcctgtctt	cttctgagcc	ctccaaactg	ttccaaacct	1620
	tgcctgttac	ccagttccaa	agtcacatac	ccatttttga	gtatctacgg	cagcacccca	1680
45	ctctactggg	accaattttag	ccactgaagt	agttggagaa	cagaagtaat	agactctggg	1740
	ttacattgta	aaagcttctc	tgtggctgct	gtgtgaagaa	aatatatgag	aatgaagccc	1800
	caagatgaag	cagggacaca	gttgcagtgg	ttagagtaag	aaatgtgctc	ggctggcact	1860
	gaagtgatag	cctggagggt	tgtgtgtgca	catgcatgtg	tatgtgtttt	acgatagtag	1920
	gcccacaga	tactgtaatc	cacacttggt	tttttttttt	ttttgagaca	gagtctcacc	1980
	tgttgccctag	actagaatgc	agtggcacaa	tcttggctca	ctacaacctc	cacctcccag	2040
50	gttcaaacaa	tccttgtgct	tcagcctccc	gagtagttgg	gattacaggt	gtgtgccacc	2100
	gtgccagct	atattttttg	tatttttagc	agagatggga	ttttgccaca	ttggccaggg	2160
	tggctctgaa	ctcctggcct	caagcaatcc	tcccacctta	gcctcccaaa	gtgctgagcc	2220
	accacacctg	gccgcaactg	atthtttaac	atgaaatgac	acatacattt	aaaaaaccca	2280
	atacctataa	tattcctggc	tagtactctt	cacatctata	tcatcaaaaa	caaagaaagt	2340
	atgtgaaact	gacacagcca	aggggagact	aaggagacat	aacaattaac	tgtaatgtgg	2400
55	tattctggag	gggatcctgg	aacagaaaaa	gacattaggg	aaaaaactaa	agaaatctga	2460

EP 1 780 217 A1

5  
10  
15

```

ataaaatgtg gatgtcagtt aataataatg tatcatatta gtccagtaat tgtaacaaat 2520
ataccacaat aatgaaagcc attaattata gggaaaatgg aggggttaat atgggtggct 2580
ggcttttgct atttctagca gctccatttt atctacaaaa gacaaacatt cattaagtcc 2640
caaaaaggtg aagaatgaca aattaagcat gtatcttatt agtaagagta atataaagat 2700
gctcactcat atttataaat atttgacaat gatgttaagg ccagaaaaga gaaaaaaggg 2760
taggggcaaa aaacgcaaag agaaaggagt tagtatcttt tctcccgac tcattagcta 2820
ttaaagagg atgtttgttt aaagctgctc agagctggta aactaatgtt aagtcactaa 2880
cggaattta aaaggtttca ttaagaactg cctgcaactag attcctccac cctgagacat 2940
taaacaatca cgataaacct cctgagtggt aagaacgtgt ccatttaaaa acaggctata 3000
gattgtcatg cagttttatc tactaatcgg ctaatgcacc gccaaaaaca aacaaaaaaa 3060
cccaaaggga tgaaagtttc atccatcaaa ggaacaac 3099

```

<210> 17

<211> 3099

<212> DNA

<213> Homo sapiens

<400> 17

20  
25  
30  
35  
40  
45  
50  
55

```

attcactatc acgagaacag cacgggtaag acctgtcccc atgattcagt tacctccac 60
tgggtccctc ccacaacgca tgggaattca ggatgagatt tgggtgggga cacaaccaa 120
ccctatcatt ccacccatgg cccctcccaa atttcatgtc ctcacatttc aaaaccaatc 180
acaccatccc aacagtcctt caaagtctta aatgatttca gcatctaactc aaaagtccac 240
agtctaattg ctcatctgag acaaggcaag tcctttccat ctatgagcct ataaaaatcca 300
aagcaagtta attacttcc tcaaaatggg agatacaatg ggggtacagg cattgggtaa atacagccat 360
tccaaatggg ataaattggt caaaacaaag aggtacagg cccatgagag tccaaaatcc 420
agtggggcag tcaaacttta aagctccaaa atgatctcct cttgactcca catctcatat 480
ccaggtcacg cagatggaag ggggtgggtc ccatggtctt gggcagctct gccctgtac 540
ctttgcaggg tacagcctcc ctctcagctg ctctcatggg ctggcattga gtgtctgcaa 600
cttttccagg tacacgggtc aagctgtcgg tggatctacc attctggggg ctggaggacc 660
tctttctaca gctccactag gtggtgcccc agtagggact gtgtgtgggg tctctgacc 720
cacatttccc ttctgcactg cctggcaga ggcctccat gagggcctg cccctgacc 780
aaacttctgc ctgggcatcc aggcatttcc gcacatctc tttaatctag gcgaaggtt 840
ccaaaccca gttcttgact tctgtgact cgcagtctca acaccacatg gaagctgtca 900
aggcttgggg cttgcaactc ccgaagctac agccaagct ctacctgoc tctgtcagt 960
catggttggg agtggtctgg atgcagggca ccaagtcctt aggtcgaca cagcatgagg 1020
accccgggcc tggccaacaa aaccattttt tcctgatata tctggacctg tgatgggagg 1080
ggttgccata aagacctctg acatgcctg gagacatttt cccattgtc ttgggaatta 1140
gcatttggct cctgttactc atgcaaattt ctgcagccag cttgaatttc tctcagaaa 1200
atgggaattt ttcttttcta tcacattgtc aggtgcaaa ttttccgaa ttttatgtc 1260
tgcttccctt ataaaactga atgtctttaa cagcacccaa gtcacctctt gaatgctttg 1320
ctgcttagaa atttctcctg ccagatactc taaatcatct ctctgaagtt caaagttcta 1380
caaatatctc gtgcaggggc aaaatgccc cagtatcttt gctaaaacat aacaagagtc 1440
ccctttgctc cagttcccaa caagttctc atttccgtct gagaccacct cagcctatgg 1500
actttatgt ccacagttt atcagcattt tgggcaaagc cattcaacaa gtccttagga 1560
agttccaaac tttcccacat ttgcctgtct cttctgagc cctccaaact gttccaaacc 1620
ctgcctgta cccagttcca aagtcacata cccatttttg agtatctacg gcagcacc 1680
actctactgg taccattta gccactgaag tagttggaga acagaagtaa tagactctgg 1740
tttacattgt aaaagcttct ctgtggctgc tgtgtgaaga aaatatatga gaatgaagcc 1800
ccaagatgaa gcagggacac agttgcagtg gttagagtaa gaaatgctgc tggctggcac 1860
tgaagtgata gcctggagg ttgtgtgtgc acatgcatgt gtatgtgttt tacgatagta 1920
ggccaacag atactgtaat ccacacttgt ttttttttt tttttgagac agagtctcac 1980
ctgttgcta gactagaatg cagtggcaca atcttggctc actacaacct ccacctcca 2040
ggttcaaca atccttgtgc ttcagcctcc cgagtagttg ggattacagg tgtgtgccac 2100
cgtgccagc tatattttt gtattttttag cagagatggg attttgccac attggccagg 2160
ctggtcttga actcctggcc tcaagcaatc ctcccacctt agcctcccaa agtgctgagc 2220
caccacact ggccgcaact gatttttaat catgaaatga cacatacatt taaaaaacc 2280
aatacctata atattcctgg ctagtactct tcacatctat atcatcaaaa acaagaag 2340
tatgtgaaac tgacacagc aaggggagac taaggagaca taacaattaa ctgtaatgtg 2400
gtattctgga ggggatcctg gaacagaaaa agacattagg caaaaaacta aagaaatctg 2460
aataaaatgt ggatgtcagt taataataat gtatcatatt agtccagtaa ttgtaacaaa 2520
tataccaat aatgaaagcc attaattata gggaaaatgg aggggttaat atgggtggct 2580
ggcttttgct atttctagca gctccatttt atctacaaaa gacaaacatt cattaagtcc 2640

```

EP 1 780 217 A1

```

caaaaaggta  aagaatgaca  aattaagcat  gtatcttatt  agtaagagta  atataaagat  2700
gctcactcat  atttataaat  atttgacaat  gatgttaagg  ccagaaaaga  gaaaaaagg  2760
taggggcaaa  aaacgcaaag  agaaaggagt  tagtatcttt  tctcccgcac  tcattagcta  2820
ttaaagagg  atgtttgttt  aaagctgctc  agagctggta  aactaatggt  aagtactaa  2880
5  cggaattta  aaaggtttca  ttaagaactg  cctgcaactag  attcctccac  cctgagacat  2940
taaacaatca  cgataaacct  cctgagtggt  aagaacgtgt  ccatttaaaa  acaggctata  3000
gattgtcatg  cagttttatc  tactaatcgg  ctaatgcacc  gccaaaaaca  acaaaaaaaa  3060
cccaaagga  tgaaagtttc  atccatcaa  ggaacaac  3099

<210> 18
<211> 9241
<212> DNA
<213> Homo sapiens

<400> 18
15  ctagaaaaca  ctttgtcatt  ttagaggtgt  tatccaatgt  tcgcgaggc  actggagtca  60
gagaaaatgg  agttgaatcc  tttctctgcc  actctttgag  gagaatctca  ccatttatta  120
tgcactgtag  aatacaaca  taaaatacag  ccatgtacca  cataacaaca  tcttggtaaa  180
caacagactg  catatatgat  ggtggtcatt  cagtaagcta  aggttaattt  attattattc  240
cctttttttt  tctttttttt  tgagatgtag  tcttactctg  tcaccaggc  tagagtgcaa  300
20  tggcaccatc  ttggctcact  gcaacctctg  cctcctgggt  tcaagcgaat  ctctgctc  360
agcctccgaa  gtagctggga  attacaggca  ccaccacat  ctggctaatt  tttgtattt  420
ttagtaaaga  tggggtttca  ccattgtggc  caggtctgat  tcaaactcct  gacctcaagt  480
gatctgcctg  cctcggcctc  ccaaagtgtc  gggaccatag  gcctgagcca  ctgtgcccg  540
ccttgtttgc  ttttttaaca  gttaacagtg  tgctcataga  aactgctttg  acatgactgc  600
aatcatgtgc  ttcatagaaa  cttaattaga  ttataccact  agagtcttca  gatttttata  660
25  cttttttttt  tgaaacggag  tctcactctg  tcaccaggct  ggagtgcagt  gccgcaatct  720
cagctcggcg  caacctccgc  ctcccagggt  caagtgatc  tctgcctca  gcctccgag  780
tagctgggat  tacaagtgca  cactaccacg  cccagctaatt  ttttgcatt  ttactagaca  840
gggtttcacc  atgttggcta  ggatagtttc  accaggatct  cttggcctca  tgatcagcct  900
gcctcggcct  cccaaagtgc  tgggattaca  ggtgtgagcc  accgtgccca  gcctatactg  960
ccttttttga  ataccatttg  gcgtttttaa  gaattaacag  ctttgtgaac  gtggcagtgc  1020
30  ttgtgattca  ggcttccact  gagaccaagg  ggagaacctg  gttgcaggac  aaacagacgg  1080
acagcgtgtg  gcagtgttta  aatgctcttc  tgaaggctga  tacgacagct  ctctgtgcac  1140
tgattgtata  cgcatcccaa  gattatatta  ttgttttcta  ttgctatgtg  tcacactttg  1200
ccaaacagga  tgtggaaaat  gaataagcgg  ttttcttagg  cacttcttaa  cagacaattg  1260
gtcaaaatga  actccattgc  ttaagaaaca  cataaacacc  atttagtcac  tgaatatagc  1320
tatatgtatg  gttgctacta  tggggaatct  gttttgcca  atttctttg  aaaatctcgg  1380
35  cagaccaagg  tcttttttgt  ttacacaata  cttgaaaaat  aaaaatgaac  aagccaaca  1440
actaccaagt  tttcacttac  ataatgttag  ttacatacag  aaaatgtgac  tgtgaatttt  1500
ttctaggact  tttaaactat  aagcactatt  tgcacgaaag  agaaccaatc  tatcaattac  1560
aaactcacat  aattttacag  atttttttt  cctacacag  cacataaac  agaaggaatt  1620
tgaagccacc  ctccaaacac  aggggaagga  ggctgtgtgt  atatcctcat  tgtctttcac  1680
40  attctaagg  ggttccactc  agtgactgaa  atccttaagt  gttgtattag  tccggttggg  1740
ctaccataac  agcagcttaa  actgttgtaa  agccactcag  acttaacaa  cagaaattta  1800
tttcttata  gttctggagg  ctggaagtcc  aaggtgccgg  caaggctggg  ttctggtgag  1860
acctctctcc  ctgtcttgca  gatggctgcc  tctcctctgt  gtctcatag  agcctgtcct  1920
ctgcttttac  acttctggtg  tcatcttctc  tttttttttt  tttttgagac  agagtctcgc  1980
tctatcgccc  aggctggagt  gcagtggccc  gatcgatctc  ggctcactgc  aacctctgcc  2040
tcccagttc  aagcaattct  cctgcctcag  cctcccgagt  agctgggact  acaggtgcc  2100
45  gccatcatgt  ctggctaatt  tttgtatttt  tagtagagac  agggtttcc  catattggcc  2160
aggctggtct  ccaactcctg  acctgtcat  ctgctgcct  cggcctocca  aagtgtcagg  2220
attacaggcg  tgagccaccg  caccggcct  ctttctcttc  ttataaggac  accagtctca  2280
ttagattagg  gctccaccct  catgacctca  tttgacctta  actattattt  ctttaagca  2340
cctatttcca  aatatagtca  cttaggggt  tagggcttca  aatatgaat  ctgagggaga  2400
50  tcaattcagt  aatagcagt  agtcattaac  ggacaatata  tacaagata  atttcgtgat  2460
tactgtcctt  atgcataaat  gtctcagtg  tctcactgcc  ttatccaga  ttactatca  2520
caaagacttt  gctctgagaa  aaatgtgatt  tcttctttt  tttttttttt  ttgagacaga  2580
gtctcactct  gtcacccagg  ctggagtgca  gtggtgcaat  ctcggtcac  tgcaatctcc  2640
gcctcccagg  ttcacgccat  tctcttgctc  cagtctcccg  agtagctggg  cctacaggcg  2700
cccgcaacc  tgcccagcta  attttttgta  ttttttagtag  agacggggt  tcaccatggt  2760
55  agccaggatg  gtctcaatct  cctgacctcg  tgatccacct  gcctcagcct  cccaaagtgc  2820

```

EP 1 780 217 A1

	tgggattaca	ggcatgagcc	accggcceca	gcagatTTTT	TTTTTTTTT	TTTTTTTTTg	2880
	agatggagtc	ttgctgtgtt	gcccagcctg	gagtgcagtg	ttatgatTTT	ggctcactgc	2940
	aacctctgtc	taccatgttc	aagcgattct	cccacctctg	cctcccgtgt	agctgggatc	3000
5	acaggcacac	gccaccacac	ctagctactt	tttgtatTTT	tagtagaaat	ggggTTTcac	3060
	catgttggcc	aggatgggtc	cgaactcctg	acctcaagtg	atcctcctgc	ctcggcctcc	3120
	caaagtgtcg	ggattacagg	tgtgagccac	tgtgcctggc	caaaaatgtg	atttcttatt	3180
	tcccacattg	ccaattccat	ttcaattaac	tataatagct	atgtctattg	agcactcaag	3240
	tgtattctag	aaactgttcc	tgattctggg	gataatcca	tgaatcaact	atagtcctcg	3300
	ttattaagta	atctgtagtc	tgactaaacc	attagaaatt	taaaaaatgg	ctactttcaa	3360
10	agacatcttg	gagttcagga	gtcccacact	gcgaaccata	ttacctata	atccaacctg	3420
	cttgtaattc	acttatttta	ccaatattta	ttgagtcca	actttgagcc	taagatacac	3480
	cagtaaacaa	atggataaag	tccctgtcct	catgaaactt	gtattctaat	ggaagaaaca	3540
	gaaaacaaac	agatatagga	tgtaatatca	ggtagggata	aatactttga	attcgaacaa	3600
	aagtatacgt	agtcagggtt	cgccaaagag	acacagccaa	tgggatacat	agatatataa	3660
	aagagggttt	atgagttaga	aagggctcac	atgattacag	aggctgagaa	gtcccacagc	3720
15	agattgtctg	caagctggag	accaggggat	actggtagca	tggctcagtc	caagtcccaa	3780
	agcctcagaa	tcaggaaagc	tgatgatata	attcttagcc	caaaggcctt	agaacccag	3840
	cggtgacgga	aaggctgatg	taggtcctgg	agtctgaga	cccaacagcc	tgggatcctg	3900
	aaatccaagg	gcaqqaatq	aaqcttat	tccagctcca	agagagtaag	accaatttgc	3960
	ctttctccg	tttttgttcc	aagccacctg	ccattgagg	gcggtgggtt	ccctcttagt	4020
	ccattcagtc	atatatcaat	ctcttctgga	aataccctca	cagacacact	aacaaataat	4080
20	gcctttccag	ttctctaggt	attctttaat	ccagtcaagc	tgacacctaa	aattaacct	4140
	cacaaaagtt	aaggagaaag	aagacaactt	gtaggggagg	ctgctatgca	agacagtgtg	4200
	tgaaggaagg	gctctctgaa	gaggttaata	tctgagcaga	gacttgaatg	aagtgaagaa	4260
	gtgagccatg	tgggtatggg	gaatacaact	tccaggtaga	gaagacaagt	gtgggtgtga	4320
	tcagggtcag	caaagaagcc	atgtgacaga	tcaaggtggg	ccagggagag	acggataaag	4380
	gatctaactc	ctgaggaggt	ggcctggcca	ggagctagag	catgaagatc	tcgtaggact	4440
25	ttattctgca	aggtgaaaag	ccattgtatt	agtctgttca	caaaccogag	actaggcaat	4500
	ttacaaaaga	aagagaggtt	taatggactt	acagttccac	atggctgggg	aggcctcaca	4560
	atcatggcga	aaggcaatga	ggagcaagtc	acgtcttacg	tggatggcag	gcaaagacaa	4620
	agacagcttg	tgcagagaaa	ctccccctta	tagagccatc	agatcctgtt	agacttatc	4680
	actatcacia	gaacagcacg	ggtaagacct	gtccccatga	ttcagttacc	ttccctggg	4740
	tocctcccac	aacgcattgg	aattoaggat	gagatttggg	tggggacaca	accaaaccct	4800
30	atcattccac	ccatggcccc	tcccaaattt	catgtcctca	catttcaaaa	ccaatcacac	4860
	catcccacaa	gtccctcaaa	gtcttaaatg	atctcagcat	taactcaaaa	gtccacagtc	4920
	taatgtctca	tctgagacaa	ggcaagtcct	ttccatttat	gagcctataa	aatccaaagc	4980
	aagttagtta	cttccctagat	acaatggggg	tacaggcatt	gggtaaacac	agccattcca	5040
	aatgggataa	attgggtcaaa	acaaagaggg	tacaggccca	tgagagtcca	aaatccagtg	5100
35	gggcagtcaa	atcttaaagc	tccaaaatga	tctcctttga	ctccacatct	cacatccagg	5160
	tcacgcagat	ggaaggggtg	ggttcccatg	gtcttgggca	gctctgcccc	tgtacctttg	5220
	caggggtacag	cctccctctc	agctgctttc	atgggctggc	attgagtgtc	tgcagctttt	5280
	ccaggtacac	ggtgcaagct	gtcgggtggat	ctaccattct	ggggctctgga	ggacctcttc	5340
	tcacagctcc	actaggtggt	gccccagtag	ggactgtgtg	tggggtctct	gacccacat	5400
	ttccctctctg	cactgcctctg	gcagaggatc	tccatgaggg	cctgctcct	gcagcaaaact	5460
40	tctgactggg	catccaggca	tttccgcaca	tctcttttaa	tctaggcgaa	ggtttccaaa	5520
	ccccattctt	tgacttctgt	greactgcag	tctcaacacc	acatggaagc	tgtcaaggct	5580
	tggggcttgc	actccccgaa	gctacagccc	aagctctacc	ttgctctccg	tcagtcattg	5640
	ttgggagtg	ctgggatgca	gggcaccaag	tccctaggct	gcacacagca	tgaggacccc	5700
	gggcctggcc	aacaaaacca	tttttctctg	atacctctgg	acctgtgatg	ggaggggttg	5760
	ccataaagac	ctctgacatg	ccctggagac	attttcccc	ttgtcttggg	aattagcatt	5820
45	tggtctctgt	tactcatgca	aatctctgca	gccagcttga	atctctctc	agaaaaatgg	5880
	aatcttctct	ttctatcaca	ttgtcaggct	gcaaatcttc	cgaactttta	tgtctgtctt	5940
	ccctataaaa	actgaatgtc	tttaacagca	ccaagtcac	ctcttgaatg	ctttgtgtct	6000
	tagaaatctc	tctgtccaga	tactctaaat	catctctctg	aagttcaaa	ttctacaaat	6060
	atctcgtgca	ggggcaaaat	gcccagtagt	tctttgctaa	aacataacaa	gagtcctctt	6120
	tgtctcagtt	cccaacaagt	tcctcatttc	cgctgagac	cacctcagcc	tatggacttt	6180
50	attgtccaca	gtgctatcag	cattttgggc	aaagccattc	aacaagtctc	taggaagttc	6240
	caaactttcc	cacatttgcc	tgtctctctc	tgagccctcc	aaactgttcc	aaacctgtcc	6300
	tgttaccag	ttccaaagtc	acatacccat	ttttgagtat	ctacggcagc	acccccctct	6360
	actgggtacca	atctagccac	tgaagtagtt	ggagaacaga	agtaatagac	tctggtttac	6420
	attgtaaaa	cttctctgtg	gctgctgtgt	gaagaaaata	tatgagaatg	aagccccag	6480
	atgaagcagg	gacacagttg	cagtggttag	agtaagaaat	gctgctggct	ggcactgaag	6540
55	tgatagcctg	gaggtttgtg	tgtgcacatg	catgtgtatg	tgttttacga	tagtaggccc	6600

EP 1 780 217 A1

aacagatact gtaatccaca cttgtttttt tttttttttt gagacagagt ctcacctgtt 6660  
 gcctagacta gaatgcagtg gcacaatctt ggctcactac aacctccacc tcccaggttc 6720  
 aaacaatcct tgtgcttcag cctcccaggt agttgggatt acaggtgtgt gccaccgtgc 6780  
 ccagctatat tttttgtatt ttttagcagag atggggatttt gccacattgg ccaggctggg 6840  
 5 cttgaaactcc tggcctcaag caatcctccc accttagcct cccaaagtgc tgagccacca 6900  
 cacctggccg caactgattt ttaatcatga aatgacacat acatttaaaa aaccaatac 6960  
 ctataatatt cctggctagt actcttcaca tctatatcat caaaaacaaa gaaagtatgt 7020  
 gaaactgaca cagccaaggg gagactaagg agacataaca attaactgta atgtggattt 7080  
 ctggagggga tcttggaca gaaaaagaca ttaggcaaaa aactaaagaa atctgaataa 7140  
 10 aatgtggatg tcagttaata ataatgtatc atattagtcc agtaattgta acaaatatac 7200  
 ccaataatga aagccattaa ttatagggaa aatggagggg ttaatattggg tggctggctt 7260  
 ttgctatttc tagcagctcc attttatcta caaaagacaa acattcatta agtcccaaaa 7320  
 aggtaaagaa tgacaaatta agcatgtatc ttattagtaa gagtaatata aagatgtcca 7380  
 ctcatattta taaatatttg acaatgatgt taaggccaga aaagagaaaa aagggtaggg 7440  
 gcaaaaaacg caaagagaaa ggagttagta tcttttctcc cgcaactcatt agctattaaa 7500  
 15 agaggatgtt tgtttaaagc tgctcagagc tggtaaacta atgttaagtc actaacggga 7560  
 atttaaaagg tttcattaag aactgcctgc actagattcc tccaccctga gacattaac 7620  
 aatcacgata aacctcctga gtggtaagaa cgtgtccatt taaaaacagg ctatagattg 7680  
 tcatgagatt ttatctacta atcggctaata gcaccgcaa aaacaacaaa aaaaaccca 7740  
 agggatgaaa gtttcatcca tcaaaggaaa caacagtcac cttggttccc atcccactca 7800  
 tatactgccg cegtacatgt caatcagatg aacctgtgcg tatctcttaa cgacaattga 7860  
 20 cccacctttt taactgaagt gaaggggggt tctgctccgc gaccacttcc tggatctctc 7920  
 ccttcacctt ctgtgttctt tgggtgcac catcgggtca aagccgcagc aacgctctc 7980  
 ctgtgtgatc gcatgtgccc ttctgcacac gaccttcccc cgagagtgc cagctaccgg 8040  
 acaggcacca aggaggggta ccgagcacct cccggaccgg cggctgcagg atcgcgagcg 8100  
 cctccgctag ggagaccgca cgttccctg gtgtccctg cgttggcgcg tcttgcaagg 8160  
 agacctcgac cctgctccct ctccggggct ggatctgact ccttgacggt gattccagac 8220  
 25 gcgagacca aactgacggc ttctagaaga gggcgagcc cggccgcaag tctttcacgt 8280  
 agctaagtca tegtgtcttc cggcttctta cgttctccc ctttgtaaac ggttacctcc 8340  
 cgaaaacca ggctctctc caacagtggt tctcaagcga ggcgatcttc cccgggaggg 8400  
 gatatttggc aaagtctggg ggcatttttg gttcactggg gctgctactt gcatccactg 8460  
 ggtagagggc ggggatgcag ctacacaacc tgcaagcac gggacagcac cctcccac 8520  
 ccagacagaa tttagcggcc caaaacctca gtagtccca ggctgagaaa cctgcctta 8580  
 30 aacaacaac aaagaaaggc caagtcccat aagtgggtca ccgcgccgag actggggtcc 8640  
 acgggacacc ccagccacgc caagccggga agtccccgcc tcttgagct gaaccgcc 8700  
 ctctcccaga ggtggagctg cggggggcgg gaacaggcac ggagaaaata aacaagacta 8760  
 aaaagtctg agtagcgtg tgtggccgca aacctgaacc caccttttgc accacgggg 8820  
 acccggcact ctctctgcca cccacctctg agagggctgc gcggccgacc ccagtactag 8880  
 35 aaaacactcg tcacctcact caagacgggt acgaaggcca acggacgct ccttttagaa 8940  
 cgctcagcac acagagcaac ttctcagcc tactctcaaa tggcgtactc caaactagca 9000  
 ctcccagcgt ccagctgtga acccagagcg gcggaaagcc cctgaacca gcgcccggg 9060  
 atgcgagac gcgttgttgt ggtgggctg gctcccctcg gaccggcgc cccgcctcc 9120  
 gccccgtgc cgcagcgcg actgagccgg gtggatggta ctgctgcac cgggtgtctg 9180  
 gaggtgtgg ccgttttgtt ttcttggtca aaatcggggg agtgaggcgg gccggcgcgg 9240  
 40 c 9241

<210> 19  
 <211> 9236  
 <212> DNA  
 45 <213> Homo sapiens

<400> 19  
 ctagaaaaca ctttgtcatt ttagaggtgt tatccaatgt tcgcgaggc actggagtca 60  
 gagaaaatgg agttgaatcc tttctctgcc actctttgag gagaatctca ccatttatta 120  
 tgcactgtag aatacaaca taaaatacag ccatgtacca cataacaaca tcttggtaaa 180  
 50 caacagactg catatatgat ggtggctcag cagtaagcta aggttaattt attattattc 240  
 cctttttttt ttcttttttt tgagatgtag tcttactctg tcaccaggc tagagtcaa 300  
 tggcaccatc ttggctcact gcaacctctg cctcctgggt tcaagcgaat ctctgcctc 360  
 agcctccgaa gtagctggga attacaggca cccaccacat ctggctaatt ttttattt 420  
 ttagtaaaga tggggtttca ccatgttggc caggctgatc tcaaactcct gacctcaagt 480  
 gatctgcctg cctcggcctc ccaaagtgc gggaccatag gctgagcca ctgtgcccg 540  
 55 ccttgtttgc ttttttaaca gttaacagtg tgctcataga aactgctttg acatgactgc 600

EP 1 780 217 A1

	aatcatgtgc	ttcatagaaa	cttaattaga	ttataccact	agagtcttca	gatttttata	660
	ctttttttt	tgaaacggag	tctcactctg	tcaccaggct	ggagtgcagt	gccgcaatct	720
	cagctcgccg	caacctccgc	ctcccagggt	caagtgatcc	tcctgcctca	gcctcccag	780
	tagctgggat	tacaagtgca	cactaccacg	ccagctaat	ttttgcattt	ttactagaca	840
5	gggtttcacc	atggtggcta	ggatagtttc	accaggatct	cttggcctca	tgatcagcct	900
	gcctcggcct	cccaaagtgc	tgggattaca	gggtgagcc	accgtgcca	gcctatactt	960
	ccctttttga	ataccatttg	gcgttttgaa	gaat taacag	ctttgtgaa	gtggcattgc	1020
	ttgtgattca	ggcttccact	gagaccaagg	ggagaacctg	gttgcaggac	aaacagacgg	1080
	acagcgtgtg	gcagtgttta	aatgctcttc	tgaaggctga	tacgacagct	ctctgtgcac	1140
10	tgattgcata	cgcatcccaa	gattatatta	ttgttttcta	ttgctatgtg	tcacactttg	1200
	ccaaacagga	tgtggaaaat	gaataagcgg	ttttcttagg	cacttcttaa	cagacaattg	1260
	gtcaaaaatga	actccattgc	ttaagaaaca	cataaacacc	athtagtcac	tgaatatagc	1320
	tatatgtatg	gttgctacta	tggggaatct	tgttttgcca	atthtctttg	aaaattctg	1380
	cagaccaagg	ttctttttgt	ttacacaata	cttgaaaaat	aaaaatgaac	aagccaacaa	1440
	actaccaagt	tttcacttac	ataaatgtag	ttacatacag	aaaatgtgac	tgtgaatttt	1500
15	ttctaggact	tttaactat	aagcactatt	tgcacgaaag	agaaccaatc	tatcaattac	1560
	aaactcacat	aatthttacag	atthtttttt	ccctacacag	caataaaaac	agaaggaatt	1620
	tgaagccacc	ctccaaacac	aggggaagga	ggctgtgtgt	atctctctat	tgtctttcac	1680
	attctagggt	ggthccactc	agtgactgaa	atctctaaat	gttgattatg	tcgcttggg	1740
	ctaccataac	agcagcttaa	actgttggtta	agccactcag	acttaaacaa	cagaaattta	1800
	tttctttata	gttctggagg	ctggaagtcc	aagggtgccg	caaggctggt	ttctggtgag	1860
20	acctctctcc	ctgtcttgca	gatggctgcc	tcctccctgt	gtctctatag	agcctgtcct	1920
	ctgctttttac	acttctgggtg	tcactctcct	tttttttttt	tttttgagac	agagctctgc	1980
	tctatcgccc	aggtctggagt	gcagtggccc	gatcgatctc	ggctcactgc	aacctctgcc	2040
	tcccaggttc	aagcaattct	cctgcctcag	ctctccgagt	agctgggact	acaggtgcc	2100
	agatgcattg	ctggctaatt	tttgtatttt	tagtagagac	aggttctcac	catattggcc	2160
25	aggtctggtct	ccaactcctg	acctgtctat	ctgcctgcct	eggctccca	aagtgtctag	2220
	attacaggcg	tgagccaccg	caccggcct	cttctctctc	ttataaggac	accagctcta	2280
	ttagattagg	gctccaccct	catgacctca	tttagacctt	actattattt	ctthaaagca	2340
	cctattttcca	aatatagtca	ctttaggggt	tagggcttca	aaatatgaat	ctgagggaga	2400
	tcaattcagt	aaatagcagt	agtcattaac	ggacaatata	tacaagata	atthctgtat	2460
	tactgtcctt	atgcataaat	gtcctcagtg	ttccactgcc	tttatccaga	ttactatca	2520
30	caaagacttt	gctctgagaa	aaatgtgatt	tctttctttt	tttttttttt	ttgagacaga	2580
	gtctcactct	gtcaccaccg	ctggagtgca	gtggtgcaat	ctcggtcac	tgcaatctcc	2640
	gcctcccagg	ttcacgccat	tctcttgcc	cagtctcccg	agtagctggg	cctacaggcg	2700
	cccgccacc	tgcccagcta	atthttttgta	tttttagtag	agacggggtt	tcaccatggt	2760
	agccaggatg	gtctcaatct	cctgacctcg	tgatccacct	gectcagcct	cccaaagtgc	2820
	tgggtattaca	ggcatgagcc	accggcccca	gcagattttt	tttttttttt	tttttttttt	2880
35	agatggagtc	ttgctgtggt	gcccagcctg	gttgagcagtg	ttatgatttt	ggctcactgc	2940
	aacctctgtc	taccatgttc	aagcgattct	cccactctg	ctctccgtgt	agctgggac	3000
	acaggcacac	gccaccacac	ctagctactt	tttgtatttt	tagtagaaat	gggtttctac	3060
	catgttggtc	aggatggtcc	cgaactcctg	acctcaagtg	atctctctgc	ctcgccctcc	3120
	caaagtgtctg	ggattacagg	tgtgagccac	tgtgcctggc	caaaaatgtg	atthcttatt	3180
	tcccacattg	ccaattccat	ttcaat taac	tataatagct	atgtctattg	agcactcaag	3240
40	tgtattctag	aaactgttcc	tgattctggg	gatatacca	tgaatcaact	atagctcag	3300
	ttattaagta	atctgtagtc	tgactaaacc	attagaaatt	taaaaaatgg	ctactttcaa	3360
	agacatcttg	gagttcagga	gtcccacact	gcgaaccata	ttacctata	atccaactg	3420
	cttgtaattc	acttattttaa	ccaatattta	ttgagtcca	actttgagcc	taagatacag	3480
	cagtaaacaa	atggataaag	tcctctgctc	catgaaactt	gtattctaat	ggaagaaaca	3540
45	gaaaacaac	agatatagga	tgtaatatca	ggtagggata	aatactttga	attcaaacaa	3600
	aagtatacgt	agtcagggtt	cgccaaagag	acacagccaa	tcggatacat	agatatata	3660
	aagagggttt	atgagttaga	aagggtcac	atgattacag	aggctgagaa	gtcccacagc	3720
	agattgtctg	caagctggag	accagggat	actggttagca	tggctcagtc	caagtcccaa	3780
	agcctcagaa	tcaggaaagc	tgatgatata	atthcttagcc	caaaggcctt	agaaccccag	3840
	cggtgacgga	aaggctgatg	taggtcctgg	agtcctgaga	cccacagcc	tgggatcctg	3900
50	aatccaagg	gcaggaatgg	aagcgtgtat	tcagctcca	agagagtaag	accaatttgc	3960
	ctttcttccg	tttttgtttc	aagccacctg	caataggag	gcggatggtt	ccctcttagt	4020
	ccattcagtc	atatatcaat	ctctctgga	aataccctca	cagacacact	aacaaataat	4080
	gcctttccag	ttctctaggt	atthctttaa	ccagtcaagc	tgacacctaa	aatttaacct	4140
	cacaaaagtt	aaggagaaag	aagacaactt	gtaggggagg	ctgctatgca	agacagtggtg	4200
	tgaaggaagg	gctctctgaa	gaggttaata	tctgagcaga	gacttgaatg	aagtgaagaa	4260
55	gtgagccatg	tgggtatggg	gaatacaact	tcagggtaga	gaagacaagt	gtggtgtgta	4320
	tcagggtcag	caaagaagcc	atgtgacaga	gaagggtggg	ccaggggagag	acggataagt	4380

EP 1 780 217 A1

	gatctaactc	ctgaggaggt	ggcctggcca	ggagctagag	catgaagatc	tcgtaggact	4440
	ttattctgca	aggtgaaaag	ccattgtatt	agtctgttca	caaacccgag	actaggcaat	4500
	ttacaaaaga	aagagaggtt	taatggactt	acagttccac	atggctgggg	aggcctcaca	4560
5	atcatggcga	aaggcaatga	ggagcaagtc	acgtcttacg	tggatggcag	gcaaagacaa	4620
	agacagcttg	tgcagagaaa	ctccccctta	tagagccatc	agatcctggt	agacttattc	4680
	actatcacaa	gaacagcacg	ggtaagacct	gtccccatga	ttcagttacc	ccccactggg	4740
	tccctcccac	aacgcgatgg	aattcaggat	gagatttggg	tggggacaca	accaaaccct	4800
	atcattccac	ccatggcccc	tcccaaattt	catgtcctca	catttcaaaa	ccaatcacac	4860
	catcccaaca	gtccctcaaa	gtctttaaag	atltcagcat	taactcaaaa	gtccacagtc	4920
10	taatgtctca	tctgagacaa	ggcaagtcct	ttccatttat	gagcctataa	aatccaaagc	4980
	aagttagtta	cttcctagat	acaatggggg	tacaggcatt	gggtaaatac	agccattcca	5040
	aatgggataa	attgggtcaa	acaaagaggc	tacaggccca	tgagagtcca	aatccagtg	5100
	gggcagtcaa	atcttaaagc	tccaaaatga	tctcctttga	ctccacatct	cacatccag	5160
	tcacgcagat	ggaaggggtg	ggttcccatg	gtcttgggca	gctctgcccc	tgtacctttg	5220
	caggggtacag	cctccctctc	agctgctttc	atgggctggc	attgagtgtc	tgcagctttt	5280
15	ccaggtacac	ggtgcaagct	gtcgggtggat	ctaccattct	ggggtctgga	ggacctcttc	5340
	tcacagctcc	actaggtggt	gccccagtag	ggactgtgtg	tggggtctct	gacccacat	5400
	ttccctctctg	cactgcctctg	gcagaggatc	tccatgaggg	cctgtctct	gcagcaaat	5460
	tctcagttgg	catccaggca	tttccgcaca	tctctttaa	tctagggcaa	ggtttccaaa	5520
	ccccaattct	tgacttctgt	gcactgcag	tcctcaacacc	acatggaagc	tgtcaaggct	5580
	tggggcttgc	actccccgaa	gctacagccc	aagctctacc	ttgcctcccg	tcagtcatgg	5640
20	ttgggagtgg	ctgggatgca	gggcaccaag	tccctaggct	gcacacagca	tgaggacccc	5700
	gggcctggcc	aacaaaacca	ttttttcctg	atacctctgg	acctgtgatg	ggaggggttg	5760
	ccataaagac	ctctgacatg	ccctggagac	atlttcccc	ttgtcttggg	aattagcatt	5820
	tggctctctg	tactcatgca	aatltctgca	gccagcttga	atltctctct	agaaaatggg	5880
	aatlttctct	ttctatcaca	ttgtcaggct	gcaaattttc	cgaactttta	tgctctgctt	5940
	cccttataaa	actgaatgtc	tttaacagca	cccaagtcac	ctcttgaatg	ctttgtgct	6000
25	tagaaatltc	tcctgccaga	tactctaaat	catctctctg	aagttcaaag	ttctacaaat	6060
	atctcgtgca	ggggcaaaat	gccgccagta	tctttgctaa	aacataacaa	gagtccccct	6120
	tgctccagtt	cccaacaagt	tcctcatttc	cgtctgagac	cacctcagcc	tatggacttt	6180
	atgtccaca	gtgctatcag	cattttgggc	aaagccattc	aacaagtctc	taggaagttc	6240
	caacttttcc	cacatttgcc	tgtcttcttc	tgagccctcc	aaactgttcc	aaacctgccc	6300
	tgttaccag	ttccaaagtc	acatacccat	ttttgagtat	ctacggcagc	acccactct	6360
30	actggtacca	atlttagccac	tgaagtagtt	ggagaacaga	agtaatagac	tctggtttac	6420
	atltgaaaag	cttctctgtg	gctgctgtgt	gaagaaaata	tatgagaatg	aagccccaa	6480
	atgaagcagg	gacacagttg	cagtggttag	agtaagaaat	gctgctggct	ggcactgaa	6540
	tgatagcctg	gaggtttgtg	tgtgcacatg	catgtgtatg	tgttttacga	tagtaggccc	6600
	aacagatact	gtaatccaca	cttgtttttt	ttttttgaga	cagagtctca	cctggttgcct	6660
35	agactagaat	gcagttggc	aatcttggct	cactacaacc	tccacctccc	aggttcaaac	6720
	aatccttgtg	cttcagcctc	ccgagtagtt	gggattacag	gtgtgtgcca	ccgtgcccag	6780
	ctatatlttt	tgtatltttt	gcagagatgg	gattttgcca	cattggccag	gctggtcttg	6840
	aactcctggc	ctcaagcaat	cctcccacct	tagcctccca	aagtgtgag	ccaccacacc	6900
	tggcgcgaac	tgatltttta	tcatgaaatg	acacatacat	ttaaaaaac	caatacctat	6960
	aatatltctg	gctagtactc	ttcacatcta	tatcatcaaa	aacaaagaaa	gtatgtgaaa	7020
40	ctgacacagc	caaggggaga	ctaaggagac	ataacaatta	actgtaatgt	ggatltctgg	7080
	aggggatcct	ggaacagaaa	aagacattag	gcaaaaaact	aaagaaatct	gaataaaatg	7140
	tggatgtcag	ttataataaa	tgtatcatat	tagtccagta	attgtaacaa	atataccaca	7200
	ataatgaaag	ccattaatta	tagggaaaat	ggaggggtta	atatgggtgg	ctggcttttg	7260
	ctatltctag	cagctccatt	ttatctgcaa	aagacaaaaca	ttcattaagt	cccaaaaagg	7320
	taaagaatga	caaattaagc	atgtatctta	ttagtaagag	taataataaag	atgctcactc	7380
45	ctatltataa	atatlttgaca	atcatgttaa	ggccacaaaa	gagaaaaaag	ggtaggggca	7440
	aaaaacgcaa	agagaaaagga	gttagtatct	tttctcccgc	actcattagc	tattaaaga	7500
	ggatgtttgt	ttaaagctgc	tcagagctgg	taaactaatg	ttaagtcaact	aacgggaatt	7560
	taaaaggttt	cattaagaac	tgcttgcact	agattctctc	accctgagac	attaacaact	7620
	cacgataaac	ctcctgagtg	gtaagaactt	gtccatttaa	aaacaggcta	tagattgtat	7680
	catgcagttt	tatctactaa	tcggctaaata	tcccgcctaaa	aacaaaaaac	cccaagggga	7740
50	tgaaaagttc	atccatcaaa	ggaaacaaca	gtcaccttgg	ttcccatctc	actcatatac	7800
	tgcccgcta	catgtcaact	agatgaacct	gtcgttatct	cttaatgaca	atgacccac	7860
	atltttgact	gaagtgaag	ggggttctgc	tgcgcgacca	cttctggat	ctccccctcc	7920
	accctctgtg	ttctttcggg	tgcaccatcg	ggtcaaagcc	gcagcaacgc	cgtctctgtg	7980
	tgatcgcatg	tgcccttctg	cacacgacct	tccccogaga	gtgaccagct	accggacagg	8040
	caccaaggag	ggctaccgag	caoctcccgg	accggcggct	gcaggatcgc	gagcgcctcc	8100
55	gctagggaga	ctgcacgttg	cgctgtgct	tcttggggtg	gcrccttctg	caaggagacc	8160

EP 1 780 217 A1

tcgaccctgc tccctctccg gggctggatc tgactccttg acggtgatto cagacgcgag 8220  
 acccaaaactg acggcttcta gaagaggggc gagcccgcc gcaagtctt cactagctta 8280  
 agtcatcggt gcttccggct tcttaccgtt ctcccccttg taaacggta cctccccgaa 8340  
 acccaggctc tcctccaaca gtgggttctca agcagggcga tcttccccgg gaggggatat 8400  
 5 ttggcaaagt ctgggggcat ttttggttca ctggggctgc tacttgcatc cactgggtag 8460  
 agggcgggga tgcagctaca caacctgcga agcagggac agcaccctcc ccaaccaga 8520  
 cagaattagc cggcccaaaa cctcagtagt gccaggctg agaaccctg ccttaaaaa 8580  
 acaacaaaga aaagccaagt cccataagtg ggtaaccgag ccgagactgg ggtccacggg 8640  
 acaccccagc cacgccaagc cgggaagtcc ccgctctctg gagctgaacc cgcccctctc 8700  
 ccagagggtg agctgcgggg ggcgggaaca ggcaaggaga aaataaaca gactaaaaag 8760  
 10 tcctgagtag cgctgtgtgg ccgcaaacct gaacccacct tttgaccac gcgggaccgg 8820  
 gcacgcttcc tgcacccac ccctgagagg gctgcggggc cgacccagt actagaaaac 8880  
 actcgtcacc tcaatcaaga cgggtacgaa gccaacgga cgcttctct tagaacgctc 8940  
 agcacacaga gcaacttctc acgcctactc tcaaatggcg tactccaaac tagcactccc 9000  
 gacgtccagc tgtgaacca gagcggcgga aagcccctga acccagcgcc cgggcatgag 9060  
 15 cagacgcggt gttgtgggtg gcgtggctcc ctccggacc ggcccccgc cctccgccc 9120  
 gtgtccgcat gcgcgactga gccgcggggg tggtagctg gcatccgggt gtctgaagat 9180  
 ccgatgaaat aacatatgca aatgattgg gtccgtgatt ggcattccag aatgg 9236

<210> 20  
 <211> 9238  
 <212> DNA  
 <213> Homo sapiens

<400> 20  
 aaacgctcat gagacgaaag tctccaatgt tcgcgaggc actggagtca gagaaaatgg 60  
 25 agttgaatcc tttctctgcc actctttgag gagaatctca ccatttatta tgcactgtag 120  
 aatacaaca taaaatacag ccatgtacca cataacaaca tcttggtaaa caacagactg 180  
 catatatgat ggtggctatc cagtaagcta aggttaattt attattattc cttgtttttt 240  
 tttttttttt tttttttttg agatgtagtc ttactctgtc acccaggcta gagtgcaatg 300  
 gcaccactct gctcactgc aacctctacc tctggggtc aagcaaatct cctccctcag 360  
 30 cctccaaagt agctgggatt acaggcacc cccacatctg gctaattttt tgtattttta 420  
 gtaaagatgg ggtttcacca tgttggccag gctgatctca aactcctgac ctcaagtgat 480  
 ctgcccgcct cggcctccca aagtgtgga accacaggcc tgagccactg tgcccagcct 540  
 tggttgcttt ttaacagat aacagtgtgc tcatagaaac tgctttgaca tgactgcaat 600  
 catgtgcttc atagaaactt aattagatta taccactaga gtcttcagat ttttatactt 660  
 tttttttttg aaacggagtc tcaactctgtc accaggctgg agtgcagtgc cgcaactctg 720  
 35 gctcactgca acctccgcct cccaggttca agcaattctc ctgcctcagc ctcccagta 780  
 gctgggatta caagtgcgca ctaccacacc cagctaattt ttgcattttt acttgacagg 840  
 gtttcacat gttggctagg atagtttcac caggatctct tggcctcatg atcagcctgc 900  
 ctccggcctcc caaagtgtg ggattacagg tbtgagccac cgtgcccagc ctatacttcc 960  
 ctttttgaat accatttggc gttttgaaga ataacagct ttgtgaacgt ggcagtgctt 1020  
 gtgattcagg ctctcattga gaccaagggg agaacctggg tcaggacaa acagacggac 1080  
 40 agcgtgtggc agtgtttaaa tgccttctg aaggctgata cgacagctct ctgtgactg 1140  
 attgcatacg catcccaaga ttatattatt gttttctact gctatgtgtc acactttgcc 1200  
 aaacaggatg tggaaaatga ataagcgggt ttcttaggca cttcttaaca gacaattggg 1260  
 caaaatgaac tccattgctt aagaacaca taaacacat ttagtactg aacatagcta 1320  
 tatgtatggt tgttactatg ggaaatcttg ttttgccaat tttctttgaa aattctggca 1380  
 gaccaagggt ctttttgttt acataatact tgaaaaaata aatgaacaa gctaacaac 1440  
 45 taccaagttt tcaactacat aaatgtagtt gcatacagaa aatgtgactg tgaattaatt 1500  
 tttctaggac ttttaacta taagcactat ttgcacaaaa gagaaccaat ctatcaatta 1560  
 caaactcaca taattttaca gatttttttt ttctacaca gcacataaaa cagaaggaat 1620  
 ttgaagccac cctccaaaca caggggaagg aggctgtgtg tatacctca ttgtctttca 1680  
 cattctaagg tggttccact cagtactga aatccttaag cgttgatta gtctgcttgg 1740  
 gctaccataa cagcagctta aactgttgtt tagccactca gacttaaaaa acagaaattt 1800  
 50 atttcttat agttctggag gctggaagtt caagggtgccc gcaagggttg tttctggtga 1860  
 gacctctctc cctgtcttgc agatggctgc ctccctctg tgcctcata gagcctgtct 1920  
 tctgctttta cacttctggt gtcactctcc tttttttttt tttttttttt ttttgagaca 1980  
 gagtctcgct ctatcgccca ggctggagtg cagtggccc atcgatctcg gctcactgca 2040  
 acctctgcct cccaggttca agcaattctc ctgcctcagc ctccaagta gctgggacta 2100  
 cagggtgccg ccatcatgcc tggctaattt ttgtattttt agtagagaca gggtttacc 2160  
 55 atattggcca ggctggcttc gaactctga cctgtctatc tgctgcctc ggcctcccaa 2220

EP 1 780 217 A1

agtgctagga ttacaggcgt gagccaaccg acccggcctc ttctctttct tataaggaca 2280  
 ccagtcctat tagattaggg ctccaccctc ataacctcat ttgaccttaa ctattatttc 2340  
 tttaaagcac ctatttccaa atatagtcac tttagggggt agggcttcaa aagatgaatc 2400  
 5 tgagggagct caattcagta aatagcagta gtcathtaag gacaatgtat acaaagataa 2460  
 tttcgtgatt actgtcctta tgcataaacg tcctcagtggt tccactgcgt ttatccagat 2520  
 ttagtatcac aaagactttg ctctgagaaa aatgtgattt tttttttttt tttttttttg 2580  
 agacggagtc tcgctctggt acccaggctg gagtgcagtg gcgcgatctc ggctcactgc 2640  
 aagctccgcc tcccgggttc acgccattct cctgcctcag cctccggagt agctgggact 2700  
 acaggcgccc gccactacgc ccggctaact tttttgtatt tttagtagag acgggggttc 2760  
 10 accatgttag ccaggatggt ctcaatctcc tgacctcgtg atccacctgc ctgagcctcc 2820  
 caaagtgctg ggattacagg catgagccac cgcgcccagc agattttttt tttttttttt 2880  
 gagatggagt cttgctctgt tgcccaacct ggagtgcagt gttatgattt tgggtcactg 2940  
 caacctctac catgttcaag cgattctccc acctctgcct cccgtgtagc tgggatcaca 3000  
 ggacacagcc accacaccta gctacttttt gtatttttag tagaaatggg gtttcacat 3060  
 gttggccagg atgggtcccga actcctgacc tcaagtgatc ctctgcctc ggctccaaa 3120  
 15 gtgctgggat tacaggtgtg agccactgtg cctggccaaa aatgtgattt cttatttccc 3180  
 acattgccaa ttccatttca attaactata atagctatgt ctattgagca ctcaagcgt 3240  
 ttctagaaac tgttctctgat tctggggata tatccatgaa tcaactatag tccctgttat 3300  
 taagtatacc qtaqtctgac taaaccatta gaaatataaa aaaaatggct actttcaaag 3360  
 acatcttgya gttcaggagt cccacactgc gaacatatt acctaataat ccaacctgct 3420  
 tgtaattcac ttatttaacc aatatttatt gagtgcacac tttgagccta agatacagca 3480  
 20 gtaaacaaat gcataaagtc cctgtcccca tgaacttgt attctaagg aaaaaacaga 3540  
 aaacaacagc atataggatg taatatcagg tagggataaa tactttgaat tcaaacaaaa 3600  
 gtatacgtag tcagggttcc ccagagagac acagccaatc gatacataga tatataagag 3660  
 agggtttatg agttagaaag ggctcacatg attacagagg ctgagaagtc ccacaacaga 3720  
 ttgtctgcaa gctggagacc cagggactat ggtagcatgg gtagtccaa gtcccaaaag 3780  
 ctcagaatca ggaaagctga tgatataaatt cttagcccaa aggccttaga accccagcgg 3840  
 25 tgacggaaag gctgatgtag gtcctggcgt cctgagacc aacagcctgg gatcctgaaa 3900  
 tccaagggca ggaatggaag cgtgtattcc agtccaaga gtagtaagacc aatttgctt 3960  
 tcttccgttt ttgttccaag ccaactgcac gttgagggcg gatggttccc tcttagccca 4020  
 ttcagtcata tatcaatctc ttctggaaat accctcacag acacactaac aaataatgcc 4080  
 tttccagtctc tctaggtatt ctttaatcca gttcaagctga cacctaaaat taacctcac 4140  
 aaaagttaag gagaaagaag acaacttgta ggagaggctg ctatgcaaga cagtgtgta 4200  
 30 aggaagggct ctctgaagag gttaatatct gagcagagac ttgaatgaag tgaagaagt 4260  
 agccatgtgg gtatggggaa tacaacttcc aggtagagaa gacaagtgtg gtgtgtatca 4320  
 cggtcagcaa agaagccatg tgacagagaa ggttgggcca gggagagacg gataagtgat 4380  
 ctaactctg agggagtggc ctggccagga gctagagcat gaggatctcg taggatttta 4440  
 ttttgcaagg tgaagagcca ttgtattagt ctgttcaca accccagact aggcaattta 4500  
 35 caaaagaaag agaggtttaa tggattaca gttccacatg gctggggagg cctcacatc 4560  
 atggcgaaag gcaatgagga gcaagtcacg tcttacgtgg atggcaggca aagacaaaga 4620  
 cagcttgtgc agagaaactc ccccttatag agccatcaga tctgttaga cttatcacta 4680  
 tcacgagaac agcacgggta agacctgtcc ccatgattca gttacctcc actgggtccc 4740  
 tcccacaacg catgggaatt caggatgaga tttgggtggg gacacaacca aaccctatca 4800  
 tttccaccat ggcccctccc aaatttcatg tctcaccatt tcaaaaacca tcacaccatc 4860  
 40 ccaacagtcc ctcaaagtct taaatgattt cagcattaac tcaaaagtcc acagtcta 4920  
 gtctcatctg agacaaggca agtcccttcc gctatgagc ctataaaaatc caaagcaagt 4980  
 taattacttc ctagatacaa tgggggtaca ggcattgggt aaatacagcc attccaaatg 5040  
 ggataaatg gtcaaaaaca agaggctaca ggccatgag agtccaaaat ccagtggggc 5100  
 agtcaaatct taaagctcca aatgatctc ctcttgactc cacatctcac atccaggtca 5160  
 tgcagatgga aggggtgggt tcccattggtc ttgggcagct ctgcccctgt acctttgcag 5220  
 45 ggtacagcct cctctcagc tgccttcatg ggctggcatt gagtgtctgc aactttcca 5280  
 ggtacagcgt gcaagctgtc ggtggatcta ccattctggg gtctggagga cctcttcca 5340  
 cagctccact aggtggtgcc ccagtaggga ctgtgtgtgg ggtctctgac cccacatttc 5400  
 ccttctgca tgccttgcca gaggatctcc atgagggccc tgcccctgca gcaacttct 5460  
 gcctgggcat ccaggcattt ccgcacatcc tctttaatct aggcgaaggt ttccaaacc 5520  
 cagttcttga cttctgtgca ctgcagctc caacaccaca tggaaagctgt caaggcttgg 5580  
 50 ggcttgact ccccgaagct acagcccaag ctctacctg cctcctgtca gtcatgggtg 5640  
 ggagtggctg ggaagctcc caccaagctc ctaggctgca cacagcatga ggacccggg 5700  
 cctggccaac aaaccattt tttcctgata tctctggacc tgtgatggga ggggttcca 5760  
 taaagacctc tgacatgccc tggagacatt ttccccattg tcttgggaat tagcatttgg 5820  
 ctctgttac tcatgcaaat ttctgcagcc agcttgaatt tctcctcaga aaatgggaat 5880  
 ttttctttc tatcacattg tcaggctgca aattttccga acttttatgc tctgctccc 5940  
 55 ttataaaact gaatgtcttt aacagcacc aagtcaacct ttgaatgctt tgctgcttag 6000

EP 1 780 217 A1

aaatctctcc tgcagatac tctaaatcat ctctctgaag ttcaaagttc taaaaatc 6060  
 tegtgcaggg gcaaaatgcc gccagtatct ttgtctaaac ataacaagag tcccctttgc 6120  
 tccagttccc aacaagttcc tcatttccgt ctgagaccac ctcagcctat ggactttatt 6180  
 5 gtccacagtg ctatcagcat tttgggcaaa gccattcaac aagtctctag gaagttccaa 6240  
 actttcccac atttgccctgt cttcttctga gccctccaaa ctgttccaaa ccctgcctgt 6300  
 taccagttc caaagtcaca taccattttt tgagtatcta cggcagcacc ccactctact 6360  
 ggtaccaatt tagccactga agtagttgga gaacagaagt aatagactct ggtttacatt 6420  
 gtaaaagctt ctctgtggct gctgtgtgaa gaaaatata gagaatgaag cccaagatg 6480  
 aagcagggac acagttgcag tggtagagt aagaaatgct gctggctggc actgaagtga 6540  
 10 tagcctggag gtttgtgtgt gcacatgcat gtgtatgtgt tttacgatag taggcccac 6600  
 agatactgta atccacactt gttttttttt tttttttgag acagagtctc acctgttggc 6660  
 tagactagaa tgcagtggca caatcttggc tctactacaac ctccacctcc caggttcaaa 6720  
 caatccttgt gttcagcct cccgagtagt tggattaca ggtgtgtgoc aacgtgcca 6780  
 gctatatctt ttgtattttt agcagagatg tggatttggc acattggcca ggctggtctt 6840  
 gaactcctgg cctcaagcaa tctctccacc ttagcctccc aaagtgtgta gccaccacac 6900  
 15 ctggccgcaa ctgattttta atcatgaaat gacacataca tttaaaaaac ccaataccta 6960  
 taatattcct ggctagtact ctccacatct atatcatcaa aaacaaagaa agtatgtgaa 7020  
 actgacacag ccaaggggag actaaggaga cataacaatt aactgtaatg tggattctg 7080  
 gaqqqqatcc tqqaacqaa aaagacatta ggcaaaaaac taaagaatc tgaataaaa 7140  
 gtggatgtca gtttaataata atgtatcata ttagtccagt aattgtaaca aatataccca 7200  
 ataatgaaag ccattaatta tagggaaaat ggagggggtta atatgggtgg ctggcttttg 7260  
 20 ctatttctag cagctccatt ttatctacaa aagacaaaca ttcattaagt cccaaaaagg 7320  
 taagaatga caaatgaagc atgtatctta ttagtaagag taatataaag atgctcactc 7380  
 atatttataa atatttgaca atgatgttaa ggccagaaaa gagaaaaaag ggtaggggca 7440  
 aaaaacgcaa agagaaagga gttagtatct tttctccgc actcattagc tattaanaa 7500  
 ggatgtttgt ttaaagctgc tcagagctgg taaactaatg ttaagtcact aacgggaatt 7560  
 25 taaaagttt cattaagaac tgcctgcaact agattcctcc acctgagac attaaacaat 7620  
 cacgataaac ctctgagtg gtaagaacgt gtccatttaa aaacaggcta tagattgtca 7680  
 tgcagtttta tctactaatc ggctaagca ccgcaaaaa caacaaaaa aacccaaagg 7740  
 gatgaaagtt tcatccatca aaggaaacaa cagtcaactt ggttccatc ccaactcat 7800  
 actgccgccc tacatgtcaa tcagatgaac ctgtgcgtat ctctaatga caattgacct 7860  
 acctttttaa ctgaagtga ggggggttct gctccgagc cacttctgg atctctctc 7920  
 30 tcacctctg tgttcttctg ggtgcaccat cgggtcaaaag ccgcaagcaac ccgctctctg 7980  
 tgtgatcgca tgtgcccttc tgcacacgac ctcccccgga gagtgaccag ctaccggaca 8040  
 ggcaccaagg agggctaccg agcacctccc ggaccggcgg ctgcaggatc gcgagcgcct 8100  
 ccgctagggg gaccgcacgt tgcgcctgtg cttcctgccc tggcgcttc tgcaaggaga 8160  
 cctcgaccct gctccctctc cggggctgga tctgactcct tgacggtgat tccagacgcg 8220  
 agacccaaac tgacggcttc tagaagaggg gcgagccgg ccgcaagtct ttcagtagc 8280  
 35 taagtcatcg ttgcttccgg ctctctaccg tttctccctt tgtaaaccgt tacctcccga 8340  
 aaaccaggg tctcctccaa cagtgggtct caagcgaggg gatcttcccc gggaggggat 8400  
 atttggcaaa gtctgggggc atttttgggt cactggggct gctacttga tccactgggt 8460  
 agaggcgggg gatgcagcta cacaacctgc gaagcacggg acagcacctt cccaaccca 8520  
 gacagaatta gccggcccaa aacctcagta gtgccaggc tgagaaacct tgccttaaac 8580  
 aaacaacaaa gaaaggccaa gtcccataag tgggtcaccg cgcagagatc ggggtccacg 8640  
 40 ggacacccca gccacgcccag gccgggaagt ccccgctcc tggagtga aacggccctc 8700  
 tcccagaggt ggagctgccc ggggcgggaa caggcacgga gaaaaataaac aagactaaaa 8760  
 agtctgaggt agcgtgtgtt ggccgcaaac ctgaacccac cttttgcacc acgcgggacc 8820  
 cggcactctt cctgccacc acccctgaga gggctgccc gccgaccca gtactagaaa 8880  
 aactctgca cctcactcaa gacgggtacg aaggccaacg gacgcctcc tttagaacgc 8940  
 tcagcacaca gagcaacttc tcacgcctac tctcaaatgg cgtactccaa actagcactc 9000  
 45 ccgacgtcca gctgtgaacc cagagcggcg gaaagcccct gaaccagcg cccgggcatg 9060  
 cgcagacggc ttgttgggtt gggcgtggct cctccggac cggcgcccc gccctccgcc 9120  
 ccgtgtccgc atgcgcgact gagccgggtg gatggtactg ctgcatccgg gtgtctggag 9180  
 gctgtggccc ttttgttttc ttggctaaaa tccggggagt gaggcgggccc ggcggcgc 9238

50 <210> 21  
 <211> 1249  
 <212> DNA  
 <213> Homo sapiens

55 <400> 21  
 ctgatctaca taggaattgt tttcaagaca tttctgcatt cctctagtga cagggtgctc 60

EP 1 780 217 A1

actacctcat gagtatttca gtggacaact gtaatgggtca ataaagtatc cactttccac 120  
 ctccctgcag ctccctggccc tggttttatt ctctggggct ccacacattc agttttacact 180  
 cagtggccag tggctggggac cattgtagaa aataaggaaa ctccaattcc ttccttcttt 240  
 5 tcttctctct tcatctcttc ctccctctct acatccctct ctctcttcc tcttctctcg 300  
 acacttacca tgtaccagac cttctgccag gcacatggat gggagcacag gggaagttgg 360  
 ctgcagggtt agaactaagt cccaagcccc cttaaagctca tgccagggga ctggactgtc 420  
 cagtactgag ggatggggat gctgaggctg gtggccttcc tcaaatgcac tgtagtgcc 480  
 caggcagagt cctgggctgc cctgtgagga ggtgaccaga ggtagagcaa cttcacccca 540  
 aggctggatc aggatccccct ccagggttttt actagagcca aaccacatc tcttttctct 600  
 10 tctgccacc ccccttaaaa tgcttagaaa cacatagatt taaatacaaa ttcaaatgta 660  
 agtaatttca actgtgtaac tatgaggagt cagtctctacg tgggtcctat ctgtatcctc 720  
 cccagggctc agctccattc tttgctttca ttcattctca ttcaatacat tgtttgtaag 780  
 agctcactgg gtgccctctc tgtcatgtag taaggtttta aaaagaaagc ctcttctgag 840  
 cttcagtttc cttattcata aataggagt attgatccgt tcttgcctt tcttacaagg 900  
 atatgctgaa gatgactgaa gtacagagta aagaaggatt atgtttgggt gtcaaaggaa 960  
 15 tagaatgcc tctttcaaac tgagcacagc aggaacctgt aacaggaaca cagcaacttg 1020  
 ttgaatgaat gacaatattg gaaaacatac atttctctcc ctccccatca tagtccctct 1080  
 gcttccgtgt taactccata gagaggccag cacaaccagc cttgcagcct gagataaggc 1140  
 ctttgggggg tgtctccctc atcgtccctc caagccctca aqtagggtgt ggagagaggg 1200  
 gtgatgcctg gtgctgggtg aacccttgca cagagacgga cacaggatg 1249

20 <210> 22  
 <211> 699  
 <212> DNA  
 <213> Homo sapiens

25 <400> 22  
 ctagagaggg aagtttttga aaattaaaca ctgtctaatt ttctgcaaag tttttattca 60  
 tgaattaaga gtatttccct tagtccatta ttccaaggc aaatatggaa gtttgatcat 120  
 atgctaatac tactaaagct ggattctctt taagagattg agaaattaa aggcaaaagc 180  
 tgataatca tgtttagtta tactgtgagt cttataagaa gctgggaggg aaccctatta 240  
 30 actcaccaga atacagaact cagtctcaca acttaaataa aattcctctc aaaccttttc 300  
 ctcaaagtta aattctgaaa ataactttgt gattaagaga agaaggctgt ccaccaatgg 360  
 acttatctgt tatttcttcc ttattgtgag cttaatggca tgacaaagca gaggcaaaga 420  
 ggcatcacatc aattcttcaa agtaggaagt caaaaaggtc agagcttcca cagcatggca 480  
 acagctttgc agatgcccac atcgtgatag ttgaaatagc aaagcccagc aaaggttaaa 540  
 gctgaaaatg ccaaaagccc tgcttggca ccttctgagc aggcattccc atgaacatag 600  
 35 tcagtaacaa cttgtccaag gccccagtga ccatgaagag tgagggctgc agccagggaa 660  
 tagtccgtcg cagagcaagg attcaaataa gcagccgga 699

40 <210> 23  
 <211> 699  
 <212> DNA  
 <213> Homo sapiens

45 <400> 23  
 ctagagaggg aagtttttga aaattaaaca ctgtctaatt ttctgcaaag tttttattca 60  
 tgaattaaga gtatttccct ttgtccatta ttccaaggc aaatatggaa atttgatcat 120  
 gtactaatca tactaaagct ggattctctt taagagattg agaaattaa aggcaaaagc 180  
 tgatatgtca tgtttagtta tattgtgagt cttataagaa gctgggaggg aaccctatta 240  
 actcaccaga atacagaact cagtctcaca acttaaataa aattcctctc aaaccttttc 300  
 ctcaaagtta aattctgaaa ataactttgt gattaagaga agaaggctgt ccaccaatgg 360  
 acttatctgt tatttcttcc ttattgtgag cttaatggca tgacaaagca gaggcaaaga 420  
 50 ggcatcacatc aattcttcaa agtaggaagt caaaaaggtc agagcttcca cagcatggca 480  
 acagctttgc agatgcccac atcgtgatag ttgaaatagc aaagcccagc aaaggttaaa 540  
 gctgaaaatg ccaaaagccc tgcttggca ccttctgagc aggcattccc atgaacatag 600  
 tcagtaacaa cttgtccaag gccccagtga ccatgaagag tgagggctgc agccagggaa 660  
 tagtccgtcg cagagcaagg attcaaataa gcagccgga 699

55 <210> 24

EP 1 780 217 A1

<211> 699  
 <212> DNA  
 <213> Homo sapiens

5 <400> 24  
 cttagagaggg aagtttttga aaattaaaca ctgtctaatt ttctgcaaag tttttattca 60  
 tgaattaaga gtatttcctt ttgtccatta ttccaaggc aaatatggaa atttgatcat 120  
 gtactaatca taataaagct ggattctctt taagagattg agaaattaa aggcaaaagc 180  
 tgatataatca tgttttagtta tattgtgagt cttataagaa gctgggaggc aacccatta 240  
 actcaccaga atacagaact cagtctcaca acttagatat aattcctctc aaaccttttc 300  
 10 ctcaaagatt aaattctgaa aataatcttg tgattaagag aagaaggctg tccaccaatg 360  
 ggcttatctg ttatttcttc cttattgtga gcttaatggc atgacaaagc agaggcaaag 420  
 aggcatacat caattcttca aagtaggaag tcaaaaaggt cagagcttc acagcatggc 480  
 aacagctttg cagatgcca catcgtgata gttgaaatag caaagcccag caaaggtaa 540  
 agctgaaaat gccaaaagcc ctgccttggc agctttctgc gaggcattcc catgaacata 600  
 15 atcagtaaca acttgttcaa ggccccagtg accatgaaga gtgagggctg cagccgggaa 660  
 tagtccgctg cagagcaag attcaaataa gcagccgga 699

<210> 25  
 <211> 994  
 <212> DNA  
 <213> Homo sapiens

20 <400> 25  
 ttgtatctct ttttacagct acctccatt tccttctat ttcaagctag taacacagtt 60  
 ttcttttaaa ttcatttatt taaatgtaaa aataagtcta tttggagaaa aaaaattttt 120  
 25 aatagcatct ctggaatgcc agtatggcta aattcatgaa tgttgcctc aaatgctgaa 180  
 atctgggaag catctggcca agctttgtgg acaggcctgc ctagttagaa tccaagagc 240  
 cactcattcc gagccacaaa acattggaat tcttggttca ctccctaac ctgaacttgt 300  
 cctctgtgaa atagggacat taatagctca ctcacagggc tgctgtgagg acatgtgtg 360  
 agctgagggc ctgcccaggg gagaccctgt gcagggagac tgttatcatg gtgatggatt 420  
 30 tctgcttcat tcatttcttt ttccagacag catcatatag aatgagttgt ggggtggcag 480  
 tcagcaggtt tgggtttatc ctctattctg ccacttatta cttaaaaaaa aaaacccaac 540  
 ttatatagta taagctatat ccagaaaagt gcaaatatca tacaagtacc atttgatgaa 600  
 tcttctgata tccccacata accaacaccc agaacctctt ctgtctcat tccaggataa 660  
 ccactaacct gacttctaac agcatcagtc agttttgtct gttttgtac attatataatg 720  
 tgatggtttg aatgtgtccc ccaaatttca tggctagaa acttaactct tcaattcata 780  
 35 tgttgatgct atttggagga agggcctttg ggaagtaatt aggattagat aaggtcatg 840  
 ggtgaggtat gatggcactg gtgacttata agaagagaaa gagaaatctg agctggcatg 900  
 ctcttgcctc ctaccctgtg gatgacttct ccatgtcatg atgcagcaag aaggccctca 960  
 ccagatggtg gcaccatgct tttggacttc ccag 994

<210> 26  
 <211> 1000  
 <212> DNA  
 <213> Homo sapiens

40 <400> 26  
 ttgtatctct ttttacagct acctccatt tccttctat ttcaagctag taacacagtt 60  
 ttcttttaaa ttcatttatt taaatgtaaa aataagtcta tttggagaaa aaaaattttt 120  
 45 aatagcatct ctggaaggcc agtatggcta aattcatgaa tgttgcctc aaatgctgaa 180  
 atctgggaag catctggcca agctttgtgg acaggcctgc ctagttagaa tccaagagc 240  
 cactcattcc gagccacaaa acattggaat tcttggttca ctccctaac ctgaacttgc 300  
 cctctgtgaa atagggacat taatagctca ctcacagggc tgctgtgagg acatgtgtg 360  
 50 agctgagggc ctggtcaggg gagaccctgt gcagggagac tgttatcatg gtgatggatt 420  
 tctgcttcat tcatttcttt ttccagacag catcatatag aatgagttgt ggggtggcag 480  
 tcagcaggtt tgggtttatc ctctattctg ccacttatta cttaaaaaaa ccccaaaaaa 540  
 cccaacttat atagtataag ctatatccag aaaagtgcaa atatcataca agtaccattt 600  
 gatgaatctt ctgatatccc cacataacca acaccagaa cctcttcttg tctcattcca 660  
 55 ggataaccac taacctgact tctaacagca tcagtcagtt ttgtctggtt ttgtacatta 720  
 tatatgtgat ggtttgaatg tgtccccc aaattcatgtg ctggaaactt aatccttcaa 780

EP 1 780 217 A1

ttcatatggt gatgctatgt ggaggaaggg cctttgggaa gtaattagga ttagataagg 840  
 tcatgggggt aggtatgatg gcaactgggtg cttataagaa gagaaagaga aatctgagct 900  
 ggcattgctct tgcctctca ctgtgtgatg acttctccat gtcattgatgc agcaagaagg 960  
 cctcaccag atggtggcac catgcttttg gacttcccag 1000

5

<210> 27  
 <211> 999  
 <212> DNA  
 <213> Homo sapiens

10

<400> 27  
 ttgtatctct ttttacagct acctcccatt tcccttctat ttcaagctag taactcagtt 60  
 ttcttttaaa ttcaattatt taaatgtaaa aataagtcta tttggagaaa aaaaatttta 120  
 atagcatctc tggaaatgcca gtatggctaa attcatgaat gttgtcctca aatgctgaaa 180  
 tctgggaagc atctggccaa gctttgtgga caggcctgcc tagtttgaat cccaagagcc 240  
 acccagtcca agccacaaaa cattggaatt cttggttcac ttccctaacc tgaacttgcc 300  
 ctctgtgaaa tagggacact aatagctcac tcacagggct gctgtgagga catgtgttga 360  
 gctgagggtc tggcagggtg agacctgtg caggagact gttatcatca tcatgattt 420  
 ctgcttcatt catttctttt tccagacagc atcataaga atgagtgtg ggggtggcag 480  
 cagcaggttt gggtttatcc tctattctgc cacttattac ttaaaaaaac ccaaaaaaac 540  
 ccaacttata tagtataagc tataaccaga aaagtgcaaa tatcatacaa gtaccatttg 600  
 atgaatcttc tgatatcccc acataaccaa caccagaac ctcttctgt ctcatccag 660  
 gataaccact aacctgactt ctaacagcat cagtcagttt tgtctgtttt tgtacattat 720  
 atatgtgatg gtttgaatgt gtccccaaa ttcatgtgc tggaaactta atccttcaat 780  
 tcatatggtg atggtttttg gaggaagggc ctttgggaag taattaggat tagataaggt 840  
 catgggggtg ggtatgatgg cactgggtgac ttataagaag agaaagagaa atctgagctg 900  
 gcatgctctt gccctctcac tgtgtgatga cttctccatg tcatgatgca gcaagaaggc 960  
 cctcaccaga tgggtggcac atgctttttg acttcccag 999

15

20

25

<210> 28  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

30

<220>  
 <223> Description of Artificial Sequence: artificial primer

35

<400> 28  
 ggctaaatat tttgatgacc aagtt 25

40

<210> 29  
 <211> 19  
 <212> DNA  
 <213> Artificial Sequence

45

<220>  
 <223> Description of Artificial Sequence: artificial primer

<400> 29  
 gcagccaact tcccctgtg 19

50

<210> 30  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

55

<220>

EP 1 780 217 A1

<223> Description of Artificial Sequence: artificial primer  
 <400> 30  
 5 gctctacctt ggtcacctcc 20  
  
 <210> 31  
 <211> 24  
 <212> DNA  
 10 <213> Artificial Sequence  
  
 <220>  
 <223> Description of Artificial Sequence: artificial primer  
 15 <400> 31  
 aggtcacatc catttatccc actg 24  
  
 <210> 32  
 <211> 24  
 <212> DNA  
 20 <213> Artificial Sequence  
  
 <220>  
 <223> Description of Artificial Sequence: artificial primer  
 25 <400> 32  
 agaagatggg ggaatctttt tcct 24  
  
 <210> 33  
 <211> 25  
 <212> DNA  
 30 <213> Artificial Sequence  
  
 <220>  
 <223> Description of Artificial Sequence: artificial primer  
 35 <400> 33  
 ttgtgactgg gctagaaaga aggtg 25  
  
 <210> 34  
 <211> 23  
 <212> DNA  
 40 <213> Artificial Sequence  
  
 <220>  
 <223> Description of Artificial Sequence: artificial primer  
 45 <400> 34  
 50 tgttgctgc atttgtacgt gag 23  
  
 <210> 35  
 <211> 21  
 <212> DNA  
 55 <213> Artificial Sequence

EP 1 780 217 A1

5  
<220>  
<223> Description of Artificial Sequence: artificial primer  
<400> 35  
ttctgtctgg gttggggagg g 21

10  
<210> 36  
<211> 21  
<212> DNA  
<213> Artificial Sequence

15  
<220>  
<223> Description of Artificial Sequence: artificial primer  
<400> 36  
ggaggggtta atatgggtgg c 21

20  
<210> 37  
<211> 22  
<212> DNA  
<213> Artificial Sequence

25  
<220>  
<223> Description of Artificial Sequence: artificial primer  
<400> 37  
tttgtcctgg ttgcctgtgg tc 22

30  
<210> 38  
<211> 23  
<212> DNA  
<213> Artificial Sequence

35  
<220>  
<223> Description of Artificial Sequence: artificial primer  
<400> 38  
caaatcctgt tgactggtct cgg 23

40  
<210> 39  
<211> 22  
<212> DNA  
<213> Artificial Sequence

45  
<220>  
<223> Description of Artificial Sequence: artificial primer  
<400> 39  
aacggctcca tcaccctaa ag 22

50  
<210> 40  
<211> 24

55

EP 1 780 217 A1

<212> DNA  
<213> Artificial Sequence

5 <220>  
<223> Description of Artificial Sequence: artificial primer

<400> 40  
cccactccta gataccaacc caag 24

10 <210> 41  
<211> 24  
<212> DNA  
<213> Artificial Sequence

15 <220>  
<223> Description of Artificial Sequence: artificial primer

20 <400> 41  
ctttatgcac tgcctcgttg aatc 24

<210> 42  
<211> 22  
<212> DNA  
25 <213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

30 <400> 42  
ttgactggtg tggttgctgt tg 22

<210> 43  
<211> 22  
<212> DNA  
35 <213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

40 <400> 43  
gcagaaagg gagttgatgc tg 22

45 <210> 44  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

50 <400> 44  
ctgacaaagt tgagagccca ctg 23

55

EP 1 780 217 A1

5 <210> 45  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

10 <400> 45  
ttaagcctac atccacatgc tgag 24

15 <210> 46  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

20 <400> 46  
ccttggctctg ccagaatttt ca 22

25 <210> 47  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

30 <400> 47  
gtttggcatc ataggagatt tggc 24

35 <210> 48  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

40 <400> 48  
cctgtcccca tgattcagtt acc 23

45 <210> 49  
<211> 19  
<212> DNA  
<213> Artificial Sequence

50 <220>  
<223> Description of Artificial Sequence: artificial primer

55 <400> 49  
acgtacaaat gcaggcaac 19

EP 1 780 217 A1

5  
5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

<210> 50  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

<400> 50  
ccttttttttg tttgtttttg gcggtgc 27

<210> 51  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

<400> 51  
agcttactgg atgaccacca 20

<210> 52  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

<400> 52  
gactggggggg aaaagcgcaa tac 23

<210> 53  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

<400> 53  
gtattgcgct tttcccccca gtc 23

<210> 54  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial primer

EP 1 780 217 A1

<400> 54  
 tgacttgctc tcatcccaca tg 22

5 <210> 55  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence

10 <220>  
 <223> Description of Artificial Sequence: artificial  
 primer

15 <400> 55  
 gggcttgaag caagtaaag gaag 24

20 <210> 56  
 <211> 29  
 <212> DNA  
 <213> Artificial Sequence

25 <220>  
 <223> Description of Artificial Sequence: artificial  
 primer

30 <400> 56  
 gctatcaata ttttcttggt tacagacac 29

35 <210> 57  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

40 <220>  
 <223> Description of Artificial Sequence: artificial  
 primer

45 <400> 57  
 gttcactgcc ataagtcttc agtgc 25

50 <210> 58  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence

55 <220>  
 <223> Description of Artificial Sequence: artificial  
 primer

<400> 58  
 tggccgcact gaagacttat gg 22

<210> 59  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: artificial

EP 1 780 217 A1

primer

5 <400> 59  
cagctgcatc tatgataatc cacc 24

10 <210> 60  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial  
primer

15 <400> 60  
atggacaagt ccgaggtgat ag 22

20 <210> 61  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial  
primer

25 <400> 61  
atcacctcgg acttgtccat tc 22

30 <210> 62  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial  
primer

35 <400> 62  
gcaatcagag atccaaaggc caac 24

40 <210> 63  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial  
primer

45 <400> 63  
gttggccttt ggatctctga ttgc 24

50 <210> 64  
<211> 26  
<212> DNA  
<213> Artificial Sequence

55

<220>  
 <223> Description of Artificial Sequence: artificial primer

5 <400> 64  
 gacatagtat accctggaat tgctgt 26

10 <210> 65  
 <211> 26  
 <212> DNA  
 <213> Artificial Sequence

15 <220>  
 <223> Description of Artificial Sequence: artificial primer

20 <400> 65  
 acagcaattc cagggtatac tatgtc 26

25 <210> 66  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: artificial primer

30 <400> 66  
 ctcccccgat tttagccaag aa 22

35 1

17

40

**Claims**

- 45
1. A nucleic acid molecular structure representing the *Rhesus* genes locus comprising the *RHD*, *SMP1*, and *RHCE* genes and/or the *Rhesus box(es)*, preferably the hybrid *Rhesus box*, the upstream *Rhesus box* and/or the downstream *Rhesus box*, the sequence of which is shown in Figures 8 to 10.
  - 50 2. The nucleic acid molecular structure of claim 1 representative of the common *RHD* positive haplotypes.
  3. The nucleic acid molecular structure of claim 1 derived from a sample comprising an *RHD* positive haplotype that is serologically classified RhD negative.
  - 55 4. The nucleic acid molecular structure of claim 3 wherein said sample is selected from a Caucasian population.
  5. The nucleic acid molecular structure of any one of claims 2 to 4 or a nucleic acid molecule being derived from the *RHD* gene comprising a single nucleotide substitution within the coding region of the *RHD* gene or within a 5' or 3'

splice site.

- 5
6. The nucleic acid molecular structure or a nucleic acid molecule of claim 5 wherein said substitution gives rise to a mutation within a 4-nucleotide sequence, a 6-nucleotide sequence or an 8-nucleotide sequence comprising the consensus splice site at the exon9/intron9 boundary.
7. The nucleic acid molecular structure or a nucleic acid molecule of claim 6 wherein said substitution gives rise to an *RHD*(K409K) mutation.
- 10
8. The nucleic acid molecular structure or a nucleic acid molecule of claim 7 wherein said substitution is a substitution at the 5' splice site intron 9 from AGgt to AAgt.
9. The nucleic acid molecular structure or a nucleic acid molecule of claim 5 wherein said substitution gives rise to a mutation within a 4-nucleotide sequence, a 6-nucleotide sequence or an 8-nucleotide sequence comprising the consensus splice site at the exon3/intron 3 boundary.
- 15
10. The nucleic acid molecular structure or a nucleic acid molecule of claim 9 wherein said substitution gives rise to an *RHD*(G486(+1)A) mutation.
- 20
11. The nucleic acid molecular structure or a nucleic acid molecule of claim 10 wherein said substitution is a substitution at the 5' splice site intron 3 from ACgt to ACat.
12. The nucleic acid molecular structure or a nucleic acid molecule of claims 6 to 11 correlating with a D<sub>e1</sub>-phenotype.
- 25
13. The nucleic acid molecular structure or a nucleic acid molecule of claim 5 wherein said nucleotide substitution gives rise to a stop-codon at codon 16.
14. The nucleic acid molecular structure or a nucleic acid molecule of claim 13 wherein said substitution gene gives rise to an *RHD*(W16X) mutation.
- 30
15. The nucleic acid molecular structure or a nucleic acid molecule of claim 14 wherein said substitution is a G → A substitution at nucleotide position 48.
16. The nucleic acid molecular structure or a nucleic acid molecule of claim 5 wherein said nucleotide substitution gives rise to a stop codon at codon 330.
- 35
17. The nucleic acid molecular structure or a nucleic acid molecule of claim 16 wherein said substitution gives rise to an *RHD*(Y330X) mutation.
- 40
18. The nucleic acid molecular structure or a nucleic acid molecule of claim 17 wherein said substitution is a C → G substitution at nucleotide position 985.
19. The nucleic acid molecular structure or a nucleic acid molecule of claim 5 wherein said substitution gives rise to a missense mutation at codon 212.
- 45
20. The nucleic acid molecular structure or a nucleic acid molecule of claim 19 wherein said substitution gives rise to an *RHD*(G212V) missense mutation.
21. The nucleic acid molecular structure or a nucleic acid molecule of claim 20 wherein said substitution is a G → T substitution at position 635.
- 50
22. The nucleic acid molecular structure or a nucleic acid molecule of claim 5 wherein said substitution gives rise to a mutation within a 4-nucleotide sequence, a 6-nucleotide sequence or an 8-nucleotide sequence comprising the consensus splice site at the exon8/intron8 boundary.
- 55
23. The nucleic acid molecular structure or a nucleic acid molecule of claim 22 wherein said substitution gives rise to an *RHD*(G1153(+1)A) mutation.

24. The nucleic acid molecular structure or a nucleic acid molecule of claim 23 wherein said substitution is a substitution at the 5' splice site intron 8 from AGgt to AGat.
- 5 25. The nucleic acid molecular structure or a nucleic acid molecule of any one of claims 13 to 24 correlating with an RhD negative phenotype.
26. A vector comprising the nucleic acid molecular structure or a nucleic acid molecule of any one of claims 1 to 25.
- 10 27. A non-human host transformed with the vector of claim 26.
28. A method producing a protein product of the *RHD* gene comprising culturing the host of claim 27 under suitable conditions and isolating the Rhesus protein produced.
- 15 29. A protein product of the *RHD* gene encoded by the nucleic acid molecule or structure of any one of claims 1 to 25 or produced by the method of claim 28.
30. An oligonucleotide hybridizing under stringent conditions to a portion of the nucleic acid molecular structure or the nucleic acid molecule on any one of claim 1 to 25 wherein said portion comprises said (missense) mutation or said stop codon or to the complementary portion thereof.
- 20 31. An antibody or aptamer or phage specifically binding to a protein product of the *RHD* gene of claim 29.
32. A method to simultaneously detect the presence of *RHD*<sub>ψ</sub> and any of the *RHD* molecular structures of any one of claims 1 to 25 comprising hybridizing the oligonucleotide of claim 30 and at least an other oligonucleotide hybridizing to an *RHD*<sub>ψ</sub> structure under stringent conditions to nucleic acid molecules comprised in the sample obtained from a human and detecting said hybridization.
- 25 33. A method for testing for the presence of a nucleic acid molecule encoding a mutant Rhesus D antigen as **characterized by** the nucleic acid molecular structure or a nucleic acid molecule of any one of claims 1 to 25 in a sample comprising hybridizing the oligonucleotide of claim 30 under stringent conditions to nucleic acid molecules comprised in the sample obtained from a human and detecting said hybridization.
- 30 34. The method of claim 33 further comprising digesting the product of said hybridization with a restriction endonuclease and analyzing the product of said digestion.
- 35 35. A method for testing simultaneously for the presence of *RHD*<sub>ψ</sub> and any of the *RHD* molecular structures of any one of claims 1 to 25 in a sample comprising determining the nucleic acid sequence of at least a portion of the nucleic acid molecular structure or nucleic acid molecule of any one of claims 1 to 25, said portion encoding said (missense) mutation or said stop codon and determining of at least a portion of an *RHD*<sub>ψ</sub> structure.
- 40 36. A method for testing for the presence of a nucleic acid molecule encoding a mutant Rhesus D antigen as **characterized by** the nucleic acid molecular structure or a nucleic acid molecule of any one of claims 1 to 25 in a sample comprising determining the nucleic acid sequence of at least a portion of the nucleic acid molecular structure or nucleic acid molecule of any one of claims 1 to 25, said portion encoding said (missense) mutation or said stop codon.
- 45 37. The method of claim 36 further comprising, prior to determining said nucleic acid sequence, amplification of at least said portion of said nucleic acid molecule or structure.
- 50 38. A method for testing simultaneously for the presence of *RHD*<sub>ψ</sub> and any of the *RHD* molecular structures of any one of claims 1 to 25 in a sample comprising carrying out an amplification reaction using a set of primers that amplifies at least a portion of said sequence wherein at least one of the primers employed in said amplification reaction is the oligonucleotide of claim 30 and at least a primer amplifying an *RHD*<sub>ψ</sub> structure and analyzing the amplified product(s).
- 55 39. A method for testing for the presence of a nucleic acid molecule encoding a mutant Rhesus D antigen as **characterized by** the nucleic acid molecular structure or nucleic acid molecule of any one of claims 1 to 25 in a sample comprising carrying out an amplification reaction using a set of primers that amplifies at least a portion of said sequence wherein at least one of the primers employed in said amplification reaction is the oligonucleotide of claim 30.

40. The method of claim 39 wherein at least one of the primers employed in said amplification reaction is the oligonucleotide of claim 30.
- 5 41. The method of any one claims 37 to 40 wherein said amplification is effected by or said amplification reaction is the polymerase chain reaction (PCR).
42. The method of claim 41 wherein said PCR is PCR-RFLP, PCR-SSP or long-range PCR.
- 10 43. The method of any one of claims 37 to 42 wherein the molecular weight of the amplification product is analyzed.
44. A method for testing for the presence of the nucleic acid molecular structure or nucleic acid molecule of any one of claims 3 to 25 encoding *RHD* positive alleles comprising the following steps:
- 15 (a) isolating the DNA from a blood sample or blood donor;  
(b) hybridizing at least two oppositely oriented primers under stringent conditions to the DNA so as to carry out a PCR;  
(c) amplifying the target sequence;  
(d) separating the amplification products on a gel; and  
20 (e) analyzing the amplicons.
45. The method of claim 44 wherein said *RHD* positive alleles are derived from a serologically RhD negative sample.
46. The method of claim 44 or 45 wherein said sample is selected from a Caucasian population.
- 25 47. A method for testing for the presence of a protein product of the *RHD* gene of claim 29 in a sample comprising assaying a sample obtained from a human for specific binding to the antibody or aptamer or phage of claim 31.
- 30 48. A method for testing for the presence of a protein product of the *RHD* gene encoding the nucleic acid molecular structure or nucleic acid molecule of any one of claims 3 to 25 in a sample comprising utilizing direct agglutination methods, indirect antiglobulin tests, monoclonal anti-D antibodies and/or adsorption/elution techniques.
- 35 49. The method of any one of claims 28 to 48 wherein said sample is blood, serum, plasma, fetal tissue, saliva, urine, mucosal tissue, mucus, vaginal tissue, fetal tissue obtained from the vagina, skin, hair, hair follicle or another human tissue.
50. The method of claim 49 comprising enrichment of fetal cells or extraction of fetal DNA or mRNA from maternal tissue, like peripheral blood, serum or plasma.
- 40 51. The method of any one of claims 28 to 50 wherein said nucleic acid molecule or proteinaceous material from said sample is fixed to a solid support.
52. The method of claim 51 wherein said solid support is a chip.
- 45 53. Use of the nucleic acid molecular structure or the nucleic acid molecule of any one of claims 1 to 25 for the analysis of a negative or a positive Rhesus D phenotype.
- 50 54. Use of the nucleic acid molecular structure or the nucleic acid molecule of any one of claims 1 to 25, the vector of claim 26 or the protein product of the *RHD* gene of claim 29 for the assessment of the affinity, avidity and/or reactivity of monoclonal antibodies or of polyclonal antisera preferably anti-D antisera, anti-C antisera, anti-globulin or anti-human-globulin antisera.
- 55 55. Use of cells, preferably red blood cells from probands carrying the nucleic acid molecular structure or the nucleic acid molecule of any one of claims 1 to 25 for the assessment of the affinity, avidity and/or reactivity of monoclonal anti-D or anti-C antibodies or of polyclonal anti-D or anti-C sera or of anti-globulin or of anti-human-globulin antisera or of preparations thereof.
56. A method for the characterization of the monoclonal antibodies or polyclonal antisera or of a preparation thereof said method comprising

- (a) testing the nucleic acid molecular structure of a sample of a proband for the presence of a mutation as defined in any one of claims 1 to 25;
- (b) correlating, on the basis of the mutation and the allelic status of the *RHD* gene, the nucleic acid with the density of the protein product of the *RHD* gene on the surface of red blood cells of said proband;
- (c) reacting said monoclonal antibodies or polyclonal antisera or said preparation thereof with a cell carrying the protein product of the *RHD* gene on its surface;
- (d) characterizing said monoclonal antibodies or polyclonal antisera or said preparation thereof on the basis of the results obtained in step (c).

57. The method of claim 56 wherein said characterization comprises the determination of reactivity, sensitivity, avidity, affinity specificity and/or other characteristics of antibodies and antisera.

58. The method of claim 56 or 57 wherein said cell carrying the protein product of the *RHD* gene on its surface is a red blood cell.

59. A method for determining whether a patient in need of a blood transfusion is to be transfused with RhD negative blood from a donor comprising the step of testing a sample from said patient for the presence of one or more nucleic acid molecular structures or nucleic acid molecules of any one of claims 3 to 25, wherein a positive testing for two different of said nucleic acid molecular structures is indicative of the need for a transfusion with Rh negative blood.

60. A method of assessing of the risk of a RhD negative mother of conceiving or carrying an RhD positive fetus or of the risk of a mother having an anti-D titer of conceiving or carrying a fetus at risk to develop haemolytic disease of the newborn comprising assessing a sample obtained from the father of the fetus for the presence of one or more nucleic acid molecular structures or nucleic acid molecules as defined in any one of claims 1 to 25.

61. The method of claim 60 said nucleic acid molecular structure carries mutations.

62. A method for assessing the possibility or likelihood of a man being the father of a child by assaying a sample obtained from said man for the presence of one or more nucleic acid molecular structure or nucleic acid molecules of any one of claims 1 to 25, wherein the test results are used to determine the homozygosity for, the heterozygosity for or the absence of any nucleic acid molecular structure or a nucleic acid molecule of claims 3 to 25 used to infer the possibility or likelihood of said man being the father of the child.

63. A method for treating a pregnant woman being Rhesus D negative, wherein the fetus does not carry two nucleic acid molecular structures or nucleic acid molecules of any one of claims 3 to 25 or is not homozygous for any nucleic acid molecular structure of any one of claims 3 to 25 comprising administering anti-D to said woman.

64. Use of an aptamer, phage, monoclonal antibody or a polyclonal antiserum as **characterized in** claim 31 or a preparation thereof for determination of the protein product of the *RHD* gene.

65. Use of claim 64 wherein said determination of the protein product of the *RHD* gene is effected in connection with blood group typing.

66. A preparation comprising the antibody or aptamer or phage of claim 31.

67. A method of identifying an antibody  $V_H$  or  $V_L$  chain or a combination thereof or an aptamer specifically binding to a protein product of the *RHD* gene of claim 29 comprising

- (a) contacting the protein product of the *RHD* gene of claim 29 with a phage library displaying  $V_H$  or  $V_L$  chains or combinations thereof on the surface of the phage or with aptamers;
- (b) identifying phage or aptamers that bind to said protein product; and optionally
- (c) repeating steps (a) and (b) one or more times.

68. A method of identifying a monoclonal antibody specifically binding to a protein product of the *RHD* gene of claim 29 comprising

- (a) contacting the protein product of claim 29 with one or more monoclonal antibodies;
- (b) identifying monoclonal antibodies that bind to said protein; and optionally

(c) repeating steps (a) and (b) one or more times.

5 69. A method of identifying an antibody V<sub>H</sub> or V<sub>L</sub> chain or a combination thereof or an aptamer specifically binding to a protein product of the *RHD* gene of claim 29 comprising

(aa) contacting said protein product; and

(ab) a normal D polypeptide

wherein the normal D polypeptide is present in a molar mass that is higher, equal or less than the protein product of the *RHD* gene of (aa) with a phage library displaying V<sub>H</sub> or V<sub>L</sub> chains or combinations thereof on the surface of the phage or with aptamers;

10 (b) identifying phage or aptamers that bind to said protein product of the *RHD* gene of (aa); and optionally

(c) repeating steps (a) and (b) one or more times.

15 70. A method of identifying a monoclonal antibody specifically binding to a protein product of the *RHD* gene of claim 29 comprising

(aa) contacting the protein product of the *RHD* gene; and

(ab) a normal D polypeptide

wherein the normal D polypeptide is present in a molar mass which is higher, equal or less than the protein product of the *RHD* gene of (a) with one or more monoclonal antibodies;

20 (b) identifying monoclonal antibodies that bind to said protein product of the *RHD* gene of (a); and optionally

(c) repeating steps (a) and (b) one or more times.

25 71. The method of any one of claims 67 to 70, wherein the protein product of the *RHD* gene is exposed on the surface of a cell.

72. The method of any one of claims 67 to 71, wherein the polypeptide or host cell is affixed to a solid support.

30 73. The method of any one of claims 67 to 72, wherein subsequent to step (b) or (c), the following step is carried out:

(d) identifying the amino acid sequence of the V<sub>H</sub> or V<sub>L</sub> chains and/or identifying the nucleic acid sequence encoding said amino acid sequence.

35 74. The method of any one of claims 67 to 70, wherein, in the case that only one round of selection is employed for the identification, the number of protein molecules of the *RHD* gene of (a) is in molar excess over the number of phage particles.

40 75. Use of cells, preferably red blood cells comprising the protein product of the *RHD* gene of claim 29 or produced by the method of claim 28, from probands for the assessment of the affinity, avidity and/or reactivity of monoclonal anti-D or anti-C antibodies of claim 31 or of polyclonal anti-D or anti-C antisera or of anti-globulin or of anti-human-globulin antisera or of preparations thereof.

76. Kit comprising

45 (a) the oligonucleotide of claim 30, and/or

(b) the antibody of claim 31; and/or

(c) the aptamer of claim 31; and/or

(d) the phage of claim 31; and/or

50 (e) a pair of primers useful for carrying out the amplification reaction of any one of claims 36 and 43.

Fig. 1

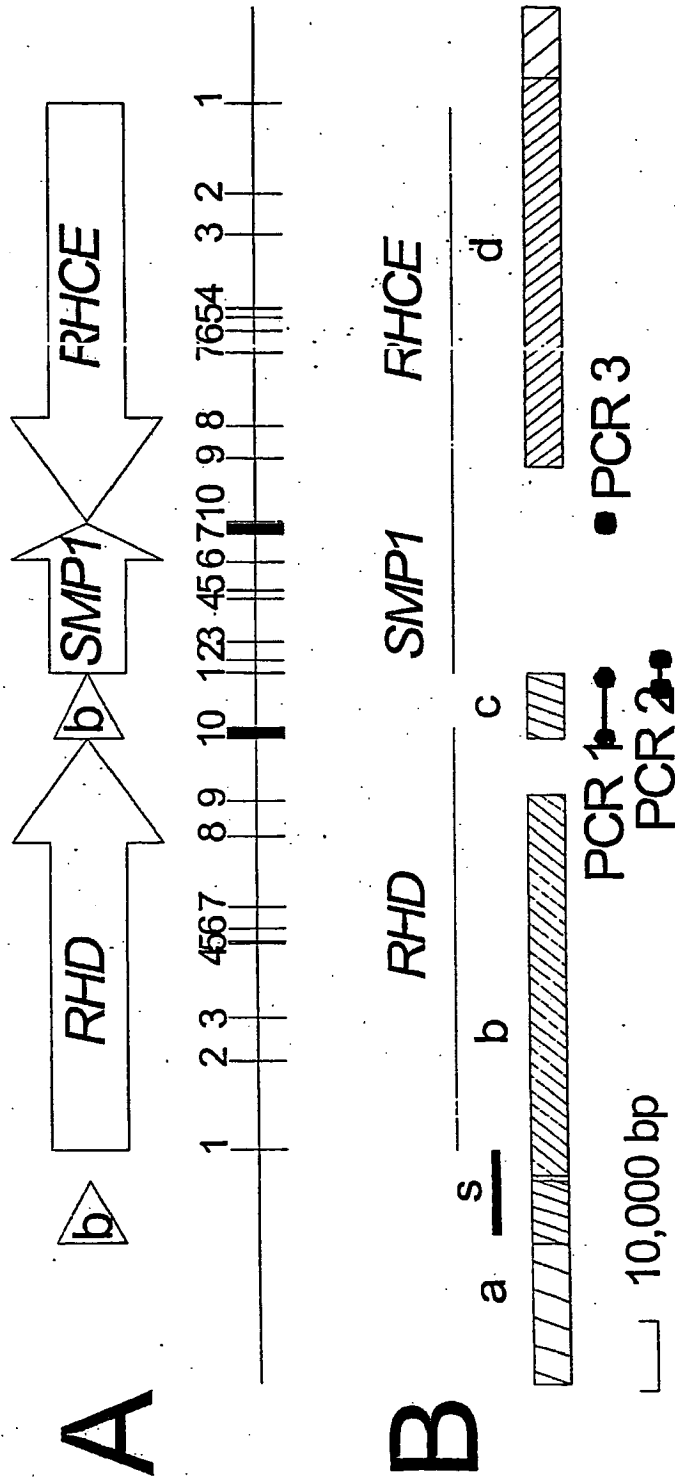


Fig. 2

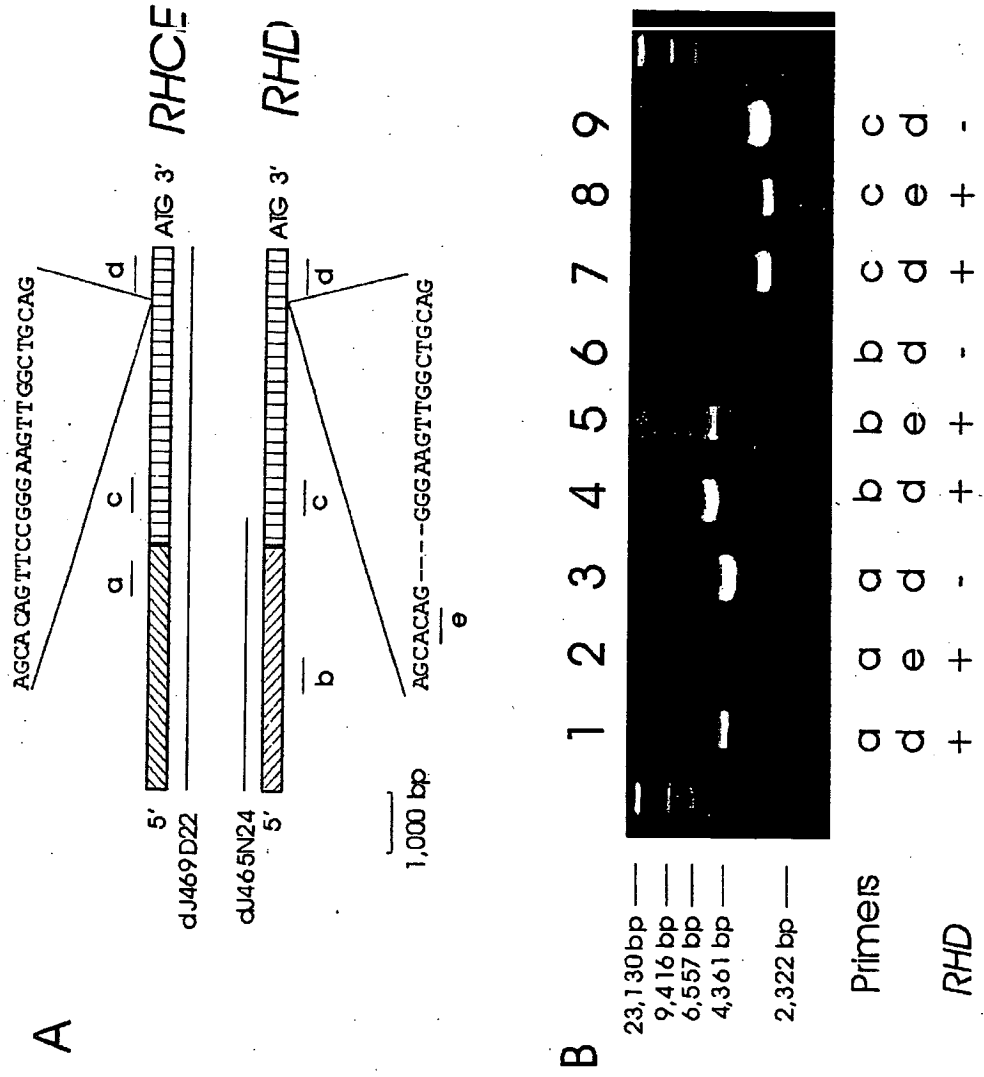


Fig. 3

```

-- Rhesus box ----->
gcatgcgactgagccgggtgatggactgctgcatccgggtgtctg

      <- SMP1 Exon 1 (91 bp, 5'UTR)-----
gaggctgtggccggttttcttcttgctaataacggggagtgagggcggg

-- SMP1 Exon 1 -----><---- Intron ~ 2,000 bp
...actgcacgacggggctggactgacgt.....

Intron ---><---- SMP1 Exon 2 (106 bp, 93 translated)-----
.....,agctgaaaaaaATGCTGGATTCTAGAGGGCTTGAGATGCTCAG...

-- SMP1 Exon 2 -----><---- Intron ~ 2,500 bp
...GCTTCCATTGCTGGTGGTACTAgT...

Intron ---><---- SMP1 Exon 3 (113 bp, all translated)-----
.....agTTTTTTACAGGCTGGTGGATTATCATAGATGCAGCTGTATT...

-- SMP1 Exon 3 -----><---- Intron ~ 8,000 bp
...TATAGCAACCATAGCCTTCCTAAAtgt.....

Intron ---><---- SMP1 Exon 4 (68 bp, all translated)-----
.....agGATTAATGCAGTATCGAATGGACAAGTCCGAGGTGATAGTTA...

-- SMP1 Exon 4 -----><---- Intron ~ 1,400 bp
...TGAAGGTTGCTGGGTCAAACAGgt.....

```

Fig. 3 cont.

```

Intron ----><--- SMP1 Exon 5 (93 bp, all translated)-----
.....agGTGCTCGCATTGGCTTTTCGTTGGTTTCATGTTGCCCTTG...

--- SMP1 Exon 5 -----><--- Intron ~ 4,000 bp
...TTTTGGAGGTTATGTTGCTAAAGgt.....

Intron ----><--- SMP1 Exon 6 (61 bp, all translated)-----
.....agAAAAAGACATAGTATACCCCTGGAAATTGCTGTATTT.....

--- SMP1 Exon 6 -----><--- Intron ~ 4,000 bp
...TCCAGAAATGCCCTTCATCTTTTTGgt

Intron ----><--- SMP1 Exon 7 (1,703 bp, 47 translated)---
.....agGAGGGCTGGTTTTTAAGTTGGCCGCACCTGAAGACTTATGGCAGT

<----- RHCE Exon 10 -----
--- SMP1 Exon 7 -----
GAaac...agcatcatcctaataatgaaactaaacatttattttaaac

----- RHCE Exon 10 -----
----- SMP1 Exon 7 ---->
ttattaaattgactctttaaactaaagttttagtctttaatttttaataatcaa

--- RHCE Exon 10 ----- Homology with RHD -
atctgtctgtaccttgtttcattatacataaggagccttctgtcatgagcgtttc

```

Fig. 3 cont.

```

----- Homology with RHD -----
----- RHCE Exon 10 -----
tcacgtacaatatgcaggcaacacagtgagaggaagttgtctgtt

-- Homology with RHD -----
-- RHCE Exon 10 -----
tttgaacaggccttgtttttcttgatgcttttgcTTAAATCCAACACAGCCCAATGAGG

```

Fig. 4

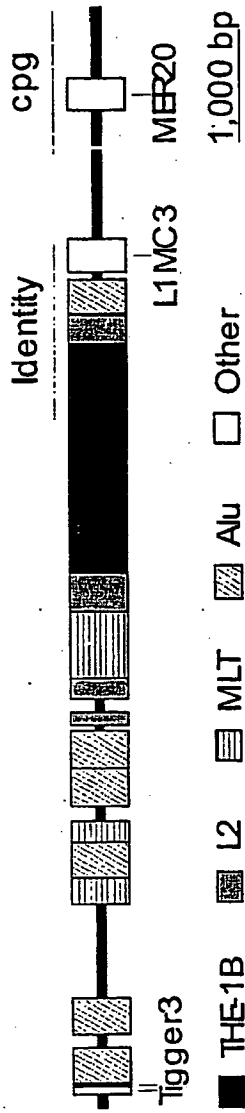


Fig. 5

```

attcactatcacaagaacagcagggtaa
*****
**_*****g*****
-----rez7----->
gacctgtcccatgattcagttacct
*****
*****
ccactgggtccctcccacaacgcatgggaattcaggatgagatt
*****
*****
gggtgggacacaaccaacctatcatccaccatggcccctccaaa
*****
*****
tttcatgtcctcacatttcaaaaccaatcacaccatcccacagtcctc
*****
*****
aaagtcttaaatgatttcagcattaactcaaaagtcacacagtctaagtc
*****
*****

```

Fig. 5 cont.

tcattctgagacaaggcaagtccttccatttatgagcctataaaatccaa  
\*\*\*\*\*  
\*\*\*\*\*  
\*\*\*\*\*c  
\*\*\*\*\*

agcaagttagttacttccctagatatacaatgggtacagggcattgggtaaa  
\*\*\*\*\*  
\*\*\*\*\*a  
\*\*\*\*\*

tacagccattccaaatggataaaatgggtcaaaacaagaggctacagggc  
\*\*\*\*\*  
\*\*\*\*\*

ccatgagagtcctccaaatccagtgaggcagtcctcaaatctaaagctccaaaat  
\*\*\*\*\*  
\*\*\*\*\*

gatctcct-ttgactccacatctcacatccaggtcacgcagatggaagg  
\*\*\*\*\*  
\*\*\*\*\*c  
\*\*\*\*\*

ggtgggttcccatggtcttgggcagctctgcccctgtaccttgcagggt  
\*\*\*\*\*  
\*\*\*\*\*

Fig. 5 cont.

```

acagcctccctctcagctgcttcatgggctggcatgagtgctgcagc
*****
*****a
*****
      ^^^^
t t t t c c a g g t a c a c g g t g c a a g c t g t c g g t g g a t c c a c c a t t c t g g g g t c
*****
*****
*****
t g g a g c c t c t t c a c a g c t c c a c t a g g t g g t g c c c a g t a g g a c t g
*****
*****
*****
t g t g t g g g t c t c t g a c c c a c a t t c c c t t c t g c a c t g c c c t g g c a g a g
*****
*****
*****
g a t c t c c a t g a g g c c c t g c t c c t g c a g c a a a c t t c t g a c t g g g c a t c c a
*****
*****c
*****
      ^^^^
g g c a t t c c g c a c a t c c t c t t a a t c t a g g g a a g g t t c c a a c c c c a a
*****
*****
*****g

```

Fig. 5 cont.

```

ttcttgacttctgtgcaactcgcagctcaacaccacatggaagctgtcaa
*****
*****
*****
ggcttggggcttgcaactccccgaagctacagcccagctctaccttgcct
*****
*****
*****
cccgtcagtcatggttgggagtggtggatgcagggccaccaagtcacctta
*****
*****
*****
ggctgcacacagcatgaggaccccgccctggccaaacaaaccattttt
*****
*****
*****
+-- breakpoint region ->
+-- identity region ->
cctgatacctctggaccctgtgatggggggttgcataaagacctctga
*****
*****
*****
catgccctggagacatttcccattgtcttgggaattagcatttggtc
*****
*****
*****

```

Fig. 5 cont.

```

ctgttactcatgcaaaatttctgcagccagcttgaatttctcctcagaaaa
*****
*****
*****
^^^^^^
tgggaatttttctttctatcacattgtcaggctgcaaatttccgaactt
*****
*****
*****
ttatgctctgcttcccttataaaactgaatgtctttaacagcaccagg
*****
*****
*****
tcaccttgaatgctttgctgcttagaaatttctcctgcccagatactct
*****
*****
*****
aatcatctcttgaagtcaaagttctacaaatatctcgtgccccggcca
*****
*****
*****
aatgcccaggatctttgctaaacataacaagagtcccccttgcctcc
*****
*****
*****

```

Fig. 5 cont.

```

agttccaacaagttcctcatctccgtctgagaccacctcagcctatgga
*****
*****
*****
ctttatgtccacagtgctatcagcatttgggcaagccattcaacaag
*****
*****
*****
tctctaggaagttccaacttccacatttgcctgttcttctgagcc
*****
*****
*****
ctccaaactgttccaaaccctgcctgttaccagttccaagtcacatac
*****
*****
*****
ccatttttgagtatctacggcagcaccctctactggtaccaatttag
*****
*****
*****
ccactgaagtagttggagaacagaagtaatagactctggtttacattgta
*****
*****
*****
aaagcttctctgtggctgtgtgaagaaataatagagaatgaagccc
*****
*****
*****

```

Fig. 5 cont.

caagatgaagcaggacacagttgcagtggttagagtaagaaatgctgct  
 \*\*\*\*\*  
 \*\*\*\*\*

ggctggcactgaagtgatagcctggaggttgtgtgacatgcatgtg  
 \*\*\*\*\*  
 \*\*\*\*\*

tatgtgtttacgatagtaggccaacagat  
 \*\*\*\*\*  
 \*\*\*\*\*

<-- breakpoint region ---+  
 actgtaatccacactgtttttttttt---ga  
 \*\*\*\*\*  
 \*\*\*\*\*

gacagagtctcacctgtgacctagactagaatgc  
 \*\*\*\*\*  
 \*\*\*\*\*

agtgacaatcttggctcactacaacctccacctccaggttcaacaatc  
 \*\*\*\*\*  
 \*\*\*\*\*

Fig. 5 cont.

```
ctgtgcttcagcctcccgagtagtggtggttacagggtgtgtgccacc
*****
*****
gtgccagctataattttgtatttttagcagagatgggattttgccacat
*****
*****
tggcaggctggtcttgaactcctggcctcaagcaatcctcccacctta
*****
*****
gcctcccaaagtgtgagccaccacacctggccgcaactgattttaatc
*****
*****
atgaaatgacacatacatttaaaaaaccaatacctataatattcctggc
*****
*****
tagtactctcacatctatatcatcaaaaaaagaagatgtgaaact
*****
*****
```

Fig. 5 cont.

gacacagccaaggggagactaaggagacataacaattaactgtaatgtgg  
 \*\*\*\*\*  
 \*\*\*\*\*

tattctggaggggatcctggaacagaaaaagacattaggcaaaaaactaa  
 \*\*\*\*\*  
 \*\*\*\*\*

agaaatctgaataaaaatgtggatgtcagttaataataatgtatcatatta  
 \*\*\*\*\*  
 \*\*\*\*\*

<-- identity region--

gtccagtaattgtaacaaatataccacaataatgaagccattaatata  
 \*\*\*\*\*  
 \*\*\*\*\*

gggaaaatggagggttaatatgggtggcttttgctatttctagcag  
 \*\*\*\*\*  
 \*\*\*\*\*

ctccatttatctgaaaaagacaaacattcattagcccaaaaaggtaaa  
 \*\*\*\*\*  
 \*\*\*\*\*





Fig. 6

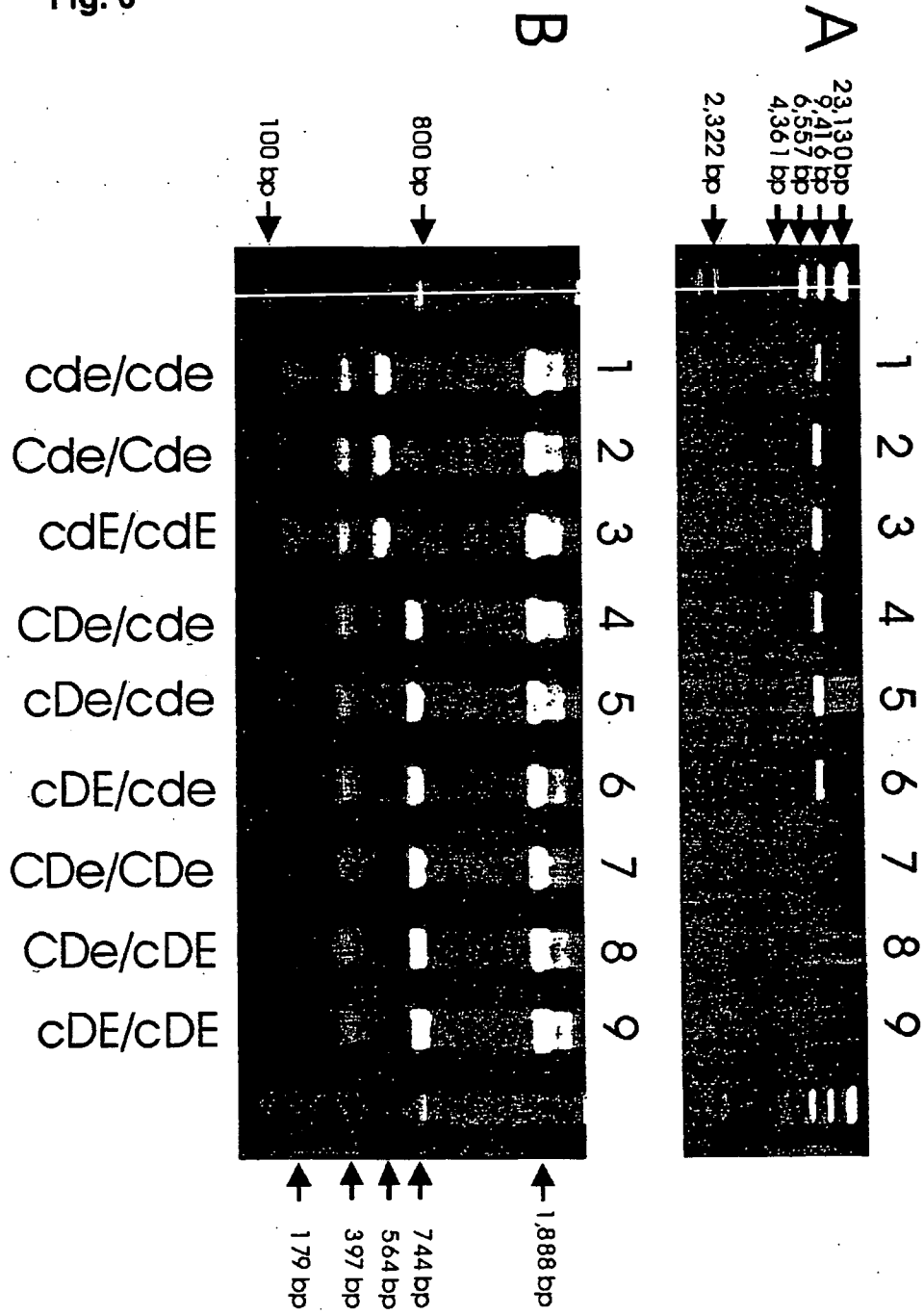


Fig. 7

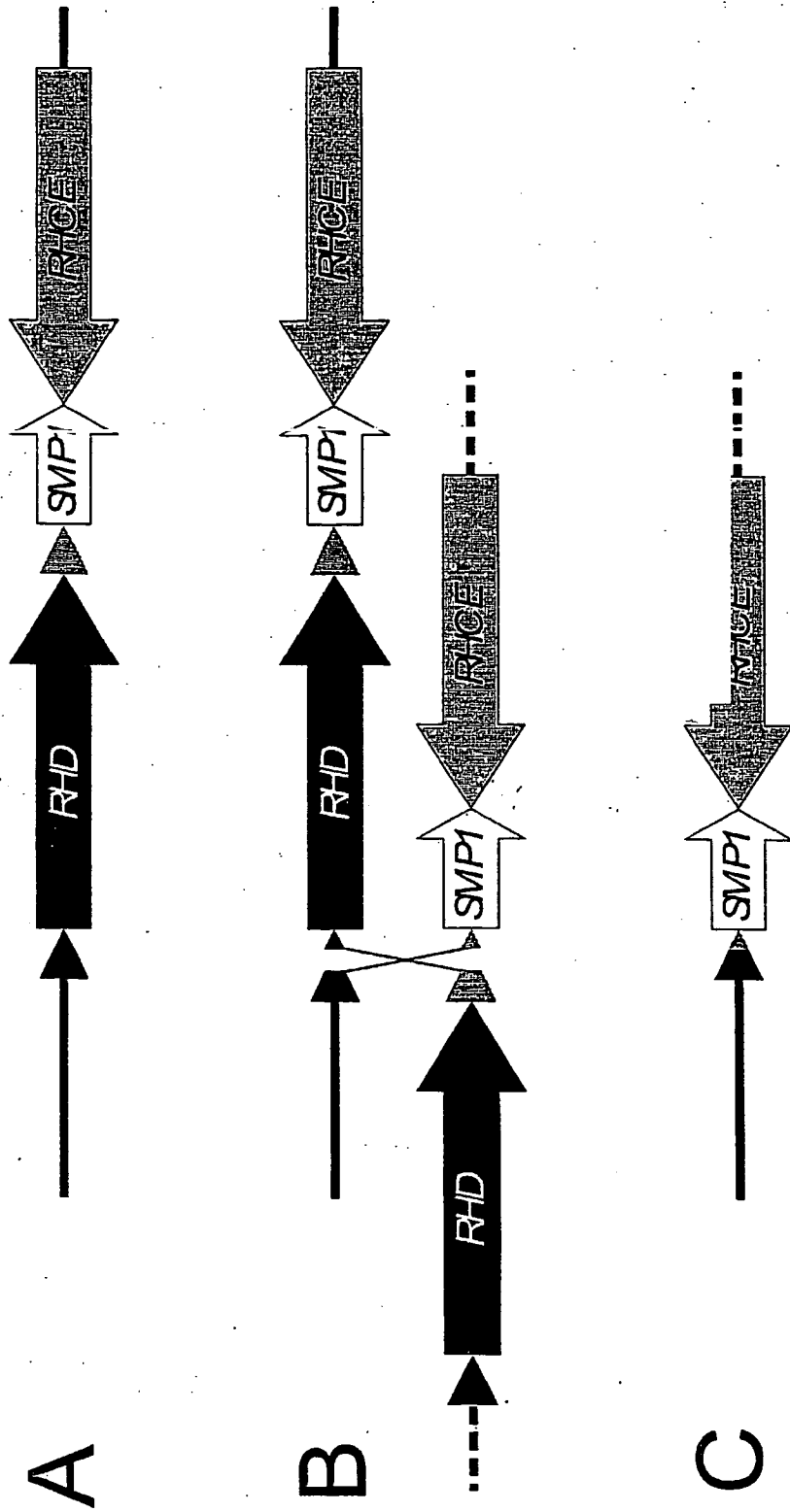


Fig. 8

Hybrid Rhesus box of RHD negatives

5' ctagaaaacactttgtcatttttagaggtgta  
 (start of Rhesus box)  
 tccaatgttcgcgaggcactggagtcagagaaaatggagttgaatcctttctctgccactc  
 ttgaggagaatctcaccatttattatgcactgtagaatacaacaataaaatacagccatgt  
 accacataacaacatcttggtaaacaacagactgcatatatgatgggggcatccagtaagc  
 taaggttaatttattattattccctttttttcttttttttgagatgtagtcttactctg  
 tccccaggctagagtgcaatggcaccatcttggtcactgcaacctctgcctcctgggtc  
 aagogaatctcctgcctcagcctccgaagtagctgggaattacaggaacccaccacatctgg  
 ctaattttttgtatttttagtaaagatggggttcaccatgttgccaggctgatctcaaac  
 tctgacctcaagtgatctgcctgcctcggcctcccaaagtgctgggacataggcctgagc  
 cactgtgcccgcccttgtttgcttttttaacagttaacagtgctcatagaaactgctttg  
 acatgactgcaatcatgtgcttcatagaaacttaattagattataccactagagcttccaga  
 tttttatacttttttttgaaacggagtctcactctgtcaccaggctggagtgcaagtgccg  
 caatctcagctcgccgcaacctccgcctcccaggttcaagtattctcctgcctcagcctcc  
 cgagtagctgggattacaagtgcacactaccagcccagctaatttttgcaattttactaga  
 cagggtttcaccatgttggttaggatagtttcaccaggatctcttgccctcatgatcagcct  
 gcctcggcctcccaaagtgctgggattacaggtgtgagccaccgtgccagcctatacttcc  
 ctttttgaataccatttggcgttttgaagaattaacagctttgtgaacgtggcagtgcttgt  
 gattcaggcttccactgagaccaaggggagaacctgggttgaggacaaacagacggacagcg  
 tgtggcagtggttaaatgctcttctgaaggctgatacagacagctctctgtgcaactgattgta  
 tacgcatccaagattatattattgttttctattgctatgtgtcacactttgccaaacagga  
 tgtggaaaatgaataagcggttttcttaggcacttcttaacagacaattgggtcaaatgaac  
 tccattgcttaagaaacacataaacaccatttagtcaactgaatagctatatgtatgggtg  
 ctactatggggaatcttgtttgccaattttctttgaaaattctggcagaccaaggttcttt  
 ttgtttacacaatacttgaaaaataaaaatgaacaagccaacaaactaccaagttttcactt  
 acataaatgtagttacatacagaaaatgtgactgtgaattttttctaggacttttaactat  
 aagcactatttgcacgaaagagaaccaatctatcaattacaaactcacataattttacagat  
 tttttttccctacacagcacataaaaacagaaggaatttgaagccaccctccaaacacaggg  
 gaaggaggctgtgtgtatcctcattgtctttcacattctaagggtgggtccactcagtgac  
 tgaaatccttaagtgttgattagtcggcttgggctaccataacagcagcttaactgttgt  
 taagccactcagacttaaacacagaaaatttatttccctatagttctggaggctggaagttc  
 aaggtgccggcaaggctggtttctgggtgagacctctctccctgtcttgagatggctgcctc  
 ctccctgtgtcctcatagagcctgtcctctgcttttacacttctgggtgtcatcttccctttt  
 ttttttttttgagacagagctctcgctctatcgcccaggctggagtgcaagtggcccgatcga  
 tctcggctcactgcaacctctgcctcccaggttcaagcaattctcctgcctcagcctcccga  
 gtagctgggactacaggtgcccgccatcatgtctggctaatttttgtatttttagtagagac  
 agggtttcaccatattggccaggctggctcceaactcctgaccttgtcatctgcctgcctcg  
 gcctcccaaagtgctaggattacaggcgtgagccaccgcaccggcctcttctcttcttat  
 aaggacaccagtcctattagattagggctccaccctcatgacctcatttgaccttaactatt  
 atttctttaagcacctatttccaaatatagtcactttaggggttagggcttcaaaatatga  
 atctgaggggagatcaattcagtaaatagcagtagtcattaacggacaatatatacaagata  
 atttcgtgattactgtccttatgcataaatgtcctcagtgttccactgcctttatccagatt  
 tactatcaciaaagactttgctctgagaaaaatgtgatttctttctttttttttttttttg  
 gacagagctcactctgtcaccaggtggagtgcaagtggtgcaatctcggctcactgcaat  
 ctccgcctcccaggttcacgccattctcttgcctcagctctcccagtagctgggcctacagg

**Fig. 8 cont.**

## Hybrid Rhesus box of RHD negatives

cgcccgccaccctgccagctaattttttagtatttttagtagagacggggtttcaccatggt  
 agccaggatgggtctcaatctcctgacctcgtgatccacctgcctcagcctcccaaagtctg  
 ggattacaggcatgagccaccgcccagcagatttttttttttttttttttttttgagat  
 ggagtcttgctgtgttgcccagcctggagtgcagtggtatgattttggctcactgcaacctc  
 tgtctaccatggtcaagcgattctcccacctctgcctcccgtgtagctgggatcacaggcac  
 acgccaccacacctagctactttttgatttttagtagaaatggggtttcaccatggtggcc  
 aggatgggtcccgaactcctgacctcaagtgatcctcctgcctcggcctcccaaagtctggg  
 attacagggtgtgagccactgtgcctggccaaaatgtgattttcttatttcccacattgcca  
 ttccatttcaattaactataatagctatgtctattgagcactcaagtgtattctagaaactg  
 ttcttgattctggggatataatccatgaatcaactatagtcctgttattaagtaactctgtag  
 tctgactaaaccattagaaatttaaaaatggctactttcaaagacatcttggagttcagga  
 gtcccacactgccaacatattacctaataatccaacctgcttgaattcacttatttaacc  
 aatatttattgagtgccaactttgagcctaagatacagcagtaaaacaaatggataaagtccc  
 tgtcctcatgaaacttgtatttctaataatggaagaaacagaaaacaaacagatataggatgta  
 atcaggtagggataaataactttgaattcaaacaaaagtatacgtagtcagggttcgccaag  
 agacacagccaatcggatacatagatatataaaaagagggttatgagttagaaagggtcac  
 atgattacagaggctgagaagtcccacagcagattgtctgcaagctggagaccagggatac  
 tggtagcatggctcagccaagtcccgaagcctcagaatcaggaaagctgatgatataattc  
 ttagcccaaaggccttagaaccccagcgggtgacggaaaggctgatgtaggtcctggagtcct  
 gagaccacacagcctgggatcctgaaatccaagggcaggaatggaagcgtgtattccagctc  
 caagagagtaagaccaatttgcctttcttccgtttttgtttcaagccacctgcacattgagg  
 ggggatgggttccctcttagtccattcagtcataatcaatctctctggaataacctcaca  
 gacacactaacaataatgcctttccagttctcttaggtattctttaatccagtcagctgac  
 acctaaaattaacctcacaaaagtaaggagaaagaagacaactttaggggaggctgcta  
 tgcaagacagtggtggaaggaagggtctctgaagagggttaatatctgagcagagacttgaa  
 tgaagtgaagaagtgagccatgtgggtatggggaatacaactccaggtagagaagacaagt  
 gtgggtgtgatcagggtcagcaagaagccatgtgacagagaagggtgggcccaggagagac  
 ggataagtgatctaactcctgaggagggtggcctggccaggagctagagcatgaagatctcgt  
 aggactttattctgcaaggtgaaaagccattgtattagctgttcacaaacccgagactagg  
 caatttacaaaagaagagaggtttaatggacttacagttccacatggctggggaggcctca  
 caatcatggcgaaaggcaatgaggagcaagtcacgtcttacgtggatggcaggcaaagaca  
 agacagcttgtgcagagaaactcccccttatagagccatcagatcctgttagacttattcac  
 tatcacaagaacagcagggtaagacctgtccccatgattcagttacctcccactgggtccc  
 tcccacaacgcatgggaattcaggatgagatttgggtggggacacaaccaaacctatcatt  
 ccacccatggcccctcccaaatttcatgtcctcacatttcaaaaccaatcacaccatcccaa  
 cagtccctcaaagctttaaattgatttcagcattaactcaaaagtccacagtctaattgtctca  
 tctgagacaaggcaagtcctttccatttatgagcctataaaaatccaaagcaagttagttact  
 tcctagatacaatgggggtacaggcattgggtaataacagccattccaaatgggataaattg  
 gtcaaaacaaagaggctacaggccatgagagtcacaaaatccagtggggagcagcaaatctta  
 aagctccaaaatgatctcctttgactccacatctcacatccaggtcacgcagatggaagggg  
 tgggttcccattggtcttgggcagctctgcccctgtacctttgcagggtacagcctccctctc  
 agctgctttcatgggctggcattgagtgctgagcttttccaggtacacgggtgcaagctgt  
 cgggtggatctaccattctggggctctggaggacctctctcacagctccactaggtgggtgcc  
 cagtagggactgtgtgtggggctctgaccccacatttcccttctgactgcctggcagag  
 gatctccatgagggccctgctcctgcagcaaacctctgactgggcatccaggcatttccgca

**Fig. 8 cont.****Hybrid Rhesus box of RHD negatives**

catcctctttaatctaggcgaagggttccaaccccaattcttgacttctgtgcaactcgag  
tctcaacaccacatggaagctgtcaaggcttggggcttgcaactccccgaagctacagccaa  
gctctaccttgccctcccgtcagtcaggttgggagtggtgggatgcagggcaccagctccc  
taggctgcacacagcatgaggacccccggcctggccaacaaaaccatttttctgatacct  
ctggacctgtgatgggaggggttgcataaagacctctgacatgacctggagacattttccc  
cattgtcttgggaattagcatttggctcctgttactcatgcaaatttctgcagccagcttga  
atctctcctcagaaaatgggaatttttctttctatcacattgtcaggctgcaaattttccg  
aactttatgctctgcttcccttataaaactgaaatgtctttaacagcacccaagtcacctct  
tgaatgcttggctgcttagaaatttctcctgccagatactctaaatcatctctctgaagttc  
aaagttctacaaatatctcgtgcaggggcaaaatgccgccagtatctttgctaaaacataac  
aagagtccccttggctccagttcccaacaagttcctcatttccgtctgagaccacctcagcc  
tatggactttattgtccacagtgctatcagcattttgggcaaagccattcaacaagtctcta  
ggaagttccaaacttccacatttgcctgtcttctctgagccctccaaactgttccaaac  
cctgcctgttaccagttccaaagtcacataccatttttgagtatctacggcagcacccca  
ctctactggtaccaattagccactgaagtagttggagaacagaagtaatagactctggttt  
acattgtaaaagcttctctgtggctgctgtgtgaagaaaatataatgagaatgaagcccaag  
atgaagcaggacacagttgcagtggttagagtaagaaaatgctgctggctggcactgaagtg  
atagcctggaggttgggtgtgcacatgcagtggtatgtgttttacgatagtaggccaaca  
gatactgtaatccacacttgttttttttttttttttggagacagagtctcacctgttgcttag  
actagaatgcagtgggcacaatcttggctcactacaacctccacctcccaggttcaacaatc  
cttgtgcttcagcctcccagtagttgggattacaggtgtgtgccaccgtgccagctatat  
ttttgtatttttagcagagatgggattttgccacattggccaggctggtcttgaactcctg  
gcctcaagcaatcctcccacttagcctcccaggtgctgagccaccacacctggccgcaac  
tgatttttaatcatgaaatgacacatacatttaaaaaacccaataacctataatattcctggc  
tagtactctcacatctatatcatcaaaaacaaagaaatgtgtgaaactgacacagccaag  
gggagactaaggagacataacaattaactgtaatgtgggtattctggaggggatcctggaaca  
gaaaaagacattaggcaaaaaactaaagaaatctgaataaaatgtggatgtcagttataat  
aatgtatcatattagtcagtaattgtaacaaatataaccaataatgaaagccattaattat  
agggaaaatggaggggttaatatgggtggctggcttttctattttctagcagctccatttta  
tctacaaaagacaaacattcattaagtcacaaaaggtaagaatgacaaattaagcatgta  
tcttattagtaagagtaataaaagatgctcactcatattataaatatttgacaatgatgt  
taaggccagaaaagagaaaaaagggtaggggcaaaaaacgaaagagaaaggagttagatc  
tttctcccgcactcattagctattaaaagaggatgtttgtttaaagctgctcagagctggt  
aaactaatgttaagtcactaacgggaatttaaaaggtttcattaagaactgcctgcactaga  
ttcctccacctgagacattaacaatcacgataaacctcctgagtggtgaagaacgtgtcca  
tttaaaaacaggctatagattgtcatgcagttttatctactaatcggctaagcaccgcaa  
aaacaaacaaaaaaacccaagggtgaaagtttcatccatcaaaggaaacaacagtcacct  
tggttcccatcccactcatatactgccgccgtacatgtcaatcagatgaacctgtgcgtatc  
tcttaacgacaattgaccacctttttaactgaagtgaaaggggggttctgctccgaccac  
ttcctggatctctccctcaccctctgtgttctttcgggtgcaccatcgggtcaaagccgca  
gcaacgccgtctctgtgtgatcgcagtgcccttctgcacacgacctcccccgagagtac  
cagctaccggacaggcaccaggagggtaccgagcaccctccggaccggcggctgcaggat  
cgcgagcgcctccgctaggagaccgcaggttgcgcctgtgcttctgcggtggcgccttct  
gcaaggagacctcgacctgctccctctccggggctggatctgactccttgacggtgattcc  
agacgcgagacccaactgacggcttctagaagaggggagcccgccgcaagctttcac

**Fig. 8 cont.****Hybrid Rhesus box of RHD negatives**

gtagctaagtcacgcttgcttccggcttcttaccgttctccccttgtaaacggttacctcc  
 cgaaaaccaggctctcctccaacagtggttctcaagcgaggcgatcttcccgggagggga  
 tatttgcaaagtctggggcatttttggttcaactggggctgctacttgcacccactgggta  
 gaggcgggggatgcagctacacaacctgcaagcaggggacagcaccctccccaaccagac  
 agaattagccggcccaaacctcagtagtgccaggctgagaaacctgccttaaaaaaca  
 acaaagaaaggccaagtcccataagtggtcaccgcgcccagactgggggtccacgggacacc  
 ccagccacgccaagccgggaagtccccgcctcctggagctgaaccgcccctctcccagagg  
 tggagctgggggggggggaaacaggcaccggagaaaataaaacaagactaaaaagtccctgagta  
 gcgctgtgtggccgcaaacctgaaccaccttttgaccacgcgggacccggcactcttctt  
 gccaccacccctgagagggctgcgcgccgaccccagtagtaaaacactcgtcacctca  
 ctcaagacgggtacgaaggccaacggacgccttctttagaacgctcagcacacagagcaac  
 ttctcagcctactctcaaattggcgtactccaaactagcactcccagcgtccagctgtgaac  
 ccagagcggcggaaagcccctgaaccagcgcggggcatgcgagacgcgttgttgggtg  
 ggcgtggctccctccggacccggcgccccgcctccgccccgtgtccgcatgcgcgactgag  
 ccgggtggatggactgctgcatccgggtgtctg  
 (end of Rhesus box)  
 gaggctgtggccgttttgttttcttggctaaaatcgggggagtgaggcgggcccggcgcggc  
 3'

Fig. 9

## Upstream Rhesus box of D-positives

5' ctagaaaacactttgtcatttttagaggtgta  
 (start of Rhesus box)  
 tccaatggttcgcgaggcactggagtcagagaaaatggagttgaatcctttctctgccactc  
 ttgaggagaatctcaccatttattatgcactgtagaatacaacaataaaatacagccatgt  
 accacataacaacatcttggtaaacaacagactgcatatatgatgggtggtcatccagtaagc  
 taaggttaatattattatttccctttttttttcttttttttgagatgtagtcttactctg  
 tcaccagggttagagtgcaatggcaccatcttggctcactgcaacctctgcctcctgggttc  
 aagcqaatctcctgcctcagcctcgaagttagctgggaattacagggcaccacacatctgg  
 ctaattttttgtatttttagtaagatgggggttcaccatggtggccaggctgatctcaaac  
 tcctgacctcaagtgatctgcctgcctcggcctcccaaagtgctgggaccataggcctgagc  
 cactgtgcccgccctggtttgcttttaacagttaacagtgtgctcatagaaactgctttg  
 acatgactgcaatcatgtgcttcatagaaacttaattagattataccactagagtcttcaga  
 tttttatactttttttttgaaacggagctctcactctgtcaccaggctggagtgcagtgccg  
 caatctcagctcggcgaacctccgcctcccagggtcaagtgattctcctgcctcagcctcc  
 cgagttagctgggattacaagtgcacactaccacgcccagctaatttttgcaattttactaga  
 cagggtttcaccatggttggttaggatagtttcaccaggatctcttggcctcatgatcagcct  
 gcctcggcctcccaaagtgctgggattacaggtgtgagccaccgtgccagcctatacttcc  
 ctttttgaaatacatttggcgttttgagaattaacagctttgtgaacgtggcagtgcttgt  
 gattcaggcttccactgagaccaaggggagaacctggttgaggacaaacagacggacagcg  
 tgtggcagtgtttaaatgctcttctgaaggctgatacgacagctctctgtgcaactgattgca  
 tacgcatcccaagattatattatgttttctattgctatgtgtcacactttgccaaacagga  
 tgtggaaaatgaataagcggttttcttaggcacttcttaacagacaattgggtcaaatgaac  
 tccattgcttaagaaacacataaacaccatttagtcaactgaatatagctatatgtatgggtg  
 ctactatggggaatcttgttttgccaattttctttgaaaattctggcagaccaaggttcttt  
 ttgtttacacaatacttgaaaaataaaaatgaacaagccaacaaactaccaagttttcactt  
 acataaatgtagttacatacagaaaatgtgactgtgaattttttctaggacttttaactat  
 aagcactatttgcacgaaagagaaccaatctatcaattacaaactcacataattttacagat  
 tttttttccctacacagcacataaaacagaaggaatttgaagccaccctccaaacacaggg  
 gaaggaggctgtgtgtatctcctcattgtctttcacattctaagggtggttccactcagtgac  
 tgaaatccttaagtgttattagtcggcttgggctaccataacagcagcttaaactgttgt  
 taagccactcagacttaacaacagaaatttatttcccttatagttctggaggctggaagtcc  
 aagggtgccggcaaggctggtttctggtgagacctctctccctgtcttgcagatggctgcctc  
 ctccctgtgtcctcatagagcctgtcctctgcttttacacttctgggtgcatcttccctttt  
 ttttttttttgagacagagtctcgctctatcgcccaggctggagtgagtgcccgatcga  
 tctcggctcactgcaacctctgcctcccagggtcaagcaattctcctgcctcagcctcccga  
 gtagctgggactacaggtgcccgccatcatgtctggctaatttttgatttttagtagagac  
 aggtttcaccaatattggccaggctggtctccaactcctgacctgtcatctgcctgcctcg  
 gcctcccaaagtgctaggattacaggcgtgagccaccgcaccggcctctttctcttctat  
 aaggacaccagtcctattagattagggctccaccctcatgacctcatttgaccttaactatt  
 atttctttaagcacctatttccaaatatagtcactttaggggttagggcttcaaatatga  
 atctgaggagatcaattcagtaaatagcagtagtcattaacggacaatatatacaaagata  
 atttcgtgattactgtccttatgcataaatgtcctcagtggtccactgcctttatccagatt  
 tactatcacaagactttgctctgagaaaaatgtgatttctttctttttttttttttttgga  
 gacagagtctcactctgtcaccaggctggagtgagtggtgcaatctcggctcactgcaat  
 ctccgcctcccagggttcacgccattctcttgcctcagctctcccgagttagctgggacctacagg

**Fig. 9 cont.**

## Upstream Rhesus box of D-positives

cgcccgccaccctgccagctaattttttgtatTTTTtagtagagacggggtttcaccatggt  
 agccaggatggctcaatctcctgacctcgtgatccacctgcctcagcctcccaaagtgtg  
 ggattacaggcatgagccaccgcgccagcagattTTTTTTTTTTTTTTTTTTTTTgagat  
 ggagtcttgctgtgttgccagcctggagtgagtggtatgattttggctcactgcaacctc  
 tgtctaccatgttcaagcgattctcccacctctgcctcccgtgtagctgggatcacaggcac  
 acgccaccacacctagctactTTTTgtatTTTTtagtagaaatggggtttcaccatggtggc  
 aggatggctccgaactcctgacctcaagtgatcctcctgcctcggcctcccaaagtgtggg  
 attacagggtgtagccactgtcctggccaaaaatgtgattttttttttccacattgccc  
 ttccatttcaattaactataatagctatgtctattgagcactcaagtgtattctagaaactg  
 ttctgattctggggatataatccatgaatcaactatagtcctgttattaagtaatctgtag  
 tctgactaaaccattagaaatttaaaaaatggctactttcaagacatcttgagggtcagga  
 gtcccacactgcaaccatattacctaataatccaacctgcttgtaattcacttatttaacc  
 aatatttattgagtgccaaactttgagcctaagatacagcagtaaacaaatggataaagtccc  
 tgtcctcatgaaacttgatttcaatggaagaaacagaaaacaaacagatataggatgtaat  
 atcaggtagggataaataactttgaattcaacaaaagtatacgtagtcagggttcgccaaag  
 agacacagccaatcggatacatagatatataaaagagggtttatgagttagaaagggctcac  
 atgattacagaggctgagaagtccacacagcagattgtctgcaagctggagaccagggatac  
 tggtagcatggctcagtcacaagtcccaaagcctcagaatcaggaaagctgatgatataatc  
 ttagcccaaaggccttagaaccccagcgggtgacggaaaggctgatgtaggtcctggagtcct  
 gagaccacaacagcctgggatcctgaaatccaagggcaggaatggaagcgtgtattccagctc  
 caagagagtaagaccaatttgcccttcttccgtttttggttcaagccacctgcacattgagg  
 gcggatggttccctcttagtccattcagtcataatcaatctcttctggaaataccctcaca  
 gacacactaacaataatgccttccagttctctaggtattctttaatccagtcaggctgac  
 acctaaaattaaccatcacaaaagttaaggagaaagaagacaactttagggggaggctgta  
 tgcaagacagtggtggaaggaagggtctctgaagaggttaatatctgagcagagacttgaa  
 tgaagtgaagaagtgagccatgtgggtatggggaatacaacttccaggtagagaagacaagt  
 gtggtgtgtatcagggtcagcaaagaagccatgtgacagagaagggtgggcccaggagagac  
 ggataagtgatctaactcctgaggaggtggcctggccaggagctagagcatgaagatctcgt  
 aggactttattctgcaaggtgaaaagccattgtattagctctgttcacaaaccgagactagg  
 caatttacaaaagaaagagaggtttaatggacttacagttccacatggctggggaggcctca  
 caatcatggcgaaaggcaatgaggagcaagtcacgtcttacgtggatggcaggcaagacaa  
 agacagcttggtcagagaaactcccccttatagagccatcagatcctgttagacttattcac  
 tatcacaagaacagcacgggtaagacctgtccccatgattcagttacctcccactgggtccc  
 tcccacaacgcatgggaattcaggatgagatttggtggggacacaaccaaacctatcatt  
 ccacccatggccccctcccaaatttcatgtcctcacatttcaaaaccaatcacaccatcccaa  
 cagtcctcctcaaagtcttaaatgatttcagcattaactcaaaagtccacagtcctaatgtctca  
 tctgagacaaggcaagtccttccatttatgagcctataaaatccaagcaagttagttact  
 tctagatacaatgggggtacaggcattgggtaaatacagccattccaaatgggataaattg  
 gtcaaaaacaaagaggctacaggccatgagagtcacaaatccagtggggcagtcacaaatctta  
 aagctccaaaatgatctccttggactccacatctcacatccaggtcacgcagatggaagggg  
 tgggttcccatggtcttgggcagctctgcccctgtaccttgcagggtacagcctccctctc  
 agctgcttcatgggctggcattgagtgctctgcagcttttccaggtacacgggtgcaagctgt  
 cgggtggatctaccattctggggtctggaggacctctctcacagctccactagggtgggtgcc  
 cagtagggactgtgtgtggggtctctgacccacatttcccttctgactgcctcggcagag  
 gatctccatgagggccctgctcctgcagcaaaacttctgactgggcatccaggcatttccgca

**Fig. 9 cont.****Upstream Rhesus box of D-positives**

catcctctttaatctaggcgaagggttccaaacccaattcttgacttctgtgactcgcag  
 tctcaacaccacatggaagctgtcaaggcttggggcttgactccccgaagctacagccaa  
 gctctaccttgctcccgtcagtcattggttgggagtggtgggatgcagggaccaagtccc  
 taggctgcacacagcatgaggacccccgggctggccaacaaaaccatTTTTCTGATACCT  
 ctggacctgtgatgggaggggttgccataaagacctctgacatgccctggagacattttccc  
 cattgtcttgggaattagcatttggctcctgttactcatgcaaatttctgcagccagcttga  
 atttctcctcagaaaatgggaattttctttctatcacattgtcaggctgcaaattttccg  
 aacttttatgctctgcttccctataaaactgaaatgtctttaacagcaccacaagtccctct  
 tgaatgctttgctgcttagaaaatttctcctgccagatactctaaatcatctctctgaagttc  
 aaagttctacaaatctcgtgcaggggcaaaatgccgccagtatcttggctaaaacataac  
 aagagtccccttggctccagttccaacaagttcctcatttccgtctgagaccacctcagcc  
 tatggactttattgtccacagtgctatcagcatttgggcaaaagccattcaacaagtctcta  
 ggaagttccaaacttcccacatttgctgtcttctctgagccctccaaactgttccaaac  
 cctgctgttaccagttccaaagtcacatacccatttttgagtatctacggcagcaccacca  
 ctctactggtaccaatttagccactgaagtagttggagaacagaagtaatagactctggttt  
 acattgtaaaagcttctctgtggctgctgtgtgaagaaaatataatgagaatgaagcccaag  
 atgaagcagggacacagttgcagtggttagagtaagaaatgctgctggctggcactgaagtg  
 atagcctggaggttgtgtgtgcacatgcagtgatgtgttttacgatagtaggccaaca  
 gatactgtaatccacactgtttttttttttgagacagagtctcacctgtgcctagacta  
 gaatgcagtgggcacaatcttggctcactacaacctccacctcccaggttcaacaatccttg  
 tgcttcagcctcccagtagttgggattacaggtgtgtgccaccgtgcccagctatattttt  
 tgtatttttagcagagatgggatttggccacattggccaggctggtcttgaactcctggcct  
 caagcaatcctcccacttagcctccaaagtgctgagccaccacacctggccgcaactgat  
 ttttaatcatgaaatgacacatacatttaaaaaaccaatacctataatattcctggctagt  
 actcttcacatctatatcatcaaaaacaaagaaagtatgtgaaactgacacagccaagggga  
 gactaaggagacataacaattaactgtaatgtggattctggaggggatcctggaacagaaa  
 aagacattaggcaaaaaactaaagaaatctgaataaaatgtggatgtcagttaataataatg  
 tatcatattagtcagtaattgtaacaaatataccacaataatgaaagccattaattatagg  
 gaaaatggaggggttaatatgggtggctggcttttgctatttctagcagctccattttatct  
 gcaaaagacaacattcattaagtcccaaaaaggtaaagaatgacaattaagcatgtatct  
 tattagtaagagtaataataaagatgctcactcctatttataaataatttgacaatcatgttaa  
 ggccacaaaagagaaaaaaggtaggggcaaaaaacgcaaagagaaaggagttagtatcttt  
 tctcccgcactcattagctattaaaaggagatgtttgtttaagctgctcagagctggtaaa  
 ctaatgttaagtactaacgggaatttaaaaggtttcattaagaactgcctgcactagattc  
 ctccacctgagacattaacaatcacgataaacctcctgagtggttaagaactgtccattt  
 aaaaacaggctatagattgtatcatgcagttttatctactaatcggctaataatcccgccaaa  
 acaaaaaaccccaagggtgaaagtttcatccatcaaaggaaacaacagtcaccttggtt  
 ccatctcactcatatactgccgccgtacatgtcaatcagatgaaacctgtgcgtatctctta  
 atgacaattgaccacatttttgactgaagtgaagggggttctgctccgcgaccacttct  
 ggatctccccctccacctctgtgttcttctgggtgcaccatcgggtcaaagccgcagcaac  
 gccgtctctgtgtgatcgcatgtgcccttctgcacacgaccttccccgagagtgaccagct  
 accggacaggcaccaggagggtaccgagcacctcccggaccggcggtgcaggatcgca  
 gcgctccgctagggagactgcacgttgcgctgtgcttctgcggtggcgcttctgcaag  
 gagacctcgacctgctccctctccggggtggatctgactccttgacgggtgattccagacg  
 cgagacccaactgacggcttctagaagaggggagcccgccgcaagtctttcacgtagc

**Fig. 9 cont.****Upstream Rhesus box of D-positives**

taagtcacggttgcttccggcttcttaccggttctcccctttgtaaacggttacctcccgaaa  
 acccaggctctcctccaacagtggttctcaagcgaggcgatcttccccgggaggggatattt  
 ggcaaagtctgggggcatTTTTGGTtactggggctgctacttgcacccactgggtagaggc  
 ggggatgcagctacacaacctgcgaagcacgggacagcacccctccccaaccagacagaat  
 tagccggcccaaaacctcagtagtgcccaggctgagaaacctgccttaacaaacaacaaa  
 gaaaagccaagtcccataagtgggtcaccgcgcccagactgggggtccacgggacaccccagc  
 cacgccaagccgggaagtccccgcctcctggagctgaàcccgcccctctcccagaggtggag  
 ctgccccggggcggaacaggcacggagaaaataaacaagactaaaaagtctgagtagcgct  
 gtgtggccgcaaacctgaaccaccttttgaccacgcccgggacccggcacgcttctgccac  
 ccaccctgagagggctgcgcccggaccccagtagaaaacactcgtcacctcaatcaa  
 gacgggtacgaaggccaacggacgcttctttagaacgctcagcacacagagcaacttctc  
 acgcctactctcaaatggcgtactccaaactagcactcccagcgtccagctgtgaaccaga  
 gccggcgaaagcccctgaaccacagcggggcatgcgcagacgcggttgttgggtgggct  
 ggctccctccggacccggcggcccgcctccgccccgtgtccgcatgcgcgactgagccgcg  
 ggggtggtagctgctgcacccgggtgtctg  
 (end of Rhesus box)  
 aagatccgatgaaataacatatgcaaaatgattgggtccgtgattggcattccagaaatgg  
 3'

Fig. 10

## Upstream Rhesus box of D-positives

5' ctagaaaacactttgtcatttttagaggtgta  
 (start of Rhesus box)  
 tccaatgttcgcgcaggcactggagtcagagaaaatggagttgaatcctttctctgccactc  
 tttgaggagaatctcaccattttattatgcactgtagaatacaacaataaaatacagccatgt  
 accacataacaacatcttggtaaacaacagactgcatatatgatgggtggatccagtaagc  
 taaggtaattttattattattccctttttttttcttttttttgagatgtagtcttactctg  
 tcaccaggctagagtgcaatggcaccatcttggctcactgcaacctctgcctcctgggttc  
 aagcgaatctcctgcctcagcctccgaagtagctgggaattacaggcaccaccacatctgg  
 ctaattttttgtatttttagtaaaagatggggtttaccatgttggccaggctgatctcaaac  
 tctgacctcaagtgatctgcctgcctcggcctccaaagtgtgggaccataggcctgagc  
 cactgtgcccggcctgtttgcttttttaacagttaacagtggtgctcatagaaaactgctttg  
 acatgactgcaatcatgtgcttcatagaaacttaattagattataccactagagtcttcaga  
 tttttatactttttttttgaaacggagtctcactctgtcaccaggctggagtgcaagtgcg  
 caatctcagctcgcgcaacctccgcctcccagggtcaagtgattctcctgcctcagcctcc  
 cgagtagctgggattacaagtgcaactaccacgcccagctaatttttgcattttactaga  
 cagggtttaccatgttggctaggatagtttaccaggatctcttggcctcatgatcagcct  
 gcctcggcctcccaaagtgtgggattacagggtgtgagccaccgtgccagcctatacttcc  
 cttttgaaataccatttggcgttttgaagaattaacagctttgtgaacgtggcagtgctgt  
 gattcaggcttccactgagaccaaggggagaacctggttgcaaggacaaacagacggacagcg  
 tgtggcagtggttaaatgctcttctgaaggctgatacgacagctctctgtgcaactgattgca  
 tacgcatcccaagattatattattgttttctattgctatgtgtcacactttgccaacagga  
 tgtggaaaatgaataagcggttttcttaggcacttcttaacagacaattgggtcaaaatgaac  
 tccattgcttaagaacacataaacaccatttagtcaactgaatatagctatatgtatgggtg  
 ctactatggggaatcttgttttgccaattttctttgaaaattctggcagaccaaggttcttt  
 ttgtttacaacaataacttgaaaaataaaaatgaacaagccaacaactaccaagttttcactt  
 acataaatgtagttacatacagaaaatgtgactgtgaattttttctaggacttttaaactat  
 aagcactatttgcacgaaagagaaccaatctatcaattacaactcacataattttacagat  
 tttttttccctacacagcacataaaaacagaaggaatttgaagccaccctccaaacacaggg  
 gaaggaggctgtgtgtatctcctcattgtctttcacattctaaaggtgggtccactcagtgac  
 tgaatccttaagtgttgatttagtcggcctgggctaccataacagcagcttaactgttgt  
 taagccactcagacttaacaacagaaatattttccttatagttctggaggctggaagttc  
 aaggtgccggcaaggctgggttctgggtgagacctctcctcctgtcttgagatggctgcctc  
 ctccctgtgtcctcatagagcctgtcctctgcttttacacttctgggtgcatcttccctttt  
 ttttttttttgagacagagctctcgctctatcgcccaggctggagtgcaagtgcccgatcga  
 tctcggctcactgcaacctctgcctcccagggtcaagcaattctcctgcctcagcctcccga  
 gtagctgggactacaggtgcccgccatcatgtctggctaatttttgtatttttagtagagac  
 aggtttcaccatattggccaggctgggtctccaactcctgacctgtcatctgcctgcctcg  
 gcctcccaaagtgttaggattacaggcgtgagccaccgacccggcctctttctctcttat  
 aaggacaccagtcctattagattagggctccaccctcatgacctcatttgaccttaactatt  
 atttctttaagcacctatttccaaatattagtcacttttaggggttagggcttcaaaatgga  
 atctgaggagatcaattcagtaaatagcagtagtcattaacggacaatatatacaaagata  
 atttcgtgattactgtccttatgcataaatgtcctcagtggtccactgcctttatccagatt  
 tactatcacaagactttgctctgagaaaaatgtgatttctttctttttttttttttttga  
 gacagagctctcactctgtcaccaggctggagtgcaagtggtgcaatctcggctcactgcaat  
 ctccgcctcccagggtcacgccattctctgcctcagctctccgagtagctgggcctacagg

**Fig. 10 cont.**

## Upstream Rhesus box of D-positives

cgcccgccaccctgccagctaattttttgtattttagtagagacggggtttcaccatggt  
 agccaggatgggtctcaatctcctgacctcgtgatccacctgcctcagcctcccaaagtgctg  
 ggattacaggcatgagccaccgcgccagcagatTTTTTTTTTTTTTTTTTTTTTgagat  
 ggagtcttgctgtggtgccagcctggagtgagtggtatgatTTTGGGTCactgcaacctc  
 tgtctaccatggtcaagcgattctcccacctctgcctcccgtagctgggatcacaggcac  
 acgccacca'cacctagctactttttgtattttagtagaaatggggtttcaccatggtggcc  
 aggatgggtcccgaactcctgacctcaagtgatcctcctgcctcggcctcccaaagtgctggg  
 attacaggtgtgagccactgtgcctggccaaaaatgtgatttcttattcccacattgccaa  
 ttccatttcaattaactataatagctatgtctattgagcactcaagtgtattctagaaactg  
 ttctgattctggggatataatccatgaatcaactatagtcctgttattaagtaatctgtag  
 tctgactaaaccattagaaatttaaaaaatggctactttcaagacatcttggagttcagga  
 gtcccacactgcgaaccatattacctaataatccaacctgcttgaattcacttatttaacc  
 aatatttattgagtgccaaacttgagcctaagatacagcagtaaacaaatggataaagtccc  
 tgtcctcatgaaacttgattctaattggaagaacagaaaaacacagatataggatgta  
 atcaggtaggataaataactttgaattcaaacaaaagtatacgtagtcagggttcgccaaag  
 agacacagccaatcggatacatagatatataaaagagggttatgagttagaaagggtcac  
 atgattacagaggctgagaagtcccacagcagattgtctgcaagctggagaccagggtac  
 tggtagcatggctcagtcacaagtcccaaagcctcagaatcaggaaagctgatgataatc  
 tttagccaaaggccttagaaccccagcgggtgacggaaaggctgatgtaggctcctggagtcct  
 gagaccacacagcctgggatcctgaaatccaagggcaggaatggaagcgtgtattccagctc  
 caagagagtaagaccaatttgctttcttccgtttttggttcaagccacctgcacattgagg  
 gcgatgggtccctcttagtccattcagtcataatcaatctcttctggaaataccctcaca  
 gacacactaacaataatgcctttccagttctctaggatattctttaatccagtcagactgac  
 acctaaaattaacatcacaaaagttaaggagaagaagacaactttaggggagggtgcta  
 tgcaagacagtggtgaaggaagggtctctgaagagggttaatatctgagcagagacttgaa  
 tgaagtgaagaagtgagccatgtgggtatggggaatacaactccaggtagagaagacaagt  
 gtgggtgtgatcagggtcagcaagaagccatgtgacagagaagggtgggcccaggagagac  
 ggataagtgatctaactcctgaggaggtggcctggccaggagctagagcatgaagatctcgt  
 aggactttattctgcaaggtgaaaagccattgtattagtctgttcacaaaccgagactagg  
 caatttacaaaagaagagaggtttaatggacttacagttccacatggctgggaggcctca  
 caatcatggcgaaaggcaatgaggagcaagtcacgtcttacgtggatggcaggcaagacaa  
 agacagcttggtgagagaaactccccttatagagccatcagatcctggttagacttattcac  
 tatcacaagaacagcacgggtaagacctgtcccctgattcagttacctcccactgggtccc  
 tcccacaacgcattgggaattcaggatgagatttgggtggggacacaaccaaccctatcatt  
 ccacctatggcccctcccaaatttcatgtcctcacatttcaaaaccaatcacaccatcccaa  
 cagtcctcctcaaagtcttaaattgatttcagcattaactcaaaagtcacagtcctaatgtctca  
 tctgagacaaggcaagtcctttccatttatgagcctataaaatccaaagcaagttagttact  
 tctagatacaatgggggtacaggcatgggttaatacagccattccaaatgggataaattg  
 gtcaaaacaaagaggctacaggccatgagagtcacaaatccagtggggcagtcacaaatctta  
 aagctccaaaatgatctcctttgactccacatctcacatccaggtcacgcagatggaagggg  
 tgggttcccattggtcttgggcagctctgcccctgtacctttgcagggtacagcctccctctc  
 agctgctttcatgggctggcattgagtgctctgcagctttccaggtacacgggtgcaagctgt  
 cgggtgatctaccattctggggtctggaggacctctctcacagctccactagggtgggtgcc  
 cagtagggactgtgtgtggggtctctgacccacattcccttctgactgcctggcagag

**Fig. 10 cont.****Upstream Rhesus box of D-positives**

gatctccatgagggccctgctcctgcagcaaacttctgactgggcatccaggcatttccgca  
catcctctttaatctaggcgaaggtttccaaaccccattcttgacttctgtgactcgcag  
tctcaacaccacatggaagctgtcaaggcttggggcttgcaactccccgaagctacagccaa  
gctctaccttgccctcccgtcagtcattggttgggagtggtgggatgcagggcaccaagtc  
taggctgcacacagcatgaggaccccggcctggccaacaaaaccatttttctgatacct  
ctggacctgtgatgggaggggttgcataaagacctctgacatgccctggagacattttccc  
cattgtcttgggaattagcatttggctcctgttactcatgcaaatttctgcagccagcttga  
atctctcctcagaaaatgggaatttttctttctatcacattgtcaggctgcaaattttccg  
aacttttatgctctgcttcccttataaaactgaatgtctttaacagcacccaagtacctct  
tgaatgcttggctgcttagaaatttctcctgccagatactctaaatcatctctctgaagttc  
aaagtctacaaatatctcgtgcaggggcaaaatgccgccagtatcttgctaaaacataac  
aagagtccttcttggctccagttcccaacaagttcctcatttccgtctgagaccacctcagc  
tatggactttattgtccacagtgtatcagcatttgggcaagccattcaacaagtctcta  
ggaagttccaaactttccacatttgcctgtcttctctgagccctccaaactgttccaaac  
cctgcctgttaccagttccaaagtcacatacccatttttgagtatctacggcagcacccca  
ctctactggtaccaatttagccactgaagtagttggagaacagaagtaatagactctggttt  
acattgtaaaagcttctctgtggctgctgtgtgaagaaaatataatgagaatgaagcccaag  
atgaagcagggacacagttgcagtggttagagtaagaaatgctgctggctggcactgaagt  
atagcctggaggtttgtgtgtgcacatgcatgtgtatgtgtttacgatagtaggccaaca  
gatactgtaatccacacttgttttttttttgagacagagtctcacctgttgcctagacta  
gaatgcagtggcacaatcttggctcactacaacctccacctcccaggttcaacaatccttg  
tgcttcagcctcccagtagttgggattacaggtgtgtgccaccgtgccagctatattttt  
tgtatttttagcagagatgggattttgccacattggccaggctggtcttgaactcctggcct  
caagcaatcctcccacttagcctccaaagtgctgagccaccacacctggccgcaactgat  
ttttaatcatgaaatgacacatacatttaaaaaacccaatacctataatattcctggctagt  
actcttcacatctataatcaaaaaacaaagaaatgtgaaactgacacagccaagggga  
gactaaggagacataacaattaactgtaatgtggtattctggaggggatcctggaacagaaa  
aagacattagggcaaaaaactaaagaatctgaataaaatgtggatgtcagttaataaatg  
tatcatattagtcagtaattgtaacaaatataccacaataatgaaagccattaattatagg  
gaaaatggaggggttaatatgggtggctggcttttgctatttctagcagctccattttatct  
gcaaaagacaaacattcattaagtcccaaaaaggtaaagaatgacaaattaagcatgtatct  
tattagtaagagtaataaaagatgctcactcctatttataaatatttgacaatcatgttaa  
ggccacaaaagagaaaaaagggtaggggcaaaaaacgcaaagagaaaggagttagtatcttt  
tctcccgcactcattagctattaaaagaggatgtttgtttaagctgctcagagctggtaaa  
ctaatgttaagtactaacgggaatttaaaaggtttcattaagaactgctgactagattc  
ctccaccctgagacattaacaatcacgataaacctcctgagtggtaagaacttgccattt  
aaaaacaggctatagattgtatcatgcagttttatctactaatcgggtaatatcccgcaaaa  
aacaacaaaccccagggatgaaagttcatccatcaaaggaaacaacagtcaccttggtt  
ccatctcactcatatactgccgccgtacatgtcaatcagatgaacctgtgcgtatctctta  
atgacaattgaccacatttttgactgaagtgaaggggggttctgctccgacacttctct  
ggatctccccctccacctctgtgttctttcgggtgcaccatcgggtcaaagccgagcaac  
gccgtctctgtgtgatcgcattgtgccttctgcacacgacctcccccgagagtgaccagct  
accggacagggcacaaggagggtaccgagcacctcccggaccggggctgcaggatcgca  
gcctcccgtagggagactgcaggttgcctgtgcttctgcggtggcgcttctgcaag

**Fig. 10 cont.**

**Upstream Rhesus box of D-positives**

gagacctgaccctgctccctctccggggctggatctgactccttgacggtgattccagacg  
 cgagacccaaactgacggcttctagaagagggcgagcccggccgcaagtctttcacgtagc  
 taagtcacgttgcttccggcttcttaccgttctcccctttgtaaacggttacctcccgaaa  
 acccaggctctcctccaacagtggttctcaagcgaggcgatcttccccgggaggggatattt  
 ggcaaagtctgggggcatttttggttcactggggctgctacttgcattccactgggttagaggc  
 ggggatgacagctacacaacctgccaagcagggacagcaccctcccccaaccagacagaat  
 tagccggccccaaaacctcagtagtgcccaggctgagaaaaccccgcttaaacaaacaacaaa  
 gaaaagccaagtcccataagtgggtcaccgcccagagactgggggtccacgggacacccccagc  
 cagccaagccgggaagtccccgcctcctggagctgaaccgcccctctcccagaggtggag  
 ctgccccggggcgggaacaggcacggagaaaataaacaaagactaaaaagtcctgagttagcgt  
 gtgtggccgcaaacctgaaccaccttttgaccacgcgggacccggcacgcttctgcccac  
 ccacctgagagggctgcccggccgaccccagtagaaaactcgtcacctcaatcaa  
 gacgggtacgaaggccaacggacgccttctttagaacgctcagcacacagagcaacttctc  
 acgcctactctcaaatggcgtactccaaactagcactcccagcgtccagctgtgaaccaga  
 gcccggaaagcccctgaaccagcgcggggcatgcccagacgcggtgtgtgggtgggcgt  
 ggctccctccggacccggcgcggccctccgccccgtgtccgcatgcccgactgagccgcg  
 ggggtggtactgctgcatccgggtgtctg  
 (end of Rhesus box)  
 aagatccgatgaataacatatgcaaaatgattgggtccgtgattggcattccagaaatgg  
 3'

Fig. 11

ctgatctaca	taggaattgt	tttcaagaca	tttctgcatt	cctctagtga	caggggtgctc	60
actacctcat	gagtatttca	gtggacaact	gtaatggtca	ataaagtatc	cactttccac	120
ctccctgcag	ctcctggccc	tggcttttatt	ctctggggct	ccacacattc	agtttacact	180
cagtggccag	tggctgggac	cattgtagaa	aataaggaaa	ctccaattcc	ttccttcttt	240
tcttcctctt	tcatactctc	ctccctctct	acatcccctc	ctctcttctc	tccttccctg	300
acacttacca	tgtaccagac	cttctgccag	gcacatggat	gggagcacag	gggaaagtgg	360
ctgcagggtt	agaactaagt	ccaagcccc	ctaaagctca	tgccagggga	ctggactgtc	420
cagtaactgag	ggatggggat	gctgaggctg	gtggccttcc	tcaaatgcac	tgtagtgccc	480
caggcagagt	cctgggctgc	cctgtgagga	ggtgaccaga	ggtagagcaa	cttcacccta	540
aggctggatc	aggatcccc	ccaggttttt	actagagcca	aaccacatc	tcctttctct	600
tctggcacc	cccctaaaa	tgcttagaaa	cacatagatt	taaatacaaa	ttcaaatgta	660
agtaatttca	actgtgtaac	tatgaggagt	cagttctacg	tgggtcctat	ctgtatcctc	720
ccpagggctc	agctccattc	tttgctttca	ttcattctca	ttcaatacat	tgttgttaag	780
agctcactgg	gtgccctctc	tgtcatgtag	taaggtttta	aaaagaaagc	ctcttctgag	840
cttcagtttc	cttattcata	aaataggagt	attgatccgt	tccttgcttt	tcttacaagg	900
atatgctgaa	gatgactgaa	gtacagagta	aagaaggatt	atgtttgggt	gtcaaaaggaa	960
tagaatgccc	tctttcaaac	tgagcacagc	aggaacctgt	aacaggaaca	cagcaacttg	1020
ttgaatgaat	gacaatattg	gaaaacatac	atctcctccc	ctccccatca	tagtccctct	1080
gcttccgtgt	taactccata	gagaggccag	cacaaccagc	cttgcagcct	gagataaggg	1140
ctttggcggg	tgtctcccct	atcgctccc	caagcccctca	agtaggtgtt	ggagagaggg	1200
gtgatgcctg	gtgctgggtg	aaccccctgca	cagagacgga	cacaggatg		1249

Fig. 12

2938 CTAGAGAGGGAAGTTTTGAAAATTAACACTGTCTAAATTTCTGCAAAGTTTTTATTCATGAAATTAAGAGTATTTCCCI TAGTCCATTATTCCTCAAGGC RHCE  
 CTAGAGAGGGAAGTTTTGAAAATTAACACTGTCTAAATTTCTGCAAAGTTTTTATTCATGAAATTAAGAGTATTTCCCI TTGTCCATTATTCCTCAAGGC Cde<sup>s</sup>  
 CTAGAGAGGGAAGTTTTGAAAATTAACACTGTCTAAATTTCTGCAAAGTTTTTATTCATGAAATTAAGAGTATTTCCCI TTGTCCATTATTCCTCAAGGC RHD

3038 AATATGGAAGTTTGGATCATATGCTAATCATACTAAAGCTGGATTCTCTTTAAGAGATTGAGAAATTAAGAGGCAAAAGCTGATATATCATGTGTTAGTTA RHCE  
 AATATGGAAGTTTGGATCATCTAATCATACTAAAGCTGGATTCTCTTTAAGAGATTGAGAAATTAAGAGGCAAAAGCTGATATGTCATGTGTTAGTTA Cde<sup>s</sup>  
 AATATGGAAGTTTGGATCATCTAATCATACTAATTAAGAGATTGAGAAATTAAGAGGCAAAAGCTGATATATCATGTGTTAGTTA RHD

\*\*\*\*\* breakpoint region \*\*\*\*\*  
 | \*\*\*\*\* breakpoint region \*\*\*\*\*

3138 TACTGTGAGCTTTATAAGAGCTGGAGGCAACCCCATTAACCTCACCAGAAATACAGAACTCAGTCTCACAACCTTAATAATAATTCCTCTCAAACCTTTTC RHCE  
 TATGTGAGCTTTATAAGAGCTGGAGGCAACCCCATTAACCTCACCAGAAATACAGAACTCAGTCTCACAACCTTAATAATAATTCCTCTCAAACCTTTTC Cde<sup>s</sup>  
 TATGTGAGCTTTATAAGAGCTGGAGGCAACCCCATTAACCTCACCAGAAATACAGAACTCAGTCTCACAACCTTAATAATAATTCCTCTCAAACCTTTTC RHD

3238 CTCAAAGTTTAAATTTGAAAATAATCTTGTGATTAAGAGAAGGCTGTCCACCAATGGACTTATCTGTGTTATTTCTTCTTATTGTGAGCTTAATGGC RHCE  
 CTCAAAGTTTAAATTTGAAAATAATCTTGTGATTAAGAGAAGGCTGTCCACCAATGGACTTATCTGTGTTATTTCTTCTTATTGTGAGCTTAATGGC Cde<sup>s</sup>  
 CTCAAAGTTTAAATTTGAAAATAATCTTGTGATTAAGAGAAGGCTGTCCACCAATGGCTTATCTGTGTTATTTCTTCTTATTGTGAGCTTAATGGC RHD

3337 ATGACAAAAGCAGAGGCAAGAGGCAATACATCAATTTCAAAGTAGGAAGTCAAAGAGGTCAGAGCTCCACAGCATGGCAACAGCTTTGCAGATGCCCA RHCE  
 ATGACAAAAGCAGAGGCAAGAGGCAATACATCAATTTCAAAGTAGGAAGTCAAAGAGGTCAGAGCTCCACAGCATGGCAACAGCTTTGCAGATGCCCA Cde<sup>s</sup>  
 ATGACAAAAGCAGAGGCAAGAGGCAATACATCAATTTCAAAGTAGGAAGTCAAAGAGGTCAGAGCTCCACAGCATGGCAACAGCTTTGCAGATGCCCA RHD

3437 CATCGTGATAGTTGAAATAGCAAAGCCAGCAAAGGTTAAAGCTGAAAATGCCAAAAGCCCTGCCCTTGGCAGCTTTCTGCAGGCATCCCCATGAACATA RHCE  
 CATCGTGATAGTTGAAATAGCAAAGCCAGCAAAGGTTAAAGCTGAAAATGCCAAAAGCCCTGCCCTTGGCAGCTTTCTGCAGGCATCCCCATGAACATA Cde<sup>s</sup>  
 CATCGTGATAGTTGAAATAGCAAAGCCAGCAAAGGTTAAAGCTGAAAATGCCAAAAGCCCTGCCCTTGGCAGCTTTCTGCAGGCATCCCCATGAACATA RHD

3537 GTCAGTAAACAATTTGTTCAAGGCCCCAGTGACCATGAAGAGTGAGGGCTGCAGCCAGGGAATAGTCCCTGCAGAGCAAGGATTCAAATAAGCAGCCGGA RHCE  
 GTCAGTAAACAATTTGTTCAAGGCCCCAGTGACCATGAAGAGTGAGGGCTGCAGCCAGGGAATAGTCCCTGCAGAGCAAGGATTCAAATAAGCAGCCGGA Cde<sup>s</sup>  
 ATCAGTAAACAATTTGTTCAAGGCCCCAGTGACCATGAAGAGTGAGGGCTGCAGCCAGGGAATAGTCCCTGCAGAGCAAGGATTCAAATAAGCAGCCGGA RHD

Fig. 13

2726 TTGTAATCTCTTTTACAGCTACCTCCCATTTCCCTTCTATTCAAGCTAGTAACAAGTTTTCTTTTAAATTCATTTTA'TTAAATGTAAAATAAGTCTA RHCE  
TTGTATCTCTTTTACAGCTACCTCCCATTTCCCTTCTATTCAAGCTAGTAACAAGTTTTCTTTTAAATTCATTTTA'TTAAATGTAAAATAAGTCTA Cde'  
TTGTATCTCTTTTACAGCTACCTCCCATTTCCCTTCTATTCAAGCTAGTAACAAGTTTTCTTTTAAATTCATTTTA'TTAAATGTAAAATAAGTCTA RHD

2826 TTGGAGAAAAAATTTTAAATAGCATCTCTGGAATGCCAGTATGGCTAAATTCATGAATGTTGTCTCCTCAAATGCTGA'AACTGGGAAGCATCTGGCCA RHCE  
TTGGAGAAAAAATTTTAAATAGCATCTCTGGAATGCCAGTATGGCTAAATTCATGAATGTTGTCTCCTCAAATGCTGA'AACTGGGAAGCATCTGGCCA Cde'  
TTGGAGAAAAAATTTTAAATAGCATCTCTGGAATGCCAGTATGGCTAAATTCATGAATGTTGTCTCCTCAAATGCTGA'AACTGGGAAGCATCTGGCCA RHD

2926 AGCTTTGTGGACAGGCCGCTGCCTAGTTTGAATCCCAAGAGCCACTCAATCCAGCCACAAAACATTTGGAAATCTTGGTTCAC'TTCCCTAACCTGAACCTTGT RHCE  
AGCTTTGTGGACAGGCCGCTGCCTAGTTTGAATCCCAAGAGCCACTCAATCCAGCCACAAAACATTTGGAAATCTTGGTTCAC'TTCCCTAACCTGAACCTTGC Cde'  
AGCTTTGTGGACAGGCCGCTGCCTAGTTTGAATCCCAAGAGCCACTCAATCCAGCCACAAAACATTTGGAAATCTTGGTTCAC'TTCCCTAACCTGAACCTTGC RHD

3026 \*\*\*\*\* breakpoint region \*\*\*\*\*  
CCTCTGTGAAATAGGGACATTTAAATAGCTCACTCACAGGGCTGTGAGGACATGTGTTGAGCTGAGGGTCTGCCAGCGGAGACCCTGTGCAGGGGAGAC RHCE  
CCTCTGTGAAATAGGGACATTTAAATAGCTCACTCACAGGGCTGTGAGGACATGTGTTGAGCTGAGGGTCTGCCAGCGGAGACCCTGTGCAGGGGAGAC Cde'  
CCTCTGTGAAATAGGGACATTTAAATAGCTCACTCACAGGGCTGTGAGGACATGTGTTGAGCTGAGGGTCTGCCAGCGGAGACCCTGTGCAGGGGAGAC RHD

3126 \*\*\*\*\* breakpoint region \*\*\*\*\*  
TGTTATCATGGTGATGGATTTCTGCTTCAATTCATTTCTTTTCCAGACAGCATCATATAGAATGAGTTGGGGTGGCG'GTCCAGCAGGTTTGGGTTTATC RHCE  
TGTTATCATGGTGATGGATTTCTGCTTCAATTCATTTCTTTTCCAGACAGCATCATATAGAATGAGTTGGGGTGGCG'GTCCAGCAGGTTTGGGTTTATC Cde'  
TGTTATCATGGTGATGGATTTCTGCTTCAATTCATTTCTTTTCCAGACAGCATCATATAGAATGAGTTGGGGTGGCG'GTCCAGCAGGTTTGGGTTTATC RHD

3226 \*\*\*\*\* breakpoint region \*\*\*\*\*  
CTCTATTCTGCCACTTATTACTTAAAAAAA-----AAAAACCAACTTATATAGTATAAGCTATATCCAGAAAAGTGC'PAATATCATACAAGTACCAATTT RHCE  
CTCTATTCTGCCACTTATTACTTAAAAAAAACCCCAAAAACCAACTTATATAGTATAAGCTATATCCAGAAAAGTGC'PAATATCATACAAGTACCAATTT Cde'  
CTCTATTCTGCCACTTATTACTTAAAAAAAACCCCAAAAACCAACTTATATAGTATAAGCTATATCCAGAAAAGTGC'PAATATCATACAAGTACCAATTT RHD

Fig. 13 cont.

\*\*\*\*\* breakpoint region \*\*\*\*\*  
 3320 GATGAATCTTCTGATATCCACATAAACCACACCCAGAACCTCTTCTTGTCTCATCCAGGATAACCCTAACCCTGACTTCTAACAGCATCAGTCAGTT RHCE  
 GATGAATCTTCTGATATCCACATAAACCACACCCAGAACCTCTTCTTGTCTCATCCAGGATAACCCTAACCCTGACTTCTAACAGCATCAGTCAGTT Cde<sup>s</sup>  
 GATGAATCTTCTGATATCCACATAAACCACACCCAGAACCTCTTCTTGTCTCATCCAGGATAACCCTAACCCTGACTTCTAACAGCATCAGTCAGTT RHD

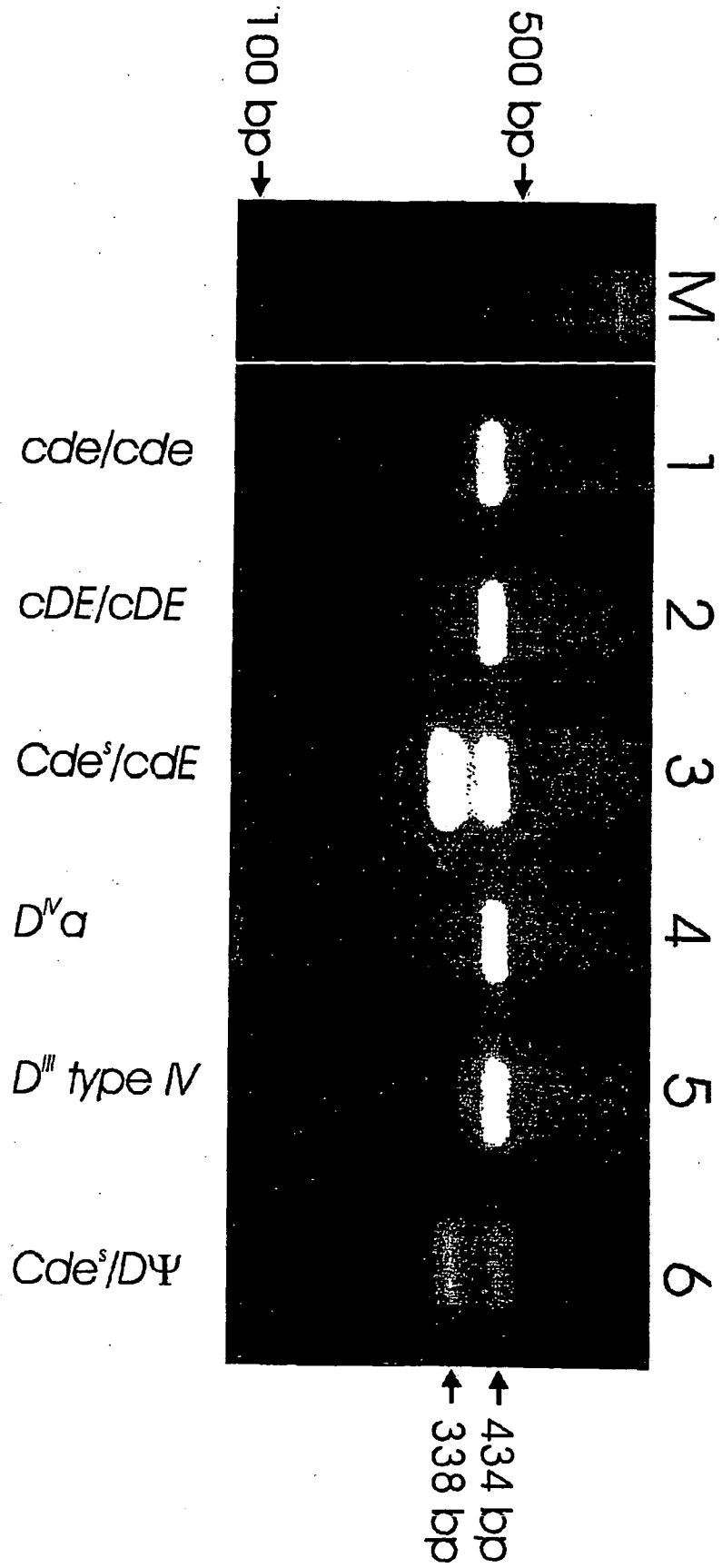
\*\*\*\*\* breakpoint region \*\*\*\*\*  
 3420 TTGTCGTGTTTTGTACATATATATATGTGATGGTTTTGAATGTGTCCCCCAAATTCATGTGCTAGAAACTTAATCCTTCAATTCATATGTTGATGCTATATTT RHCE  
 TTGTCGTGTTTTGTACATATATATATGTGATGGTTTTGAATGTGTCCCCCAAATTCATGTGCTAGAAACTTAATCCTTCAATTCATATGTTGATGCTATATTT Cde<sup>s</sup>  
 TTGTCGTGTTTTGTACATATATATATGTGATGGTTTTGAATGTGTCCCCCAAATTCATGTGCTAGAAACTTAATCCTTCAATTCATATGTTGATGCTATATTT RHD

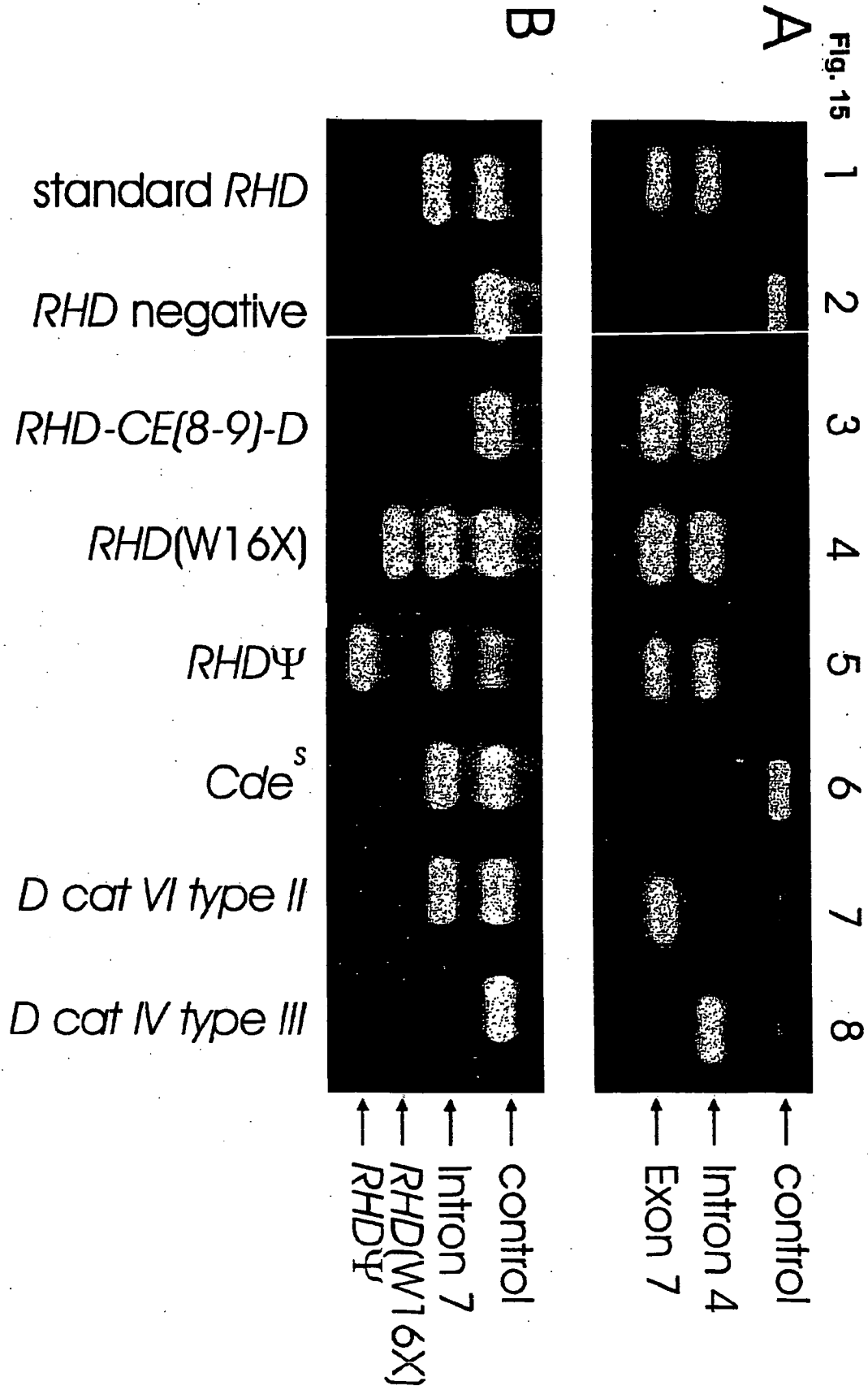
\*\*\*\*\* breakpoint region \*\*\*\*\*  
 3520 GGAGGAAGGGCCTTTGGGAAGTAATTAGGATTAGATAAGGTCAATGGGTGAGGTATGATGGCCTGCTGACTTATAAGAAAGAGAAAGATACTGAGCT RHCE  
 GGAGGAAGGGCCTTTGGGAAGTAATTAGGATTAGATAAGGTCAATGGGTGAGGTATGATGGCCTGCTGACTTATAAGAAAGAGAAAGATACTGAGCT Cde<sup>s</sup>  
 GGAGGAAGGGCCTTTGGGAAGTAATTAGGATTAGATAAGGTCAATGGGTGAGGTATGATGGCCTGCTGACTTATAAGAAAGAGAAAGATACTGAGCT RHD

\* breakpoint region \*|

3620 GGCATGCTTTGCCCTCTCACCGTGTGATGACTTCTCCATGTCATGATGCAGCAAGAGGGCCCTCACCAGATGGTGGCACCATGCTTTTGGACTTCCCAG RHCE  
 GGCATGCTTTGCCCTCTCACCGTGTGATGACTTCTCCATGTCATGATGCAGCAAGAGGGCCCTCACCAGATGGTGGCACCATGCTTTTGGACTTCCCAG Cde<sup>s</sup>  
 GGCATGCTTTGCCCTCTCACCGTGTGATGACTTCTCCATGTCATGATGCAGCAAGAGGGCCCTCACCAGATGGTGGCACCATGCTTTTGGACTTCCCAG RHD

Fig. 14







DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X,D	CARRITT, B. ET AL.: "Evolution of the Rh (rhesus) blood group genes: a 50 year old predication (partially) fulfilled" HUMAN MOLECULAR GENETICS., vol. 6, no. 6, 1997, pages 843-850, XP000998694 OXFORD UNIVERSITY PRESS, SURREY., GB ISSN: 0964-6906 * page 846 * * page 847; figure 4 * -----	1,30	INV. C07K14/47 C12N15/12 C12Q1/68
X	CARRITT, B. ET AL.: "Rh null phenotypes are not due to a gross deletion and can occur on different Rh genetic backgrounds" ANNALS OF HUMAN GENETICS, vol. 57, no. 4, 1993, pages 273-279, XP000926510 * abstract * * page 274, paragraph 3 * -----	1	
X,D	OKUDA H ET AL: "SEQUENCE ANALYSIS OF THE SPACER REGION BETWEEN THE RHD AND RHCE GENES" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS., vol. 263, 24 September 1999 (1999-09-24), pages 378-383, XP000998148 ACADEMIC PRESS INC. ORLANDO, FL., US ISSN: 0006-291X * abstract * * page 378, column 2, last paragraph - page 380, column 1, last paragraph * * page 382, column 1 * ----- -/--	1	TECHNICAL FIELDS SEARCHED (IPC) C07K C12Q G01N
The present search report has been drawn up for all claims			
Place of search Berlin		Date of completion of the search 23 March 2007	Examiner De Kok, Ad
<b>CATEGORY OF CITED DOCUMENTS</b> X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons ..... & : member of the same patent family, corresponding document	

7  
EPO FORM 1503 03 82 (P04C01)



DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	WO 99/37763 A (FLEGEL WILLY A ;WAGNER FRANZ F (DE); DRK BLUTSPENDEDIENST BADEN WU) 29 July 1999 (1999-07-29)	2-27,30,63	
A	* page 1, paragraph 2 - page 27, paragraph 2 *	47-57,59-62,76	
A,D	----- HU, G.: "Human small membrane protein 1 (SMP1) mRNA" EMBL SEQUENCE DATABASE, 4 March 1999 (1999-03-04), XP002170737 HEIDELBERG DE Accession Nr.: AF081282 * abstract *	1,30	
A	----- US 5 972 602 A (SAUL ALLAN ET AL) 26 October 1999 (1999-10-26)	1,26,27,30,48-57,59-63	
A	----- WO 99/47555 A (GENETICS INST) 23 September 1999 (1999-09-23) * page 102 - page 103; claim 26 *	1	
X	----- LE VAN KIM CAROLINE ET AL: "Molecular cloning and primary structure of the human blood group RhD polypeptide." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 89, no. 22, 1992, pages 10925-10929, XP002134526 1992 ISSN: 0027-8424 * the whole document *	1,5-25,30	TECHNICAL FIELDS SEARCHED (IPC)
The present search report has been drawn up for all claims			
Place of search Berlin		Date of completion of the search 23 March 2007	Examiner De Kok, Ad
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons ..... & : member of the same patent family, corresponding document	

7  
EPO FORM 1503 03.82 (P04C01)



DOCUMENTS CONSIDERED TO BE RELEVANT				
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)	
A	<p>CHERIF-ZAHAR B ET AL: "Organization of the gene ( RHCE) encoding the human blood group RhCcEe antigens and characterization of the promoter region." GENOMICS, vol. 19, no. 1, 1994, pages 68-74, XP000926511 ISSN: 0888-7543 * the whole document *</p> <p>-----</p>	1,48		
P,X	<p>WAGNER, F.F. ET AL.: "RHD gene deletion occurred in the Rhesus box" BLOOD, vol. 29, no. 12, 15 June 2000 (2000-06-15), pages 3662-3668, XP000926502 WASHINGTON US * the whole document *</p> <p>-----</p>	1-27,30, 49-57, 59-63,76		
X	<p>BLUNT, T. ET AL.: "Serotype switching in a partially deleted RHD gene" VOX SANGUINIS, vol. 67, no. 4, 1994, pages 397-401, XP000926512 * abstract *</p> <p>-----</p>	2-5,30, 44		TECHNICAL FIELDS SEARCHED (IPC)
X	<p>AVENT ND ET AL: "Evidence of genetic diversity underlying Rh D-, weak D (Du), and partial D phenotypes as determined by multiplex polymerase chain reaction analysis of the RHD gene" BLOOD, W.B. SAUNDERS, PHILADELPHIA, VA, US, vol. 89, no. 7, 1 April 1997 (1997-04-01), pages 2568-2577, XP002108680 ISSN: 0006-4971 * the whole document *</p> <p>-----</p>	2-25, 32-46, 53,59-62		
-/--				
The present search report has been drawn up for all claims				
Place of search Berlin		Date of completion of the search 23 March 2007	Examiner De Kok, Ad	
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p>		<p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons ..... &amp; : member of the same patent family, corresponding document</p>		

7  
EPO FORM 1503 03.92 (P04C01)



DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X,D	<p>DATABASE EM_HUM [Online] EMBL; 13 August 1998 (1998-08-13), COBLEY. V. ET AL.: "Human DNA sequence clone RP3-469D22 on chromosome 1p35.1-36.13, contains the 5'part of the gene for RHCE." XP002178290 retrieved from EBI accession no. HS469D22 Database accession no. AL031284 * abstract *</p> <p style="text-align: center;">-----</p>	2-4,30, 44	
X	<p>WAGNER FF ET AL: "Molecular basis of weak D phenotypes" BLOOD, W.B. SAUNDERS, PHILADELPHIA, VA, US, vol. 93, no. 1, January 1999 (1999-01), pages 385-393, XP002108682 ISSN: 0006-4971 * the whole document *</p> <p style="text-align: center;">-----</p>	5-25	
The present search report has been drawn up for all claims			TECHNICAL FIELDS SEARCHED (IPC)
Place of search		Date of completion of the search	Examiner
Berlin		23 March 2007	De Kok, Ad
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons ..... &amp; : member of the same patent family, corresponding document</p>			

7  
EPO FORM 1503 03.02 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT  
ON EUROPEAN PATENT APPLICATION NO.**

EP 06 02 6286

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on  
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

23-03-2007

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9937763	A	29-07-1999	AT 292176 T	15-04-2005
			AU 748222 B2	30-05-2002
			AU 2274699 A	09-08-1999
			BR 9815090 A	17-10-2000
			CA 2318486 A1	29-07-1999
			DE 69829587 D1	04-05-2005
			DE 69829587 T2	09-02-2006
			ES 2241195 T3	16-10-2005
			JP 2002500884 T	15-01-2002
			PT 1047777 T	31-08-2005
			US 7005276 B1	28-02-2006
			-----	
US 5972602	A	26-10-1999	WO 9506137 A1	02-03-1995
-----				
WO 9947555	A	23-09-1999	AU 3099599 A	11-10-1999
			CA 2324504 A1	23-09-1999
			EP 1064302 A1	03-01-2001
-----				

## REFERENCES CITED IN THE DESCRIPTION

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

## Patent documents cited in the description

- WO 9429469 A [0110]
- WO 9700957 A [0110]

## Non-patent literature cited in the description

- **FILBEY.** *Acta Obstet Gynecol Scand*, 1995, vol. 74, 687 [0003]
- **BOWMAN. J.** *Semin Perinatol*, 1997, vol. 21, 39 [0003]
- **MOURANT.** The distribution of the human blood groups and other polymorphisms. Oxford University Press, 1976 [0003]
- **URBANIAK.** *Transfusion*, 1981, vol. 21, 64 [0003]
- **CHERIF-ZAHAR.** *Hum. Genet.*, 1991, vol. 86, 398 [0004]
- **MACGEOCH.** *Cytogenet. Cell Genet.*, 1992, vol. 59, 261 [0004]
- **CARRITT.** *Hum. Mol. Genet.*, 1997, vol. 6, 843 [0004] [0007]
- **FLEGEL.** *Transfus. Med.*, 1998, vol. 8, 281 [0004] [0006] [0090]
- **OKUDA.** *Biochem. Biophys. Res. Commun.*, 1999, vol. 263, 378 [0004] [0209]
- **GASSNER.** *Transfusion*, 1997, vol. 37, 1020 [0005] [0094]
- **LO.** *Lancet*, 1993, vol. 341, 1147 [0006]
- **COSSU.** *Electrophoresis*, 1996, vol. 17, 1911 [0007] [0090]
- **DÖSCHER.** *Infusionsther. Transfusionsmed.*, 1999, vol. 26 (1), 31 [0007]
- **HUANG.** *Am. J. Hum. genet.*, 1996, vol. 58, 133 [0007]
- **FUJIWARA.** *Hum. genet.*, 1999, vol. 104, 301 [0007]
- **ONDA.** *Gene*, 1995, vol. 159, 225 [0007] [0029]
- **HUANG.** *Blood*, 1996, vol. 88, 2326-33 [0008]
- **FAAS.** *Transfusion*, 1997, vol. 37, 38-44 [0008] [0008] [0225] [0225]
- **FAAS.** *Transfusion*, 1996, vol. 36, 506-11 [0008]
- **AVENT.** *Blood*, 1997, vol. 89, 2568-77 [0008] [0008] [0008] [0234]
- **ANDREWS.** *Blood*, 1998, vol. 92, 1839-40 [0008]
- **CHERIF-ZAHAR.** *Br. J. Haematol.*, 1998, vol. 102, 1263-70 [0008]
- **SINGLETON.** *Blood*, 2000, vol. 95, 12-8 [0008] [0225]
- **SINGLETON.** *Blood*, 2000, vol. 95, 12-18 [0008]
- **OKUDA.** *J. Clin. Invest.*, 1997, vol. 100, 373-9 [0008]
- **FLEGEL.** *Transfus. Med.*, 1998, vol. 8, 281-302 [0008]
- **HYLAND.** *Blood*, 1994, vol. 84, 321-4 [0008]
- **BLUMENFELD.** *Hum. Mutat.*, 1995, vol. 6, 1999 [0029]
- **WAGNER.** *Blood*, 1998, vol. 91, 2157 [0029]
- **HARLOW ; LANE.** *Antibodies, A Laboratory Manual.* CSH Press, 1988 [0045] [0121] [0144]
- **ISSITT, P.D.** *Applied Blood Group Serology.* Scientific Publications, 1985, 239 [0062]
- **BLUNT, T. ; DANIELS, G. ; CARRITT, B.** Serotype switching in a partially deleted RHD gene. *Vox Sang.*, 1994, vol. 67, 397-401 [0062]
- **FAAS, B.H.W. ; BECKER, E.A.M. ; WILDOER, P. ; LIGTHART, P.C. ; OVERBEEKE, M.A.M. ; ZONDERVAN, H.A. ; VON DEM BORNE, A.E.G.K. ; VAN DER SCHOOT, C.E.** Molecular background of VS and weak C expression in blacks. *Transfusion*, 1997, vol. 37, 38-44 [0062]
- **DANIELS, G.L. ; FAAS, B.H. ; GREEN, C.A. ; SMART, E. ; MAASKANT-VAN WIJK, P.A. ; AVENT, N.D. ; ZONDERVAN, H.A. ; VON DEM BORNE, A.E. ; VAN DER SCHOOT, C.E.** The VS and V blood group polymorphisms in Africans: a serologic and molecular analysis. *Transfusion*, 1998, vol. 38, 951-958 [0062]
- **GASSNER.** *Transfusion*, 1997, vol. 37, 1020ff [0063]
- **REID, M.E. ; LOMAS-FRANCIS, C.** *The Blood Group Antigen Facts Book.* Academic Press, 1997 [0065]
- **FLEGEL.** *Transfusion Medicine*, 1998, vol. 8, 281-302 [0094]
- **MAASKANT.** *Transfusion*, 1998, vol. 38, 1015-1021 [0094]
- **LEGLER.** *Transfusion*, 1996, 426-31 [0094]
- **HAMES ; HIGGINS.** *Nucleic acid hybridization, a practical approach.* IRC Press, 1985 [0104]
- **SAMBROOK.** *Molecular Cloning A Laboratory Manual.* Cold Spring Harbor Laboratory, 1989 [0107]
- **AUSUBEL.** *Current Protocols in Molecular Biology.* Green Publishing Associates and Wiley Interscience, 1989 [0107]
- **GIORDANO.** *Nature Medicine*, 1996, vol. 2, 534-539 [0110]
- **SCHAPER.** *Circ. Res.*, 1996, vol. 79, 911-919 [0110]

- **ANDERSON.** *Science*, 1992, vol. 256, 808-813 [0110]
- **ISNER.** *Lancet*, 1996, vol. 348, 370-374 [0110]
- **MUHLHAUSER.** *Circ. Res.*, 1995, vol. 77, 1077-1086 [0110]
- **WANG.** *Nature Medicine*, 1996, vol. 2, 714-716 [0110]
- **SCHAPER.** *Current Opinion in Biotechnology*, 1996, vol. 7, 635-640 [0110]
- **SAMBROOK et al.** *Molecular Cloning, A Laboratory Handbook*. CSH Press, 1989 [0118]
- **HAMES ; HIGGINS.** *Nucleic acid hybridization, a practical approach*. IRL Press, 1985 [0118]
- **OSBORNE et al.** *Curr. Opin. Chem. Biol.*, 1997, vol. 1, 5-9 [0122]
- **STALL ; SZOKA.** *Pharm. Res.*, 1995, vol. 12, 465-483 [0122]
- **SINGLETON, B.K. ; GREEN, C.A. ; AVENT, N.D. ; MARTIN, P.G. ; SMART, E. ; DAKA, A. ; NARTER-OLAGA, E.G. ; HAWTHORNE, L.M. ; DANIELS, G.** The presence of an RHD pseudogene containing a 37 base pair duplication and a nonsense mutation in africans with the Rh D- negative blood group phenotype. *Blood*, 2000, vol. 95 (1), 12-18 [0132]
- **WINTER et al.** *Annu. Rev. Immunol.*, 1994, vol. 12, 433-455 [0187]
- **REBOUL.** *Genome res.*, 1999, vol. 9, 242 [0201]
- **WAGNER.** *Infusionsther Transfusionsmed*, 1995, vol. 22, 285-90 [0211]
- **GASSNER et al.** *Transfusion*, 1997, vol. 37, 1020 [0211]
- **GASSNER.** *Transfusion*, 1997, vol. 37, 1020-6 [0222] [0222]
- **DANIELS.** *Transfusion*, 1998, vol. 38, 951-8 [0225]
- **WAGNER, F.F. ; FROHMAJER, A. ; LADEWIG, B. ; EICHER, N.I. ; LONICER, C.B. ; MÜLLER, T.H. ; SIEGEL, M.H. ; FLEGEL, W.A.** Weak D alleles express distinct phenotypes. *Blood*, 2000, vol. 95, 2699-2708 [0226]
- **ROUILLAC, C. ; COLIN, Y. ; HUGHES-JONES, N.C. ; BEOLET, M. ; D'AMBROSIO, A.-M. ; CARTRON, J.P. ; LE VAN KIM, C.** Transcript analysis of D category phenotypes predicts hybrid Rh D-CE-D proteins associated with alteration of D epitopes. *Blood*, 1995, vol. 85, 2937-2944 [0226]
- **FLEGEL.** *Transfusion*, 2000, vol. 40, 428-434 [0233]
- **WAGNER.** *Infusionsther. Transfusionsmed.*, 1995, vol. 22, 285-90 [0234]
- **ANDREWS KT ; WOLTER LC ; SAUL A ; HYLAND CA.** The RhD- trait in a white patient with the RhCCee phenotype attributed to a four-nucleotide deletion in the RHD gene. *Blood*, 1998, vol. 92, 1839 [0235]
- **AVENT ND ; MARTIN PG ; ARMSTRONG-FISHER SS ; LIU W ; FINNING KM ; MADDOCKS D ; UR-BANIAK SJ.** Evidence of genetic diversity underlying Rh D negative, weak D (Du) and partial D phenotypes as determined by multiplex PCR analysis of the RHD gene. *Blood*, 1997, vol. 89, 2568 [0235]
- **AVENT ND ; LIU W ; JONES JW ; SCOTT ML ; VOAK D ; PISACKA M ; WATT J ; FLETCHER A.** Molecular analysis of Rh transcripts and polypeptides from individuals expressing the DVI variant phenotype: an RHD gene deletion event does not generate all DVIccEe phenotypes. *Blood*, 1997, vol. 89, 1779-86 [0235]
- **BLUMENFELD OO ; HUANG CH.** Molecular genetics of the glycophorin gene family, the antigens for MNSs blood groups: multiple gene rearrangements and modulation of splice site usage result in extensive diversification. *Hum Mutat*, 1995, vol. 6, 199 [0235]
- **BLUNT T ; DANIELS G ; CARRITT B.** Serotype switching in a partially deleted RHD gene. *Vox Sang*, 1994, vol. 67, 397 [0235]
- **BOWMAN J.** The management of hemolytic disease in the fetus and newborn. *Semin Perinatol*, 1997, vol. 21, 39 [0235]
- **CARRITT B ; KEMP TJ ; POULTER M.** Evolution of the human RH (rhesus) blood group genes: a 50 year old prediction (partially) fulfilled. *Hum Mol Genet*, 1997, vol. 6, 843 [0235]
- **CARRITT B ; STEERS FJ ; AVENT ND.** Prenatal determination of fetal RhD type. *Lancet*, 1994, vol. 344, 205-6 [0235]
- **CHERIF-ZAHAR B ; MATTEI MG ; LE VAN KIM C ; BAILLY P ; CARTRON JP ; COLIN Y.** Localization of the human Rh blood group gene structure to chromosome region 1p34.3-1p36.1 by in situ hybridization. *Hum Genet*, 1991, vol. 86, 398 [0235]
- **CHERIF-ZAHAR B ; BONY V ; STEFFENSEN R ; GANE P ; RAYNAL V ; GOOSENS D ; LAURSEN JS ; VARMING K ; JERSILD C ; CARTRON JP.** Shift from Rh-positive to Rh-negative phenotype caused by a somatic mutation within the RHD gene in a patient with chronic myelocytic leukaemia. *Br J Haematol*, 1998, vol. 102, 1263-70 [0235]
- **COLIN Y ; CHERIF-ZAHAR B ; LE VAN KIM C ; RAYNAL V ; VAN HUFFEL V ; CARTRON J.-P.** Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by southern analysis. *Blood*, 1991, vol. 78, 2747-52 [0235]
- **COSSU G ; ANGIUS A ; GELFI C ; RIGHETTI PG.** Rh D/d genotyping by quantitative polymerase chain reaction and capillary zone electrophoresis. *Electrophoresis*, 1996, vol. 17, 1911 [0235]
- **DANIELS GL ; FAAS BH ; GREEN CA ; SMART E ; MAASKANT-VAN WIJK PA ; AVENT ND ; ZONDERVAN HA ; VON DEM BORNE AE ; VAN DER SCHOOT CE.** The VS and V blood group polymorphisms in Africans: a serologic and molecular analysis. *Transfusion*, 1998, vol. 38, 951-8 [0235]

- **DÖSCHER A ; SCHUNTER F ; MÜLLER TH.** Quantitative PCR to assess RHD zygosity. *Infusionsther Transfusionsmed*, 1999, vol. 26 (1), 31 [0235]
- **FAAS BHW ; BECKER EAM ; WILDOER P ; LIGHART PC ; OVERBEEKE MAM ; ZONDERVAN HA ; VON DEM BORNE AEGK ; VAN DER SCHOOT CE.** Molecular background of VS and weak C expression in blacks. *Transfusion*, 1997, vol. 37, 38 [0235]
- **FAAS BHW ; BECKERS EAM ; SIMSEK S ; OVERBEEKE MAM ; PEPPER R ; VAN RHENEN DJ ; VON DEM BORNE AEGK ; VAN DER SCHOOT CE.** Involvement of Ser103 of the Rh polypeptides in G epitope formation. *Transfusion*, 1996, vol. 36, 506 [0235]
- **FILBEY D ; HANSON U ; WESSTROM G.** The prevalence of red cell antibodies in pregnancy correlated to the outcome of the newborn: a 12 year study in central Sweden. *Acta Obstet Gynecol Scand*, 1995, vol. 74, 687 [0235]
- **FLEGEL WA ; KHULL SR ; WAGNER FF.** Primary anti-D immunization by weak D type 2 RBCs. *Transfusion*, 2000, vol. 40, 428-434 [0235]
- **FLEGEL WA ; WAGNER FF ; MÜLLER TH ; GAS-SNER C.** Rh phenotype prediction by DNA typing and its application to practice. *Transfus Med*, 1998, vol. 8, 281 [0235]
- **FUJIWARA H ; OKUDA H ; OMI T ; IWAMOTO S ; TANAKA Y ; TAKAHASHI J ; TANI Y ; MINAKAMI H ; ARAKI S ; SATO I.** The STR polymorphisms in intron 8 may provide information about the molecular evolution of RH haplotypes. *Hum Genet*, 1999, vol. 104, 301 [0235]
- **GASSNER C ; SCHMARDA A ; KILGA-NOGLER S ; JENNY-FELDKIRCHER B ; RAINER E ; MÜLLER TH ; WAGNER FF ; FLEGEL WA ; SCHÖNITZER D.** RhesusD/CE typing by polymerase chain reaction using sequence-specific primers. *Transfusion*, 1997, vol. 37, 1020 [0235]
- **HUANG CH.** Human DVI category erythrocytes: correlation of the phenotype with a novel hybrid RhD-CE-D gene but not an internally deleted RhD gene. *Blood*, 1997, vol. 89, 1834-9 [0235]
- **HUANG CH ; REID ME ; CHEN Y ; COGLAN G ; OKUBO Y.** Molecular definition of red cell Rh haplotypes by tightly linked SphI RFLPs. *Am J Hum Genet*, vol. 58, 133 [0235]
- **HUANG CH.** Alteration of RH gene structure and expression in human dCCee and DCW- red blood cells: phenotypic homozygosity versus genotypic heterozygosity. *Blood*, 1996, vol. 88, 2326 [0235]
- **HYLAND CA ; WOLTER LC ; SAUL A.** Three unrelated Rh D gene polymorphisms identified among blood donors with Rhesus CCee (r'r) phenotypes. *Blood*, 1994, vol. 84, 321-4 [0235]
- **KEMP TJ ; POULTER M ; CARRITT B.** Microsatellite variation within the human RHCE gene. *Vox Sang*, 1999, vol. 77, 159 [0235]
- **LO Y-MD ; BOWELL PJ ; SELINGER M ; MAC-KENZIE IZ ; CHAMBERLAIN P ; GILLMER MDG ; LITTLEWOOD TJ ; FLEMING KA ; WAINSCOT JS.** Prenatal determination of fetal RhD status by analysis of peripheral blood of rhesus negative mothers. *Lancet*, 1993, vol. 341, 1147 [0235]
- **MAASKANT-VAN WIJK PA ; HEMKER MB ; DOUGLAS-BERGER L et al.** Ethnic variability of the Rhesus system. *Transfusion*, 1999, vol. 39 (10S), 103S-4S [0235]
- **MAASKANT-VAN WIJK PA ; FAAS BHW ; DE RUIJTER JAM ; OVERBEEKE MAM ; VON DEM BORNE AEGK ; VAN RHENEN DJ ; VAN DER SCHOOT CE.** Genotyping of RHD by multiplex polymerase chain reaction analysis of six RHD-specific exons. *Transfusion*, 1998, vol. 38, 1015-21 [0235]
- **MAASKANT-VAN WIJK PA ; BECKERS EAM ; VAN RHENEN DJ ; MOURO I ; COLIN Y ; CARTRON J-P ; FAAS BHW ; VAN DER SCHOOT CE ; APOIL PA ; BLANCHER A et al.** Evidence that the RHD(VI) deletion genotype does not exist. *Blood*, 1997, vol. 90, 1709-11 [0235]
- **MACGEOCH C ; MITCHELL CJ ; CARRITT B ; AVENT ND ; RIDGWELL K ; TANNER MJ ; SPURR NK.** Assignment of the chromosomal locus of the human 30-kDal Rh (rhesus) blood group-antigen-related protein (Rh30A) to chromosome region 1p36.13----p34. *Cytogenet Cell Genet*, 1992, vol. 59, 261 [0235]
- **MOURANT AE ; KOPEC AC ; DOMANIEWSKA-SOBCZAK K.** The distribution of the human blood groups and other polymorphisms. Oxford University Press, 1976 [0235]
- **OKUDA H ; SUGANUMA H ; TSUDO N ; OMI T ; IWAMOTO S ; KAJII E.** Sequence analysis of the spacer region between the RHD and RHCE genes. *Biochem Biophys Res Commun*, 1999, vol. 263, 378 [0235]
- **OKUDA H ; KAWANO M ; IWAMOTO S ; TANAKA M ; SENO T ; OKUBO Y ; KAJII E.** The RHD gene is highly detectable in RhD-negative Japanese donors. *J Clin Invest*, 1997, vol. 100, 373 [0235]
- **ONDA M ; FUKUDA M.** Detailed physical mapping of the genes encoding glycoporphins A, B and E, as revealed by P1 plasmids containing human genomic DNA. *Gene*, 1995, vol. 159, 225 [0235]
- **REBOUL J ; GARDINER K ; MONNERON D ; UZE G ; LUTFALLA G.** Comparative genomic analysis of the interferon/interleukin-10 receptor gene cluster. *Genome Res*, 1999, vol. 9, 242 [0235]
- **SALVIGNOL I ; CALVAS P ; SOCHA WW ; COLIN Y ; LE VAN KIM C ; BAILLY P ; RUFFIE J ; CARTRON JP ; BLANCHER A.** Structural analysis of the RH-like blood group gene products in nonhuman primates. *Immunogenetics*, 1995, vol. 41, 271 [0235]

- **SINGLETON BK ; GREEN CA ; AVENT ND ; MARTIN PG ; SMART E ; DAKA A ; NARTER-OLAGA EG ; HAWTHORNE LM ; DANIELS G.** The presence of an RHD pseudogene containing a 37 base pair duplication and a nonsense mutation in africans with the Rh D- negative blood group phenotype. *Blood*, 2000, vol. 95, 12 [0235]
- **URBANIAK SJ ; ROBERTSON AE.** A successful program of immunizing Rh-negative male volunteers for anti-D production using frozen/thawed blood. *Transfusion*, 1981, vol. 21, 64 [0235]
- **WAGNER FF ; GASSNER C ; MÜLLER TH ; SCHÖNITZER D ; SCHUNTER F ; FLEGEL WA.** Three molecular structures cause Rhesus D category VI phenotypes with distinct immunohematologic features. *Blood*, 1998, vol. 91, 2157 [0235]
- **WAGNER FF ; KASULKE D ; KEROWGAN M ; FLEGEL WA.** Frequencies of the blood groups ABO, Rhesus, D category VI, Kell, and of clinically relevant high-frequency antigens in South-Western Germany. *Infusionsther Transfusionsmed*, 1995, vol. 22, 285 [0235]
- **WAGNER FF ; GASSNER C ; MÜLLER TH ; SCHÖNITZER D ; SCHUNTER F ; FLEGEL WA.** Molecular basis of weak D phenotypes. *Blood*, 1999, vol. 93, 385 [0235]
- **WAGNER FF ; FROHMAJER A ; LADEWIG B ; EICHER NI ; LONICER CB ; MÜLLER TH ; SIEGEL MH ; FLEGEL WA.** Weak D alleles express distinct phenotypes. *Blood*, 2000, vol. 95, 2699-2708 [0235]
- **WAGNER FF ; FLEGEL WA.** RHD gene deletion occurred in the Rhesus box. *Blood*, 2000 [0235]
- Paul-Ehrlich-Institut. Richtlinien zur Blutgruppenbestimmung und Bluttransfusion (Hämotherapie). *Wissenschaftlicher Beirat der Bundesärztekammer*, 1996 [0235]
- **LEGLER TJ ; KOHLER M ; MAYR WR ; PANZER S ; OHTO H ; FISCHER GF.** Genotyping of the human platelet antigen systems 1 through 5 by multiplex polymerase chain reaction and ligation-based typing. *Transfusion*, May 1996, vol. 36 (5), 426-31 [0235]

专利名称(译)	RHD阴性基因座的分子结构		
公开(公告)号	<a href="#">EP1780217A1</a>	公开(公告)日	2007-05-02
申请号	EP2006026286	申请日	2000-10-31
[标]申请(专利权)人(译)	DRK BLUTSPENDEDIENST巴符州HESSEN GEMEINNUTZIGE		
申请(专利权)人(译)	DRK-BLUTSPENDEDIENST巴登 - 符腾堡州 , HESSENGEMEINNÜTZIGEGMBH		
当前申请(专利权)人(译)	DRK-BLUTSPENDEDIENST巴登 - 符腾堡州 , HESSENGEMEINNÜTZIGEGMBH		
[标]发明人	FLEGEL WILLY A WAGNER FRANZ F		
发明人	FLEGEL, WILLY A. WAGNER, FRANZ F.		
IPC分类号	C07K14/47 C12N15/12 C12Q1/68 G01N33/53 A61K39/395 A61P15/00 C07K14/705 C07K16/18 C12N5/10 C12N7/00 C12N15/09 C12P21/02 C12P21/08 C12Q1/02 G01N33/566		
CPC分类号	A61P15/00 C07K14/705 C12Q1/6883 C12Q2600/156 C12Q2600/16 C12Q2600/172 C12N15/52		
代理机构(译)	法思博事务所		
优先权	1999121686 1999-11-02 EP 2000111696 2000-05-31 EP		
其他公开文献	EP1780217B1		
外部链接	<a href="#">Espacenet</a>		

摘要(译)

本发明涉及代表Rhesus基因座的核酸分子结构，其包含RHD，SMP1和RHCE基因和/或恒河猴盒，优选混合恒河猴盒，上游恒河猴盒和/或下游恒河猴盒。。此外，本发明涉及特异性检测常见RHD阴性单倍型的方法。本发明还涉及D阴性个体中RHD阳性降低型的检测。已经鉴定了RHD基因中的各种突变，其允许开发诊断工具。本发明还涉及寡核苷酸，其特异性杂交杂交盒，优选断裂点或断裂点区域或上游和下游恒河猴盒。另外，本发明涉及包含或使用上述本发明化合物的试剂盒。

```

SEQUENCE LISTING
<110> DRK Blutspendedienst Baden-Wuerttemberg gGmbH
<120> Molecular structure of RHD negative
<130> D2600 EP/1
<141> 2000-10-31
<160> 66
<170> PatentIn Ver. 2.1
<210> 1
<211> 100
<212> DNA
<213> Homo sapiens
<400> 1
gctctgggga ctgagccggg tggatggtae tctgcacac ggggtgtctgg aggctgtggc 60
cgtctctgttc tcttggctaa aactcggggga ctgagccggg 100
<210> 2
<211> 26
<212> DNA
<213> Homo sapiens
<400> 2
actgcacgac ggggctggac tgaagt 26
<210> 3
<211> 45
<212> DNA
<213> Homo sapiens
<400> 3
agcggaaaa aatgtctgga ttctagagg gcttgatag etcag 45
<210> 4
<211> 26
<212> DNA
<213> Homo sapiens
<400> 4
gcttccattg ctactgtgtt actagt 26
<210> 5
<211> 44
<212> DNA
<213> Homo sapiens
<400> 5
agctttttac aggetgttg attatcatag atcagctgt tatt 44
<210> 6
<211> 26
<212> DNA

```