



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>6</sup> : <b>C07H 21/00</b></p>	<p><b>A1</b></p>	<p>(11) International Publication Number: <b>WO 99/14228</b></p> <p>(43) International Publication Date: 25 March 1999 (25.03.99)</p>
<p>(21) International Application Number: PCT/US98/19325</p> <p>(22) International Filing Date: 16 September 1998 (16.09.98)</p> <p>(30) Priority Data: 60/059,304 17 September 1997 (17.09.97) US 60/066,172 18 November 1997 (18.11.97) US</p> <p>(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications US 60/059,304 (CIP) Filed on 17 September 1997 (17.09.97) US 60/066,172 (CIP) Filed on 18 November 1997 (18.11.97)</p> <p>(71) Applicant (for all designated States except US): AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): LIPSHUTZ, Robert, J. [US/US]; 970 Palo Alto Avenue, Palo Alto, CA 94301 (US). CHEE, Mark [AU/US]; 3199 Waverley Street, Palo Alto, CA 94306 (US). FAN, Jian-Bing [CN/US]; Apartment 20, 275 Ventura Avenue, Palo Alto, CA 94306 (US). BERNO,</p>	<p>Anthony [CA/US]; 570 South 12th Street, San Jose, CA 95112 (US).</p> <p>(74) Agents: LIEBESCHUETZ, Joe et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US).</p> <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	
<p>(54) Title: GENETIC COMPOSITIONS AND METHODS</p>		
<p>(57) Abstract</p> <p>The invention provides nucleic acid segments of the human genome including polymorphic sites. Allele-specific primers and probes hybridizing to regions flanking these sites are also provided. The nucleic acids, primers and probes are used in applications such as forensics, paternity testing, medicine and genetic analysis.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<b>AL</b>	Albania	<b>ES</b>	Spain	<b>LS</b>	Lesotho	<b>SI</b>	Slovenia
<b>AM</b>	Armenia	<b>FI</b>	Finland	<b>LT</b>	Lithuania	<b>SK</b>	Slovakia
<b>AT</b>	Austria	<b>FR</b>	France	<b>LU</b>	Luxembourg	<b>SN</b>	Senegal
<b>AU</b>	Australia	<b>GA</b>	Gabon	<b>LV</b>	Latvia	<b>SZ</b>	Swaziland
<b>AZ</b>	Azerbaijan	<b>GB</b>	United Kingdom	<b>MC</b>	Monaco	<b>TD</b>	Chad
<b>BA</b>	Bosnia and Herzegovina	<b>GE</b>	Georgia	<b>MD</b>	Republic of Moldova	<b>TG</b>	Togo
<b>BB</b>	Barbados	<b>GH</b>	Ghana	<b>MG</b>	Madagascar	<b>TJ</b>	Tajikistan
<b>BE</b>	Belgium	<b>GN</b>	Guinea	<b>MK</b>	The former Yugoslav Republic of Macedonia	<b>TM</b>	Turkmenistan
<b>BF</b>	Burkina Faso	<b>GR</b>	Greece	<b>ML</b>	Mali	<b>TR</b>	Turkey
<b>BG</b>	Bulgaria	<b>HU</b>	Hungary	<b>MN</b>	Mongolia	<b>TT</b>	Trinidad and Tobago
<b>BJ</b>	Benin	<b>IE</b>	Ireland	<b>MR</b>	Mauritania	<b>UA</b>	Ukraine
<b>BR</b>	Brazil	<b>IL</b>	Israel	<b>MW</b>	Malawi	<b>UG</b>	Uganda
<b>BY</b>	Belarus	<b>IS</b>	Iceland	<b>MX</b>	Mexico	<b>US</b>	United States of America
<b>CA</b>	Canada	<b>IT</b>	Italy	<b>NE</b>	Niger	<b>UZ</b>	Uzbekistan
<b>CF</b>	Central African Republic	<b>JP</b>	Japan	<b>NL</b>	Netherlands	<b>VN</b>	Viet Nam
<b>CG</b>	Congo	<b>KE</b>	Kenya	<b>NO</b>	Norway	<b>YU</b>	Yugoslavia
<b>CH</b>	Switzerland	<b>KG</b>	Kyrgyzstan	<b>NZ</b>	New Zealand	<b>ZW</b>	Zimbabwe
<b>CI</b>	Côte d'Ivoire	<b>KP</b>	Democratic People's Republic of Korea	<b>PL</b>	Poland		
<b>CM</b>	Cameroon	<b>KR</b>	Republic of Korea	<b>PT</b>	Portugal		
<b>CN</b>	China	<b>KZ</b>	Kazakstan	<b>RO</b>	Romania		
<b>CU</b>	Cuba	<b>LC</b>	Saint Lucia	<b>RU</b>	Russian Federation		
<b>CZ</b>	Czech Republic	<b>LI</b>	Liechtenstein	<b>SD</b>	Sudan		
<b>DE</b>	Germany	<b>LK</b>	Sri Lanka	<b>SE</b>	Sweden		
<b>DK</b>	Denmark	<b>LR</b>	Liberia	<b>SG</b>	Singapore		
<b>EE</b>	Estonia						

## GENETIC COMPOSITIONS AND METHODS

5

10

## BACKGROUND OF THE INVENTION

15

20

25

30

35

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution generating variant forms of progenitor sequences (Gusella, *Ann. Rev. Biochem.* 55, 831-854 (1986)). The variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers a lethal disadvantage and is not transmitted to subsequent generations of the organism. In other instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. In many instances, both progenitor and variant form(s) survive and co-exist in a species population. The coexistence of multiple forms of a sequence gives rise to polymorphisms.

Several different types of polymorphism have been reported. A restriction fragment length polymorphism (RFLP) means a variation in DNA sequence that alters the length of a restriction fragment as described in Botstein et al., *Am. J. Hum. Genet.* 32, 314-331 (1980). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have been widely used in human and animal genetic analyses (see WO 90/13668; W090/11369; Donis-Keller, *Cell* 51, 319-337 (1987); Lander et al., *Genetics* 121, 85-99 (1989)). When a heritable trait can be linked to a particular RFLP, the

presence of the RFLP in an individual can be used to predict the likelihood that the animal will also exhibit the trait.

Other polymorphisms take the form of short tandem repeats (STRs) that include tandem di-, tri- and tetra-  
5 nucleotide repeated motifs. These tandem repeats are also referred to as variable number tandem repeat (VNTR) polymorphisms. VNTRs have been used in identity and paternity analysis (US 5,075,217; Armour et al., *FEBS Lett.* 307, 113-115 (1992); Horn et al., WO 91/14003; Jeffreys, EP 370,719), and  
10 in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPs, STRs and VNTRs. Some single nucleotide polymorphisms occur in protein-  
15 coding sequences, in which case, one of the polymorphic forms may give rise to the expression of a defective or other variant protein and, potentially, a genetic disease. Examples of genes, in which polymorphisms within coding sequences give rise to genetic disease include  $\beta$ -globin (sickle cell anemia) and CFTR (cystic fibrosis). Other single nucleotide  
20 polymorphisms occur in noncoding regions. Some of these polymorphisms may also result in defective protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects.

25 Single nucleotide polymorphisms can be used in the same manner as RFLPs, and VNTRs but offer several advantages. Single nucleotide polymorphisms occur with greater frequency and are spaced more uniformly throughout the genome than other forms of polymorphism. The greater frequency and uniformity  
30 of single nucleotide polymorphisms means that there is a greater probability that such a polymorphism will be found in close proximity to a genetic locus of interest than would be the case for other polymorphisms. Also, the different forms of characterized single nucleotide polymorphisms are often  
35 easier to distinguish than other types of polymorphism (e.g., by use of assays employing allele-specific hybridization probes or primers).

Despite the increased amount of nucleotide sequence data being generated in recent years, only a minute proportion of the total repository of polymorphisms in humans and other organisms has so far been identified. The paucity of polymorphisms hitherto identified is due to the large amount of work required for their detection by conventional methods. For example, a conventional approach to identifying polymorphisms might be to sequence the same stretch of oligonucleotides in a population of individuals by dideoxy sequencing. In this type of approach, the amount of work increases in proportion to both the length of sequence and the number of individuals in a population and becomes impractical for large stretches of DNA or large numbers of persons.

#### SUMMARY OF THE CLAIMED INVENTION

The invention provides nucleic acid segments of between 10 and 100 bases from a fragment shown in Table 1, column 1 including a polymorphic site. Complements of these segments are also included. The segments can be DNA or RNA, and can be double- or single-stranded. Some segments are 10-20 or 10-50 bases long. Preferred segments include a diallelic polymorphic site. The base occupying the polymorphic site in the segments can be the reference (Table 1, column 3) or an alternative base (Table 1, column 5).

The invention further provides allele-specific oligonucleotides that hybridizes to a segment of a fragment shown in Table 1, column 8 or its complement. These oligonucleotides can be probes or primers. Also provided are isolated nucleic acids comprising a sequence of Table 1, column 8, or the complement thereto, in which the polymorphic site within the sequence is occupied by a base other than the reference base shown in Table 1, column 3.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the polymorphic sites shown in Table 1. Optionally, a set of bases occupying a set of the polymorphic sites shown in Table 1 is determined. This type of analysis can be performed on a plurality of individuals who



Hybridization probes are oligonucleotides capable of binding in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen et al., *Science* 254, 1497-1500 (1991).

5           The term primer refers to a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (*i.e.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or  
10 RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form  
15 sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair  
20 means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3', downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

Linkage describes the tendency of genes, alleles, loci  
25 or genetic markers to be inherited together as a result of their location on the same chromosome, and can be measured by percent recombination between the two genes, alleles, loci or genetic markers.

Polymorphism refers to the occurrence of two or more  
30 genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected  
35 population. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats,

trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms.

A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations).

A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific probe hybridizations.

An isolated nucleic acid means an object species invention that is the predominant species present (*i.e.*, on a molar basis it is more abundant than any other individual species in the composition). Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods).

Linkage disequilibrium or allelic association means the preferential association of a particular allele or genetic marker with a specific allele, or genetic marker at a nearby chromosomal location more frequently than expected by chance for any particular allele frequency in the population. For example, if locus X has alleles a and b, which occur equally frequently, and linked locus Y has alleles c and d, which occur equally frequently, one would expect the combination ac to occur with a frequency of 0.25. If ac occurs more frequently, then alleles a and c are in linkage disequilibrium. Linkage disequilibrium may result from natural selection of certain combination of alleles or because an allele has been introduced into a population too recently to have reached equilibrium with linked alleles.

A marker in linkage disequilibrium can be particularly useful in detecting susceptibility to disease (or other phenotype) notwithstanding that the marker does not cause the disease. For example, a marker (X) that is not itself a causative element of a disease, but which is in linkage disequilibrium with a gene (including regulatory sequences) (Y) that is a causative element of a phenotype, can be used detected to indicate susceptibility to the disease in circumstances in which the gene Y may not have been identified or may not be readily detectable.

The present invention includes the use of any of the polymorphic forms shown in Table 1 as a means to determine susceptibility to a phenotype resulting from an allele or marker in linkage disequilibrium with such polymorphic forms.

## DESCRIPTION

### 30 I. Novel Polymorphisms of the Invention

The novel polymorphisms of the invention are listed in Table 1. The first column of the Table lists the names assigned to the fragments in which the polymorphisms occur. The fragments are all human genomic fragments. SGC, TIGR and WI respectively stand for Stanford Genome Center, The Institute for Genome Research and the Whitehead Institute.

The sequence of one allelic form of each of the fragments (arbitrarily referred to as the prototypical or reference form) has been previously published. These sequences are listed at <http://www-genome.wi.mit.edu/> (all STS's (sequence tag sites)); <http://shgc.stanford.edu> (Stanford STS's); and <http://ww.tigr.org/> (TIGR STS's). The Web sites also list primers for amplification of the fragments, and the genomic location of fragments. Some fragments are expressed sequence tags, and some are random genomic fragments. All information in the websites concerning the fragments listed in Table 1 is incorporated by reference in its entirety for all purposes.

The second column lists the position in the fragment in which a polymorphic site has been found. Positions are numbered consecutively with the first base of the fragment sequence as listed in one of the above databases being assigned the number one. The third column lists the base occupying the polymorphic site in the sequence in the data base. This base is arbitrarily designated the reference or prototypical form but is not necessarily the most frequently occurring form. The fifth column in the table lists the alternative base(s) at the polymorphic site. The eighth column of the Table lists about 15 bases of sequence on either side of the polymorphic site in each fragment. The indicated sequences can be either DNA or RNA. In the latter, the T's shown in the Table are replaced by U's. The base occupying the polymorphic site is indicated in EUPAC-IUB ambiguity code.

The fourth and sixth columns of the table show the frequency with which reference and alternative alleles occur at a polymorphic site. The seventh column in the table indicates the population frequency of heterozygotes of the polymorphic site.

Fragment	Position	"Ref Allele"	"Freq (P)"	"Alt Allele"	"Freq (Q)"	"H"	"Sequence Tag"
19201	179	T	0.25	C	0.75	0.38	GGTGCACCGAAAGGAYTGGGGGATAAAATTC
19212	46	T	0.94	A	0.06	0.12	GAGACTAGAGTGACAWGTTTTCAGAACCCAAA
19222	179	C	0.94	T	0.06	0.12	AGGGACTCTGCGGAAYTTTCACACCTCTTC
19224	112	C	0.94	T	0.06	0.12	ACAGAGGAGATAATCYCAGGATGCCTGTGAA

	19235	173	A	0.81	G	0.19	0.30	G TTCACAATGGTGGARGCTTCATGTAATATG
	19236	54	G	0.63	A	0.38	0.47	TGGAAGGGGGAAAAGRATGGAGACCTGCTC
	19269	85	A	0.56	T	0.44	0.49	ATTTTGGAGTGTGTCWTTGGGTAGCAATGTG
5	19307	196	T	0.94	C	0.06	0.12	CCTTTAAAGAGACCCYTGAAATGGGCCATG
	19348	98	G	0.56	A	0.44	0.49	GACTGTTGGTCATGGRGTGACGTCCTTCTCC
	19348	103	C	0.50	T	0.50	0.50	TTGGTCATGGCGGTGYGTCCTTCTCCAGGCT
	19359	39	T	0.56	C	0.44	0.49	TGAATACTTTGTTTTYCATGTTCAAAAAAAG
	19415	161	A	0.56	G	0.44	0.49	CCTTAGCTGATCTCARAAGTCCACCTCATGA
10	19591	45	T	0.69	A	0.31	0.43	ATCACATATACCTGAWATAAGGTAAC TCCAA
	19591	156	C	0.38	A	0.63	0.47	GTGGGGAGCTCTTCCMCTACCACTCCCACC
	19599	230	C	0.56	G	0.44	0.49	TTAAAAGTAAAGGGCSTTCCAAGAGTAACAC
	19635	98	A	0.63	T	0.38	0.47	AAAAATACAGTATTAWATCTTATTGTGTAGC
	19641	46	A	0.88	G	0.13	0.22	TTGTGATAAGCACTARTATTATAGTCTCATG
15	18012	112	C	0.20	T	0.80	0.32	GCCACTTTTGCCCCTYGTGAAGTGTTTCTG
	18014	40	A	0.90	G	0.10	0.18	TTGAATAGCTACAGARGAATGAAAGTGCACC
	18036	27	T	0.43	C	0.57	0.49	GAGTCAAGTACCAAGYAAACTTCTAGAAATA
	18036	97	T	0.93	A	0.07	0.12	TTTAATTCTTTCATAWCTGACAGGTCAAGTA
	18046	72	C	0.80	T	0.20	0.32	TTTCAGGCCAATGTGYTGTGGGTCTGAGAT
20	18052	50	T	0.40	C	0.60	0.48	CTTTCATGTACGAATYTGTTACACATCTTA
	18052	67	A	0.40	G	0.60	0.48	TGGTTACACATCTTARACAGCAGAGCTGCCT
	18054	46	G	0.13	A	0.87	0.23	GAGTGGGGGAGTAAARTGGAAGCAGGGTGAC
	18063	105	G	0.77	A	0.23	0.36	TAAACTTAAAATTTGRTCTTTAACAATATA
	18064	54	G	0.87	A	0.13	0.23	TAAGCTGTATTT CAGRGAATGTCACAATCAT
25	18078	86	A	0.97	T	0.03	0.06	TTTTTTTCAGCATCAWGTCCACTAGCCAAGT
	18080	41	T	0.47	C	0.53	0.50	ATCAAAGTCTCTCTYTTGTAATTA AATCT
	18080	65	G	0.53	A	0.47	0.50	TAAAATCTACTATGCRTGTTTGACTTTTATC
	18080	80	C	0.73	T	0.27	0.39	CGTGTTTGACTTTTAYTCTTATGTAAATTGA
	18086	63	G	0.10	A	0.90	0.18	CAGAAAGCATACTTCRTGGCTTTGTTACAG
30	18091	90	T	0.97	C	0.03	0.06	CTCTAGAAGTTTGACYGGGCCTTTT TATAC
	18115	70	C	0.87	T	0.13	0.23	CCTTTGGTATTCCCTYCTTTGGTATGAAAGA
	18115	71	C	0.87	T	0.13	0.23	CCTTTGGTATTCCCTYTTTGGTATGAAAGAC
	18119	38	T	0.83	C	0.17	0.28	GTGGTATTACAGAGGYTTGAAAATGGATTG
	18136	78	A	0.97	G	0.03	0.06	TCTTTAGGTAATTTGRTAAGAACAATAAAAG
35	18142	66	T	0.97	G	0.03	0.06	AAAATAATCTATATAKCCCAATAAACTCACA
	18169	115	A	0.70	G	0.30	0.42	ATCTTTCCGGAAGCTRTGGAGCACAAGCAGA
	18175	27	A	0.20	G	0.80	0.32	ACGCTGCCCTTTTARTAGAACATTATCAAA
	18178	68	T	0.83	C	0.17	0.28	AGGTTAGTCTGGGGYCGGCGGGATGGACAC
	18181	100	A	0.60	C	0.40	0.48	ACACTCCCTTCAGATMCAAAAAGCTTAACAAA
40	18190	26	G	0.90	A	0.10	0.18	CGACACAGCGGACACRTCATAAGTGGAACAA
	18190	62	G	0.67	A	0.33	0.44	TGAAGCTAATCATGGRGCAAGCTCCCTGGAG
	18215	78	G	0.75	A	0.25	0.38	CAGAGTTCCTGCCCTRGTGTGCGGGGGGAGA
	18232	60	T	0.91	A	0.09	0.17	TTGTGATACACTTAAWGAACCCCTGAAAACC
	18243	36	T	0.94	C	0.06	0.12	CAGCAGCAGAATGCAYTTTGCAGAAACACAC
45	18244	35	G	0.59	T	0.41	0.48	TAAGCCAGCATGGGKGGGGAGGTGATTATG
	18245	115	G	0.97	A	0.03	0.06	GGACAGAGAAACATGRCTGGGGAGTAGGCTC
	18247	19	G	0.09	A	0.91	0.17	CACACCACAAACGCARGTTAGTGAGCTGCTA
	18261	26	G	0.78	A	0.22	0.34	GATTGCTTTATTTAARTGAAAAGCGTGATAG
	18266	97	C	0.16	T	0.84	0.26	ATGGACTATCTTCAAYTGACAAATGATGCA
50	18266	119	C	0.75	T	0.25	0.38	AAATGATGCATGAATYACATTTGAGACCCGC
	18266	124	T	0.16	C	0.84	0.26	ATGCATGAATCCACAYTGAGACCCGCAACTC
	18268	88	C	0.75	T	0.25	0.38	TACTTCCCCCATAGAYCCTGACAATGTGCTG

	18299	48	C	0.56	T	0.44	0.49	TGTCTAAGATCATTAYTTGGTTTGCCAATTT
	18299	52	G	0.75	A	0.25	0.38	TAAGATCATTAACRTRTTTGCCAATTTTTTTT
	18299	67	T	0.56	G	0.44	0.49	GGTTTGCCAATTTTTKATCTATTTGGGTCTG
	18299	77	G	0.78	A	0.22	0.34	TTTTTTTATCTATTTTRGTCTGAGAATTCCAC
5	18299	101	A	0.38	G	0.63	0.47	AATTCCACAATTTTGRGAATTCTTTTGCCAA
	18299	107	C	0.78	A	0.22	0.34	ACAATTTTGAAGAATMTTTTGCCAATTATTG
	18307	76	G	0.94	A	0.06	0.12	AACTCAGTTTCCGCTRTGCTATGTAAAGCAT
	18324	72	C	0.97	T	0.03	0.06	GGGGTACTGATTTATYTAGATCCAAATAAAG
10	18327	104	G	0.41	A	0.59	0.48	TTTCGTTAGGCTAGTRGCTGAGCCATTGTAT
	18330	49	G	0.47	A	0.53	0.50	GAAATCAGGGATAAAGRCTGAGGAACAAGAGG
	18330	66	A	0.50	G	0.50	0.50	CTGAGGAACAAGAGGRTATGTAGGCAGTGAG
	18350	48	T	0.97	C	0.03	0.06	AAAGAATGTTTTTTCAGYTAATCTATGAAAAG
	18357	89	C	0.66	G	0.34	0.45	CAGCCCTTAGCATCASTCATCTTCAGTCTTT
15	18369	58	G	0.84	A	0.16	0.26	AATCTGTCACACAATRAAATGGATAAAGCCT
	18387	57	A	0.66	G	0.34	0.45	TTTGGTGACCCCATARTTTGTGGTCACATGC
	18387	84	A	0.94	C	0.06	0.12	CATGCTTTAGCCATAMCATGGTAACATTGAC
	18395	77	G	0.41	C	0.59	0.48	AAATTGATTATTCAASTGTGCATTGGTTTAT
	18396	21	C	0.91	A	0.09	0.17	TGGTATTCTCTCATCMTTCTTTTCGCTCTT
20	18398	62	G	0.84	T	0.16	0.26	AAACAACCTCAAGGGTKGATAACATTGCCAGT
	18399	28	A	0.16	T	0.84	0.26	CTGTATCCAGTGGCAWTTTTGGCTGCTGGTT
	18399	99	C	0.34	T	0.66	0.45	ACCTCAAGGGACACCYCCACCCGACACTGTT
	18409	20	C	0.44	A	0.56	0.49	TGGGAAAGAGGAAATMTTTTTCTTACTAGAG
	18420	38	C	0.09	T	0.91	0.17	GGGAAAATGGGAAGAYAGAGTGAAATTAAG
25	18420	108	T	0.56	C	0.44	0.49	CTCAAAAAAATCAAYGCTTATAGCAATGCT
	18425	81	A	0.06	C	0.94	0.12	GTCCTAGACAGATTTCMTGCACACAACAACAG
	18425	101	T	0.84	C	0.16	0.26	ACACAACAACAGGAGYGGGGTGCACACGGGC
	18442	62	C	0.78	T	0.22	0.34	CAAAATAAGTTTCTGYTTGGCTGATCTGGGT
	18452	38	G	0.97	A	0.03	0.06	GGAGATCGGCTAAAAARAAGCATAGTTATTA
30	18457	120	T	0.97	C	0.03	0.06	CACATTGGGGCCACAYAAATAGGCTAAAAGG
	18462	39	A	0.70	G	0.30	0.42	ACAATGGCAGAGGTGRTAGAAACCATCTCAA
	18489	102	A	0.93	C	0.07	0.12	TGCAAGGATTCAAACMGGTTATGGCAATAGA
	18491	109	G	0.83	A	0.17	0.28	TAAATCCCAGAATGARGGATTACAAGAAAAT
	18520	75	G	0.90	A	0.10	0.18	TTGTTCTTTTACTACRCCGGAGTGGTAAATA
35	18533	91	T	0.80	C	0.20	0.32	TCATTTTTTTCATCCTAYTTACTGAAGCCATTT
	18535	107	G	0.93	A	0.07	0.12	CTTCACGGGAGAGCTRTTGTTTAAAGCAGTG
	18562	29	G	0.93	A	0.07	0.12	TAATGATGGAAATATRACAAATATTCAACAT
	18563	94	A	0.93	G	0.07	0.12	GGTTCACTAATGTGARGACATGGTGTGGCTC
	18582	69	T	0.97	A	0.03	0.06	TTTTTCATTGTGAGAWGTGCCATAATTTATT
40	18612	37	A	0.73	G	0.27	0.39	TCAAGTTTGGAAATGRTATTTGCAAGCAGCA
	18618	51	A	0.97	C	0.03	0.06	AGGCCGAGCTAAGAAMCGCTCAGCTTCGTTA
	18619	44	G	0.97	A	0.03	0.06	ACTAACAAGCTTCTGRACAGGAGGTAACATT
	18640	121	T	0.50	C	0.50	0.50	GTGGGGGGGTGCAGAYGTGTCCTCTTCAGTG
	18658	52	T	0.97	C	0.03	0.06	CTGCAACTTCTGCTTYCTCCTTTGCCTCTGC
45	18668	76	C	0.13	T	0.87	0.23	AAAAACTAGGCAAAAYAGCAAAAAGTGCAGT
	18673	29	A	0.50	G	0.50	0.50	TGTTTTAATTGCAARACTTAATTTACAGCA
	18680	75	T	0.67	C	0.33	0.44	ACTCTAGCATCTGGAYGCTCCGTTGTATATT
	18683	22	C	0.87	T	0.13	0.23	GTTCAGGACTGGACTYGGTCCCTTTATTGAG
	18694	41	A	0.56	T	0.44	0.49	CAGCCAGCTCTGACTWCTCTCTGTTTCTGTC
	18704	99	A	0.63	C	0.38	0.47	GTTCTCCGAGGGGTAMCCAGCAGGGCCTTCA

	18715	76	G	0.94	A	0.06	0.12	GAGCTTTGTACATGGRCTGGGAGACAAGGGA
	18723	71	T	0.50	C	0.50	0.50	AGATTTTTGAAAGTGYAACAGGTACATAGGT
	18723	94	G	0.69	A	0.31	0.43	TACATAGGTAACCAARTATATAGCTTATTTG
5	18723	96	A	0.63	G	0.38	0.47	CATAGGTAACCAAAGRTATAGCTTATTTGGT
	18740	96	C	0.56	G	0.44	0.49	TTTACCATCATGTATSAGTAGTGGATAATTC
	18740	104	G	0.50	T	0.50	0.50	CATGTATCCAGTAGTKATAATTCATTTTGAT
	18741	23	T	0.88	G	0.13	0.22	GTCAAGGCTTTGGACKCTCTTCAGTCATCAG
	18741	38	G	0.75	C	0.25	0.38	ATCTCTTCAGTCATCSACAGAGTATCTCTGC
10	18741	64	G	0.88	A	0.13	0.22	CTCTGCTCTAGACCTCTGGAGTTCAAGCTT
	18742	51	C	0.94	T	0.06	0.12	CACTTTTGCCAATGTATCGGGTTTGGTTTT
	18746	114	G	0.94	A	0.06	0.12	TTTTGTAATATTCTRTCCACATTCTACTTC
	18763	38	A	0.50	G	0.50	0.50	GTACAATGGTGTGGGRTGACGATGATGTGAA
	18763	53	A	0.88	G	0.13	0.22	AATGACGATGATGTGRTATTTAGAATGTACC
15	18768	120	C	0.63	T	0.38	0.47	CATGTGCACCCTTGGYTTTCGCTCCATCGCCT
	18771	57	A	0.81	G	0.19	0.30	CTCGGAGGATGCCTARAGATGTTGGGAACAG
	18771	75	G	0.88	A	0.13	0.22	GATGTTGGGAACAGARAAAATAAACTGAGTTT
	18790	49	A	0.56	T	0.44	0.49	GTCAACCAGGACAGAWGCATGGACAAGGGAT
	18820	70	T	0.56	C	0.44	0.49	ATGAAATTCTGAGGCTGATTTAAATCTTTC
20	18821	69	C	0.44	T	0.56	0.49	CCTCTCTCGGAGGCCYAGAGGCTGGGGGTAG
	18821	76	T	0.38	C	0.63	0.47	CGGAGGCCACAGAGGYGGGGGTAGCCATTGT
	18846	49	G	0.94	A	0.06	0.12	AGAGCAGGAGGTGCCRAAAGCTGGGAGCGTG
	18851	90	T	0.88	A	0.13	0.22	TTTCCTTATTGTATTWGTAAATATAGGATCCT
	18882	94	C	0.81	T	0.19	0.30	ACAACATCCTCTGCCYACACAACAAAACGTA
25	18908	70	G	0.25	C	0.75	0.38	AAAAGGGTCAGTATGSTTAGGGAAAACATTC
	18910	112	T	0.63	C	0.38	0.47	ATCACTGTGCTGCTTYGGCTCATGGCAGAGC
	18919	26	C	0.50	T	0.50	0.50	CCACAGGGATTCCGGYGCCAGACCCCATTTT
	18922	74	G	0.88	A	0.13	0.22	TCACCTGGACTTAAGRTCTGGCTCTAATTCA
	18932	177	C	0.69	T	0.31	0.43	ATATCTTTGAGTTCAYCTTTAGTACGTGTGG
30	18944	147	A	0.13	G	0.88	0.22	CCCAAATGGCTAGAARTGTTTAATTAATTTT
	18952	232	G	0.38	A	0.63	0.47	TTGGGAAAAGGTGTARACAGTAGCCCCATCA
	18959	123	G	0.56	A	0.44	0.49	TCGAGAAAAGAGGCACRGAAGCCGTCCTGGC
	18972	112	A	0.56	G	0.44	0.49	TGGGCTGGGGAAGCARTGCTTGCTGGCCATG
	18984	208	A	0.94	C	0.06	0.12	GTGATGCATTTATCTMATAAAAATGCTAAATG
35	18985	105	C	0.13	T	0.88	0.22	TACAGAGGTAGCACAYTGATTCCAACACAAA
	18987	35	G	0.19	A	0.81	0.30	CCTGCCAGCAGCCTRTGTGGCCAAGCCCAGA
	19016	161	C	0.75	T	0.25	0.38	TTAGATACATAGCCGYTGATACAGAGGTTT
	19016	184	C	0.75	A	0.25	0.38	CAGAGGTTTCTCTCAMCTCAACACTATTGAC
	19021	20	C	0.44	G	0.56	0.49	CTCTGCTGCTGTCCASACTGTCTTTTGAAC
40	19034	45	T	0.69	C	0.31	0.43	GATGAGGATAGGGGAYACTTCTATTACATTA
	19037	47	C	0.94	A	0.06	0.12	TCTGGTCCTAGCCACMCCTGTATGACCGCGC
	19037	155	A	0.75	G	0.25	0.38	TCCCCTTACGAACACRAAACCAGCCCACAT
	19041	198	T	0.50	C	0.50	0.50	CCTTCTCAATACAGCYGCCCTTGCACTCCCT
	19042	193	A	0.81	C	0.19	0.30	TAAATAACTCTAACCMGGCTGTGTTTAGATT
45	19057	175	G	0.50	A	0.50	0.50	CAGATCCCCACAGCTRTCTCTTTCATCTTGGT
	19064	66	T	0.25	C	0.75	0.38	TGCTGGGCTGTGTTTCYCGGGCTCTTCTGGAC
	19066	72	C	0.56	T	0.44	0.49	CAGTGAGCCACAAGCYTTAAAACCCATGAAC
	19066	87	C	0.44	T	0.56	0.49	ACTTAAAACCCATGAYCTTCAGCTGATCGTC
	19066	100	G	0.94	A	0.06	0.12	GAACCTTCAGCTGATRTCTTAGCCAGTCCA
	19066	147	G	0.81	C	0.19	0.30	TGGCATATGTTCTTGSTTGGTCACCCTGTAG

	19066	148	T	0.75	C	0.25	0.38	GGCATATGTTCTTGTCYTGGTCACCCTGTAGC
	19066	184	C	0.38	T	0.63	0.47	TTACTTCTCCATATTYGGATGCTCAATTACA
	19066	239	A	0.38	G	0.63	0.47	CTTAAACGCCTTCACRGTTTTCTTTTTATCGT
5	19067	57	C	0.88	G	0.13	0.22	GGCTGCTGCAGCCTCSCTGGCTGTGCACATT
	19067	151	T	0.56	C	0.44	0.49	CTTGGGCTCTAGGTCTYGGAGAATGTTGTGAG
	19067	153	G	0.50	C	0.50	0.50	TGGGCTCTAGGTCCSAGAATGTTGTGAGGG
	19067	202	T	0.50	G	0.50	0.50	AGTGTTTCATAAAGAAKACATAGTATTCTTCT
	19076	40	G	0.69	A	0.31	0.43	AAAAAGCAGTTTTAARGTATTCAAATACCT
10	19087	37	A	0.94	G	0.06	0.12	AGCTAAGCTCAAATGRTATTTAACTTCTAGT
	19092	232	A	0.69	C	0.31	0.43	AAAGATCATAATTTTMMATGATTAGCCGTGTA
	19102	25	C	0.44	G	0.56	0.49	GTCACGCTGAGGAGASCTTCACTCAGGAGTT
	19106	247	T	0.94	C	0.06	0.12	GAACCTCCTATTTTAYTGAATTCTGGATCT
	19112	212	G	0.88	A	0.13	0.22	TTGAGGGTGACAAGGRTCTCTTCAAACAGTT
15	19117	134	A	0.38	G	0.63	0.47	ACATAATTGCATGAARTAGCTATTTTTTTCC
	19134	162	T	0.25	C	0.75	0.38	AGCCAGGGCTAGAGGYGCACGGTGGCTAGAG
	19134	263	C	0.94	T	0.06	0.12	GGAAAGGGTTGATGCYATCATTATTGAGGG
	19135	20	G	0.75	A	0.25	0.38	TACCCTGCTTGCCTRAAAGTGCATCAATT
	19139	66	C	0.88	T	0.13	0.22	TTTACACGAGGGTAGYGGCAGATGCCTGACA
20	19139	110	C	0.63	A	0.38	0.47	GCAGACAACACACTAMATTTTACGGGTGTG
	19144	222	G	0.38	C	0.63	0.47	GGCTCTGCTGGAGCGSTGGGAACCAAACACC
	19179	170	G	0.19	A	0.81	0.30	ATAAACATATCAACCRTAGCATTAAACCATT
	19183	210	G	0.50	C	0.50	0.50	GCTCTGCCCTTGGASTGCATTTGACCTGCT
	19642	52	C	0.38	A	0.63	0.47	GACACATTATCCCCMGGGTAAACCAGGACT
25	19673	35	G	0.69	A	0.31	0.43	GATGAAGAACATGATRTCCTAGTAGGTAAC
	19673	180	C	0.94	T	0.06	0.12	TGTGAAAACATTTTTYCTTGGACCAGCTGAA
	19724	35	A	0.25	G	0.75	0.38	ATTGTAATTTGGGTARCTGAGTCACGGTGGC
	19765	57	T	0.94	C	0.06	0.12	GTATACCTTGTCTCYATGTATCTTGTCCCT
	19766	31	G	0.81	A	0.19	0.30	GTACATTGGAGAAGCRTGCAGCAGCATCCTT
30	19766	93	A	0.94	G	0.06	0.12	ATAGGAGCCAAAAGTRGACAAAACAGAAGAAG
	19856	63	C	0.63	T	0.38	0.47	TCCCCCTCCTGGAGAYGCTGCGTCCCCAGC
	19909	29	T	0.94	C	0.06	0.12	CTGAATATTCTCTTYTTAAAAATATAATTT
	19911	116	A	0.94	G	0.06	0.12	AACAATGCAATTTTTTRACACTGTTTTGAAAA
	19946	122	C	0.69	T	0.31	0.43	AGACGCACAGAGAGGYTCTTCTGACCCAGA
35	19956	141	G	0.94	A	0.06	0.12	GTCTGGACCTCAATGRCTCTCGGAGAAGCAG
	19970	126	T	0.50	C	0.50	0.50	CCTGCCAGTTCCTCAYGCGGGGACCAGCAAA
	19970	167	G	0.94	A	0.06	0.12	ACTGGGTTGGTCAAARTAGTCACCTTGGCCT
	19984	47	A	0.19	G	0.81	0.30	CACTGACAGGTAAARTATAACATTAGAAAA
	20014	214	T	0.81	C	0.19	0.30	AGTCACCAAGCATACYTCTGGCTCCCCAAG
40	20096	21	T	0.81	C	0.19	0.30	TGGGGGCATTTATTTYGATAGAGACTGGCAC
	20103	168	C	0.56	T	0.44	0.49	AGCTGGGTCTCCCCYTTCTTCTGCTCAAA
	20113	60	T	0.75	C	0.25	0.38	AAGACCTGAAATACTYGGAAACAGTAAAAGC
	20122	135	T	0.88	C	0.13	0.22	CATTCAAGTTTGACAYTGAAAAACCAACTGG
	20146	31	T	0.88	C	0.13	0.22	TCATTGAGCAGTTAGYCATTGAGATAAAGT
45	20218	26	T	0.94	C	0.06	0.12	TGGTTTTATAAAGCTYAGGACAGAGCAGAGA
	20295	154	T	0.25	G	0.75	0.38	CCAGTCTATTGCCAGKCCAGAGAAAGCGCGG
	20310	125	G	0.38	A	0.63	0.47	CTCTCTAGAGGCTCCRTCAGAACTGGACCCT

	20907	241	A	0.63	C	0.38	0.47	CTAAAAAACATTTTTMAATTATCTAAACAAA
	20964	87	G	0.44	A	0.56	0.49	GGTAGTCCACAGAATRGACACAAGAAACCTC
	20993	139	A	0.75	G	0.25	0.38	AAAACCTGGGCTTCRTAACAAAGTGAGTATA
5	21006	106	A	0.69	G	0.31	0.43	ACACATGTGCACACARAGAGGCAAGTACAAA
	21016	207	G	0.94	A	0.06	0.12	GTGCGCTGTGGGTCRRTGGCGTGGTGATGT
	21028	121	A	0.75	C	0.25	0.38	TTGAGCAATCTAGGGMTATGTGACAGGGGTT
	21028	139	A	0.75	G	0.25	0.38	ATGTGACAGGGGTTTTRTGCAGTGGTACAGAA
	21031	31	C	0.75	T	0.25	0.38	GACCTCTGACATGTGYCTCTGGTCCCCATTT
10	21034	148	T	0.88	C	0.13	0.22	TGGGAGATTGGATAGYGCCTAACCTATCTCA
	21054	23	G	0.13	T	0.88	0.22	CTGCATGGGTACAAAKTCCAATTCATACTTA
	21059	63	C	0.56	T	0.44	0.49	TTCCCACTGAGCCTGYTGAACACAGCTGCC
	21059	181	T	0.50	C	0.50	0.50	AGTCATTTCTCTATTYATTGTAGCCAGGGCA
	21079	50	G	0.94	A	0.06	0.12	ATGCATGCAACTGTGRCGCAAAATCAAGTTG
15	21079	166	G	0.94	A	0.06	0.12	AATATCTGCTAGTGGRAATTTACAACCCACT
	21122	42	C	0.75	T	0.25	0.38	AAGCTAAAGTTATTCTYTTAACAGGAACTCTG
	21139	165	T	0.44	C	0.56	0.49	TAAAGGAACTAATACYGTACAGCACTTCAGC
	21149	167	G	0.13	A	0.88	0.22	TGGAAAGCTTTTACARTGCTTCAGAATGCGG
	21155	36	A	0.75	G	0.25	0.38	TTGATGGAAAATTGGRTCTGTGTAGAATGAT
20	21186	95	G	0.25	A	0.75	0.38	CTGAGGTGGGGCTTARAATTAGTATTTTCGAA
	21190	39	T	0.56	C	0.44	0.49	TGTTTGTATAAACTAYGTGGGGTAAGCCCTT
	21202	61	T	0.94	C	0.06	0.12	GTATAAGCTAAATATYTGATCTGTTTTATGA
	21202	156	A	0.94	C	0.06	0.12	GGAGAGAGTTGACCAMGTCTACATGCATAGA
	21235	43	T	0.06	C	0.94	0.12	GGGAGGAGGGCAGTYCTCGGGCCGATGTTC
25	21242	115	G	0.44	A	0.56	0.49	GGGAGGGGCAGAGAARCACTAGCTTGGGGGT
	21254	53	A	0.88	G	0.13	0.22	AACTATTCCACAGGARCAAGGAGAAGCTGTT
	21376	188	A	0.25	G	0.75	0.38	TTAGGGATGAGTTCRGAAGTGATTCTGAAC
	21399	75	C	0.75	T	0.25	0.38	TCTAAAGTTTCAGTYYTTCACCAGTAAAGGA
	21444	39	A	0.31	G	0.69	0.43	AAGAAATACTCTCAARAGTTCTTTTTTATG
30	21485	82	C	0.69	T	0.31	0.43	GCAATTTTCATGCAGYTGTCACACAGTACA
	21504	147	C	0.81	T	0.19	0.30	TCCCAGATGCAACAA YCGGTTCTGGCTTCT
	21512	54	C	0.94	G	0.06	0.12	ACAAAAATATTTCTGSTAGAGAGGGAAAGAG
	21513	192	G	0.31	A	0.69	0.43	TAAAGAGGCAGGTARAGTAGTATTCTCTAC
	21524	35	A	0.94	C	0.06	0.12	AATGAAAAGGTGTAAMGCCTGATGTACGACC
35	21524	97	C	0.81	T	0.19	0.30	GCATTCCTGTCTACCYGTAGTATGCTTCTCTC
	21552	66	G	0.69	A	0.31	0.43	TACAGATACACAATGRTAATAATTACTTCAG
	21552	166	C	0.88	A	0.13	0.22	ATTATTTAAAAATGTMAATTAATTTATTAT
	21561	55	T	0.63	G	0.38	0.47	CTATACCTTCGAAACKCCTCTTAACCTCTCC
	21627	106	A	0.50	G	0.50	0.50	TCAACTTGAGTACCTRTTATGGATATTTATG
40	21627	153	A	0.94	G	0.06	0.12	GTAAGGGCATTGCAARTCCAAAGTCATCTAA
	21636	71	A	0.81	G	0.19	0.30	TACCAGCTTTTTTAARTAGCAATATCTATTA
	21660	120	C	0.94	T	0.06	0.12	CAAGCCTAACCTGGCYTGTCTTTTTTCAGGCT
	21661	117	G	0.94	C	0.06	0.12	ATTTTAAAAATAAAATSTTTAGTCACAGTCAC
	21703	134	A	0.31	G	0.69	0.43	CATTGGAGCCTACACRCCTTGTGCTTTCTCA
45	21703	197	A	0.31	G	0.69	0.43	GGAGTGAGTCTGGGARGTGGGCAGAGCACAG
	21723	82	G	0.94	A	0.06	0.12	ATGGACTTTAAAGCTRACATAAAAATTAGTAG
	21723	125	A	0.25	G	0.75	0.38	TTAGTCATATTTCCCRCAACAGCATGATAAA
	21763	135	T	0.38	C	0.63	0.47	GACTGTTCTCAGTCAYGCTCTCCACAGCTG
	21763	154	A	0.38	G	0.63	0.47	TCTCCACAGCTGATRCAGACATTGCCTGTG
50	21778	155	T	0.56	C	0.44	0.49	TGGCTTCTGAGGTCYGGTAGAAGGAGGGCA
	21863	47	C	0.69	T	0.31	0.43	GCCCTGGCCCTGCCYAGCTGCATGCCACCC
	21909	153	A	0.25	T	0.75	0.38	TCTTAACATACCAA WAGTGGAAATCAATAGA
	21930	146	G	0.56	C	0.44	0.49	TCCCCATTTTGAGTCS CATAGTCCATTATAT
	21956	26	T	0.63	G	0.38	0.47	TCTCTTTCAAGTGAAKTTCCTTTCGTTCTG
55	21961	73	G	0.13	A	0.88	0.22	TTACTTTTATTTTCRTAAGTTATTGGGGTA
	21961	200	T	0.94	G	0.06	0.12	TTTTATCCCTCGCCCKTCCCCTTTTCCCC
	21965	112	A	0.25	G	0.75	0.38	GACCTCCCCACAGCRCCCCACAGGGTTCT
	21966	148	G	0.69	A	0.31	0.43	AGGGGATTGCAATGGRAACAGGATAAAAAGG
	21980	25	T	0.63	C	0.38	0.47	ACACATTCATCAAGGYAGATTAATTAATGTC
	21981	61	T	0.31	A	0.69	0.43	TCTTGAAGAAAAAA WGTCTCCCTTATGGGT
60	22012	57	T	0.56	C	0.44	0.49	GCCTACATCTGGAATYCATACATCAACGTT
	22020	27	C	0.75	G	0.25	0.38	TGCAGTGGGATGAASTTATCATGATGCTAA

	22082	67	C	0.19	T	0.81	0.30	AGTTATTGGTTGTGYGTTTTCCTTTTTGCA
	22082	179	G	0.88	A	0.13	0.22	GCCGAAGGACGTATTRCTGAACTGGGACGAG
	22091	205	G	0.94	A	0.06	0.12	TTACTTGAGGGCAACRAATTACGGCTTAACA
5	22132	99	T	0.81	G	0.19	0.30	GCCTTTTACTATCCTKCCCATTTCTTCTAA
	18017	87	C	0.25	A	0.75	0.38	GGCAACCCCNNGAACMACTGCTGGATAAATC
	22202	128	A	0.94	G	0.06	0.12	TGAAATCTGAATTTTCRTTAATACTCTGGTGC
	22283	109	T	0.94	C	0.06	0.12	CTGCAGGCTCTGGTTYTTTCATTTGCAAAATA
	22292	53	A	0.94	G	0.06	0.12	ATTGCTCAGTACCAGRGTGAGTACGGTCTG
10	22387	186	C	0.81	T	0.19	0.30	AAGGCAGGATTGTGGYCCTTGTGTTTTCTGA
	22405	90	A	0.88	C	0.13	0.22	ATTGGCTGTAAAGTCMGATCAGGTGCTCTCC
	22440	64	A	0.94	C	0.06	0.12	TTAAGCCACTTGGGTMTCCATTCCAGCTCTG
	22457	112	G	0.75	A	0.25	0.38	AGGCATGAAGGATACRCAGTTAATTAACATA
	22585	56	A	0.63	G	0.38	0.47	TGACAAGTGAACAATRCAGAAGCAGCAGTGA
15	22631	52	T	0.81	C	0.19	0.30	CTGGCTTCAGTTCTGYAGCACCATTTTCAAG
	22652	32	G	0.50	T	0.50	0.50	GCCACTTTGGAGAAKAAGAGAATGCTATTA
	22663	38	C	0.81	T	0.19	0.30	CTTCTCACTGCACTGYGAGGTGAGCCGGCGC
	22663	55	C	0.56	T	0.44	0.49	GAGGTGAGCCGGCGCYGCTAATCTTATTTCC
	22663	139	G	0.81	A	0.19	0.30	TGGTGCACCTACAGGRGAAGAGCTTCTCAT
	22714	212	C	0.63	A	0.38	0.47	GAGCTTACCAACCCMTGATAGGGGCCAAA
20	22724	117	A	0.56	G	0.44	0.49	AAAGCTTGCTAAGGGRTTATTCTATTTTTG
	22750	48	G	0.88	A	0.13	0.22	AGCTGAGGCAGCTAARGGCTCATACAAAGGT
	22775	60	A	0.69	G	0.31	0.43	TTCCATTGTTTACATRTAGTAGGAAAGGGAA
	22808	143	C	0.50	T	0.50	0.50	ACCAGGAGGATGAAGYAGCAAACCTGATTAAG
25	18148	101	A	0.13	G	0.88	0.22	CGATTCTGAATATCCRTGGCGGCATATGCAA
	18254	64	T	0.56	C	0.44	0.49	AGAGCAGTAAATCAYGCCAAAATTCCTCT
	18265	117	C	0.88	A	0.13	0.22	AGCCATGAACTGGCTMGTTTTCAACCTTTCC
	18295	40	C	0.94	T	0.06	0.12	TGTGGAGAACAACAYTTGGGAAGTAAAGGT
	18459	64	T	0.31	C	0.69	0.43	GGGTGGGAGACACAAYGAGTAATTAACAACA
30	18501	121	C	0.88	T	0.13	0.22	GCAGGACAGAGGGGCGYGGACAGCAGCGCATG
	18548	62	G	0.56	A	0.44	0.49	AGTCCCCTCACTGGGRAAAAAAAGCATCTN
	18548	65	A	0.94	G	0.06	0.12	CCCCTCACTGGGGGARAAAAAGCATCTNTCA
	18700	97	T	0.13	C	0.88	0.22	TGCTGAGAGCAGAGCYAAGATCCACAATTGC
	18829	35	T	0.0000	A	1.00	0.0000	GGGGAAAAATCCTAGWAATAACTTATGTGTA
	18829	58	A	0.44	G	0.56	0.49	TTATGTGTACTTCTTRTTTCATCATAAAGA
35	18916	35	G	0.75	C	0.25	0.38	CCAAACATCTTCAGCSCTCAGCCGGTCTCCC
	18916	42	C	0.75	T	0.25	0.38	TCTTCAGCAGCTCAGYGGCTTCCCACCTTCT
	19105	33	T	0.19	C	0.81	0.30	GGACAGAAAGAATATYGTGGTCCATGTGGTT
	19105	211	C	0.94	T	0.06	0.12	ATCTCCCCACAACCTYTCCAGGGGCAGGATT
40	19576	113	A	0.81	G	0.19	0.30	AAAAAATTTAACATTRTCTAGTTCAGTGATT
	19828	200	A	0.56	G	0.44	0.49	CACCACCACCCAAAARCTTTAATTCTGGAA
	19860	51	C	0.50	G	0.50	0.50	AATGTTTCCAAAGATSCTGCATCAGTATCTC
	19888	98	C	0.13	T	0.88	0.22	TAGAAAGTAGCAGTGYTGGACAACGTGTGAA
	19889	80	C	0.56	T	0.44	0.49	CAAGAGGAGTGAGGGYTACAGCATTATTTTC
45	19891	172	C	0.75	G	0.25	0.38	GCCATCTGTCTGACTSCGTCTTCCCGGGGCG
	19937	185	C	0.75	T	0.25	0.38	GTGTTCTCAGCAAGYTCCAAACCTTCCAA
	19937	186	G	0.81	A	0.19	0.30	TGTTCTCAGCAAGTRTCCAAACCTTCCAAA
	19941	71	C	0.38	G	0.63	0.47	ACAAGGTGAAAGGTASGGTCTGGTGAGACA
	20059	59	T	0.63	A	0.38	0.47	ACAGAGTGGATAACCWACATTGGCTGGAATG
50	20116	22	C	0.75	G	0.25	0.38	ATTTCTGTCCACCASCTGTCCACCAGTTAT
	20116	59	T	0.75	A	0.25	0.38	CCTTCAATATATGGCWTAGAACATATATAAAA
	20116	69	T	0.81	A	0.19	0.30	ATGGCGTTAGAACATWATAAATCTATATCAT
	20155	81	C	0.75	T	0.25	0.38	CATTCCTTGCGGGYGCAAACTGCTTTGA
	20258	157	G	0.88	T	0.13	0.22	CCGCGGGGTGTTCAKCGCGTTGACGCAGGT
55	20270	53	G	0.94	A	0.06	0.12	ACAGGAGTGGGGACGRTCAGTGTACAATACA
	20270	91	T	0.31	G	0.69	0.43	TCCAGGATAAGGAGCKACACCAGGATTTATA
	20317	217	G	0.38	T	0.63	0.47	AAACCATCATCAGAAKTATTAATTAATTGC
	20329	68	G	0.94	A	0.06	0.12	AGACAAGACATCAARTCTGTAGCAGCGAG
	20442	37	T	0.63	C	0.38	0.47	AAAANGGGGGGGGGCYTAAGGTGGCACAAAT
	20466	133	G	0.63	A	0.38	0.47	TGAAGTGAATAAACGRTGTGAACTAATGTTT
60	20561	25	A	0.69	G	0.31	0.43	TTTAAGATGGCTGTRTAAGTATAAAGCAGT
	20561	94	T	0.31	C	0.69	0.43	GAAAAATCCTTACATYGGAATCAATGTCTTT

	20601	125	T	0.56	C	0.44	0.49	ATTAGTCTTCTCTGTYCTTGGTGCAAGTTTG
	20622	130	T	0.50	C	0.50	0.50	TATCTTAAAAGTTGAYTACTAATTTTTATGA
	20768	71	C	0.94	T	0.06	0.12	CCTGCCTGCCTGCTCYGACTGATTACTTTCA
5	20768	190	C	0.94	T	0.06	0.12	ACACATACTGCTGGGYCAGGGACTCGTAATT
	20893	179	T	0.63	C	0.38	0.47	CTGGGNAAAACCTGCCYTTTCTTCTCTTTTA
	20893	207	A	0.38	G	0.63	0.47	TTTACAATGCAGTTTRACATAACATTGGTAG
	20934	72	T	0.88	G	0.13	0.22	ATTTGTATTCAGAGAKTCTAAGACAAATGGT
	21117	227	C	0.81	T	0.19	0.30	TCTACAGTCTGTATTYTCTACTGAATCTTG
	21187	94	A	0.19	G	0.81	0.30	CACACATAAAGACACRGGNTCTCAGTAATGC
10	21249	155	T	0.56	C	0.44	0.49	TCTAGGTGTATACTTYATGGAAGTACTTTAT
	21314	122	A	0.63	T	0.38	0.47	CTCTGTCAAACTTTTWTTTTGTTTATAAACT
	21342	59	T	0.38	C	0.63	0.47	ATNAGCAATACACTGYTGGAAATCTGCATGA
	21382	125	C	0.81	G	0.19	0.30	TGGGATNTGGCTTCCSAGGTTGCAACCCCAA
	21437	201	G	0.88	A	0.13	0.22	TCACCTTTACCAGGGRCAGGCATAGTGTGGC
15	21449	222	C	0.75	T	0.25	0.38	ACCCCTCAGCTTCCCYGTGACAGAGCCAGTGT
	21475	117	A	0.81	T	0.19	0.30	AAACCCAGGCTTCTWCTTGCTTACTAAGCA
	21475	181	A	0.75	G	0.25	0.38	GTCTTTGGAGAAGGCRAAAAGCCACAGCAGC
	21514	100	A	0.56	G	0.44	0.49	CATTACAAAACCCCRCTTCAAGGAAAGGA
20	21514	133	C	0.13	T	0.88	0.22	CACATTACCATGGAGYACAGGACTCCAAAGG
	21558	157	G	0.50	A	0.50	0.50	TGGTGGGGGGCAGTARAGCCAGGGACTCCCT
	21569	198	T	0.69	C	0.31	0.43	AGAAATTTATCTCTAYAGAGACAATTCATAG
	21574	235	C	0.44	T	0.56	0.49	TTACTGCCTACTTCCYGTCTGTGTCAGGTGGGA
	21609	42	C	0.94	T	0.06	0.12	TCTCCCTTGTAACAAAYGTGCAGTCCGTTAC
25	21609	146	G	0.88	A	0.13	0.22	AAAGGATGTTTCAAARAGGGTCCCGGCTATG
	21614	55	G	0.69	A	0.31	0.43	TTTGANTATAGCTATRTTTTAAACAAACCTCA
	21615	151	C	0.38	T	0.63	0.47	TTTCACTGAGTATTAAYAGGACACAATCGACG
	21644	151	T	0.81	A	0.19	0.30	TTTCATAAATAAGGGWTTCAATCAAGATCCA
	21687	115	C	0.44	G	0.56	0.49	GGACTTCTTCCCTAATGTTCTATGATCAGA
	21695	141	A	0.88	C	0.13	0.22	CCTTTCCAAGGGAATMTACTACACTAAGCCCT
30	21760	35	A	0.75	G	0.25	0.38	GATGCAAATGATTTGRGGTGTCTTCTAGCT
	21760	81	C	0.75	A	0.25	0.38	GGGACCTCTGACTGCMCCTCTGTCTCAGTTT
	21761	138	C	0.94	G	0.06	0.12	TAAACGTGCCGTGGCSAATACACACCAAAG
	21805	45	A	0.69	T	0.31	0.43	TTTATAATCCTATATWAAAAAAAAATCTATAG
35	21941	79	A	0.13	G	0.88	0.22	AGAGTGAGGGGCAGARGGATGAGGCTCTTCT
	22129	45	T	0.50	G	0.50	0.50	AACTTTAAGGAAAAKTATATAACAGTCAT
	22130	165	C	0.94	T	0.06	0.12	ACCCGCGCGCTTGCYGTGTTAATCCAGGT
	22187	110	C	0.13	A	0.88	0.22	ACATTTAAAAACCAAMCAAAACAAAAA
	22187	178	G	0.69	A	0.31	0.43	TCTATTGGTAATGGTRAAATTTTATGAAAT
40	22189	70	C	0.88	T	0.13	0.22	TGAAGTGTCTATGAYGAGGCGAGGAATGGG
	22250	89	G	0.50	A	0.50	0.50	GGAAATGTCATTCACRTAGTGGGTTATTATG
	22250	132	C	0.94	T	0.06	0.12	TCCTGGCTGTGTTATYGGANCCAGGAGTGGA
	22290	136	C	0.88	T	0.13	0.22	TCAGGACCTTGCTTTYTTCCAATCTCTCCTT
	22374	149	T	0.94	C	0.06	0.12	TTATTAGTAACCTAAYAGGNTCCTGCATCAT
45	22395	127	A	0.69	G	0.31	0.43	GGGGCAACTCTTAAARAAGGAAATGTTACCA
	22419	67	T	0.13	C	0.88	0.22	AGGCACAGCCCAGTYCTGGATGGCATCAGC
	22449	74	T	0.94	C	0.06	0.12	AATACAGTACTTCTTYGAAAAAATACACAAT
	22512	104	T	0.94	G	0.06	0.12	GGTCTTTGTGATCTKACCTCACCCATGTCT
	22668	99	A	0.69	G	0.31	0.43	AGTTTTCTGTAATATRTTCTAGTCCATTAG
50	22734	44	G	0.75	A	0.25	0.38	GGGTCTGGGAAGGCCRTCTTAGAAGACATTA
	stCSF2RB	149	C	0.94	T	0.06	0.12	GGAGCCCAGAGGTTTTYGGGACTCCCAGCCA
	stCSF2RB	192	G	0.94	C	0.06	0.12	CCAGCCCCAGAACCTTSAAGTCTTCTTTGACG
	stD22S100	88	G	0.94	A	0.06	0.12	CTGGCAGGAAGAAGRGATCCAGCAGTGAG
	stFIBB	341	T	0.69	C	0.31	0.43	CCCACCTTTGAGCTYACCTGCCCCACCCCA
	stFIBB	412	G	0.56	C	0.44	0.49	TTGCCCTTCCCTGAASTGCCTTCTTGTGGCT
55	stIGLV2	61	T	0.56	C	0.44	0.49	CTCTGCTGCTCCTCAYCTCCTCACTCAGGAC
	stSG10017	33	G	0.81	A	0.19	0.30	ACTCCTGGTGTCAAGRATCCTCCCACCTCGA
	stSG10017	70	T	0.44	C	0.56	0.49	CAGGGTGTGGGATYAGGCATGAGCCCCCA
	stSG10023	63	A	0.31	T	0.69	0.43	CCAATATCATTGAGGWAACAGTTTGGGCTGT
	stSG10096	36	G	0.44	C	0.56	0.49	CTCCCTCCCCATGACSGGCTTCCCAGGCA
60	stSG10118	107	C	0.50	A	0.50	0.50	TGCCCATTCCTTGCMTCTCAGCCCTCAGTTC
	stSG10120	89	T	0.94	C	0.06	0.12	CACGAACACTTTAATYGTGTTGTAATCTGA

	stSG10178	42	C	0.75	T	0.25	0.38	CTGGACATTAAGTCCYGGGAGGAGAAGTGAA
	stSG10193	136	G	0.75	A	0.25	0.38	TATACAAACTTTTACRTTTGAAAAGTGGAT
	stSG10202	143	G	0.94	T	0.06	0.12	CTGCTTCTCGCTGTCKCAAGACCACAAGGCA
5	stSG10209	34	C	0.56	T	0.44	0.49	CTCAGTCACCATGATYAAAATAAACTAATTCT
	stSG10209	75	A	0.94	G	0.06	0.12	CCCCTTTATTTTTTRCTCCAATAAATGTAA
	stSG10218	29	T	0.38	C	0.63	0.47	AAATGAGAAGATTACYGTGAATATTTAAAGA
	stSG10252	108	A	0.63	C	0.38	0.47	CCTTCCCCTTGATCMAGTGAAGATATGATA
	stSG10266	55	T	0.94	C	0.06	0.12	GAATTGTCTTCTGTGYGACAGTTGAAGTGGG
10	stSG10282	70	T	0.88	G	0.13	0.22	TGAAATCTTTACAAGKAAGCACAGTAGTACA
	stSG10310	128	C	0.38	A	0.63	0.47	AAATAATTTTTTACMTGTGCAATGCCAATG
	stSG10331	107	A	0.81	T	0.19	0.30	TAGACCTCAAACACCWCACCTCCATGCATTT
	stSG10331	116	T	0.94	C	0.06	0.12	AACACCAACACCTCCYGCATTTCCCTTTTGG
	stSG1243	225	G	0.13	A	0.88	0.22	AAAAGAAATCTGTTRAAAGTATTTTCAGACC
15	stSG1345	54	T	0.50	G	0.50	0.50	TTGAACTAGTTTGCKCTTACCGCGCTTCAAC
	stSG1345	60	G	0.63	A	0.38	0.47	CTAGTTTGCTTCTTARCGCTTACATTTTTAG
	stSG1385	117	T	0.94	G	0.06	0.12	GAGACTTGGTATTTTCTCAATCATTAAAGAAG
	stSG139	69	T	0.19	C	0.81	0.30	ACAGCACTTTGTGTCYGCTTTGAGCACTTGC
	stSG1427	103	T	0.25	C	0.75	0.38	TTGGCTTCTGTCTCYAGTCTCTCTCCATGT
20	stSG1471	50	A	0.13	G	0.88	0.22	GTCATGTGTAGGTCRCTCCCTTGATGAAA
	stSG1483	44	T	0.06	C	0.94	0.12	TACTATTTAGTCTAAYTTTAAATTCAAAGGTT
	stSG1696	67	C	0.94	G	0.06	0.12	GCAAAACCAGTGTGCSAATGTGGAGGATGTC
	stSG1847	49	C	0.38	A	0.63	0.47	CAACACAAATGCTACMCTAAAATGAAAGAAT
	stSG1847	95	G	0.38	A	0.63	0.47	AAACAAGTGAGAGACRTTACTTACATCAGT
25	stSG1897	83	A	0.56	G	0.44	0.49	AGGAGGACACAGGACRGCACCACCTTCTC
	stSG2022	86	T	0.00	C	1.00	0.00	TTAACATTAATATACYATTCATAATCTCAT
	stSG2034	166	T	0.81	A	0.19	0.30	AAAATAGTACATGTTWGTGTAATAAAATTA
	stSG2076	104	C	0.94	G	0.06	0.12	AATATATTTTGACATSACATCACAGTGGGGC
	stSG2108	49	T	0.19	C	0.81	0.30	CCAACCAAAAAATGAYGAGGGGCTCCACAGA
30	stSG2108	71	A	0.19	G	0.81	0.30	GCTCCACAGAGAGARTAAGGGGAAGACTTT
	stSG2141	113	C	0.94	T	0.06	0.12	ATGGCAGCACCCTGYATGGCGATGGTGCAG
	stSG2141	173	A	0.75	G	0.25	0.38	GCTTGAAGAGAGAAAARAAGTTCCTATTATT
	stSG2148	50	A	0.88	G	0.13	0.22	TTTAGACCGTGATTTTAAAGAAAACAATAAAT
	stSG2175	68	C	0.94	T	0.06	0.12	AAATCTGTTGTGTGTCYGCCGCTGACTCAGC
35	stSG2189	41	C	0.69	T	0.31	0.43	CCTGATATTCACACTYCTACATTCCTCCAG
	stSG2200	49	T	0.25	C	0.75	0.38	CTGGTCTGTATGATYTTTATATTTATGTAT
	stSG2218	48	C	0.81	T	0.19	0.30	AAGAAAAAAATCCTCYTTAAAAAAACAATAA
	stSG2218	139	G	0.94	T	0.06	0.12	GCATTTTGAATTTAKGTTTGAATAAATAAC
	stSG2218	201	A	0.44	T	0.56	0.49	AAACATTTCTGGTATGWTATTGTGAGTGGTGC
40	stSG2243	85	G	0.81	T	0.19	0.30	ATGGTCAGTAGAAAAGKAGAGCATCTCCTCAG
	stSG2257	65	A	0.94	C	0.06	0.12	GCTATCAGAAGGGCAGTGTGAGGAACTCTC
	stSG2306	67	A	0.13	G	0.88	0.22	TGGGAATATTTTACRTATGCTCCCATTGGG
	stSG2334	70	T	0.38	G	0.63	0.47	CGCAAAAAACAATAAATGTCAGTGGAGGGGGC
	stSG2339	63	T	0.44	C	0.56	0.49	AAGTAACTGCTGTCAYGTTTCTCAGAGTACC
45	stSG2465	76	C	0.13	T	0.88	0.22	CAAAATGCAGAAACCYTACAGATTAATAAGAG
	stSG2549	140	T	0.69	C	0.31	0.43	GCAGCTAAAGGAATAYTACACCAACCCACCC
	stSG2577	121	C	0.13	T	0.88	0.22	AACCGAACTGTGAAAAYATGAACAATCCCGGC
	stSG2577	123	T	0.88	G	0.13	0.22	CCGAAGTGTGAAAGCKGAACAATCCCGGCC
	stSG2700	58	G	0.31	A	0.69	0.43	TGAACTGTCCGGCCCRAGTCACTCAGCGTTT
50	stSG2724	101	T	0.38	G	0.63	0.47	ATTGCTTGCATAATCKTTTTTTAATCCTGG
	stSG2776	65	G	0.50	A	0.50	0.50	AAAGTCTCGAATATGRTATTGGCCCTTTTGG
	stSG2791	100	A	0.44	G	0.56	0.49	TAAACTAGCAATTTTTTAAATATTGGGGTCC
	stSG2791	109	G	0.88	T	0.13	0.22	AAATTTAATAAATAKGGGTCCACTTAAATC
	stSG2826	85	C	0.50	T	0.50	0.50	CTCCCTCAAAAACAAAYGAACAAAAATAAGA
55	stSG2850	88	G	0.56	A	0.44	0.49	CCCAAGGGAGACGGCRGGCTCACACATCCCA
	stSG3031	71	T	0.94	C	0.06	0.12	CTGTGGTGTGAGCAAYGCCCTTTATTTTAA
	stSG3058	81	G	0.75	A	0.25	0.38	TGAAAAAAGTCAAAARTGAAGAAGCATCAAA
	stSG3092	94	T	0.94	G	0.06	0.12	TAATAAATGAACGTGKGATAAACATTCTTCT
	stSG3230	95	A	0.63	G	0.38	0.47	AATTGTCAGTGGAGTRGTGGGGTGGCTAAGTG
60	stSG3245	160	G	0.81	C	0.19	0.30	CCTACCTGGGAGGTTSTGTACTTGGCTTAAG
	stSG3265	42	T	0.88	C	0.13	0.22	ATTTATTTATAAGGAYGCATTGTGAATAGTT

	stSG3269	24	A	0.50	G	0.50	0.50	CTGTGTCATCCTATCRTTCCCTTCCCTGAGC
	stSG3269	141	C	0.81	T	0.19	0.30	CCATGCTAAAGCATGYTGTAGATCCCCAAGT
	stSG3284	130	C	0.75	T	0.25	0.38	CACTCAGACTTCCCCYTCCCTAACTTTTGTT
5	stSG3292	99	A	0.63	T	0.38	0.47	TGACTTAAATATCTAWTACAAATCAAATAGC
	stSG3323	26	C	0.94	A	0.06	0.12	ATCTTTAGCTCTCACMCCAGTGTATCCATTT
	stSG3369	69	C	0.63	T	0.38	0.47	AGGACCACTCAGAGGYATAAGGGAACCCTCT
	stSG3398	125	G	0.56	T	0.44	0.49	CTGTCACCTTTGTAGKCTGGGTCAAAGTCTA
	stSG3416	43	A	0.06	G	0.94	0.12	AAAGGATGCAATCACRCTCACTGTAGCCTGG
10	stSG3424	173	T	0.44	A	0.56	0.49	TGCTGGGTAACACTGWCAAGTTGCTTAACCT
	stSG3436	88	T	0.31	A	0.69	0.43	TGGCAGAGAGGGCCCWAAAATAGCTTACTCT
	stSG3463	103	C	0.19	T	0.81	0.30	CAGCTCAATGGGTCAYTGGAACAAAACCTTGCT
	stSG3470	123	A	0.81	C	0.19	0.30	TTACGATCATTTTAAMATTTTAAGAAGCTGAG
	stSG3491	71	G	0.81	A	0.19	0.30	AAGGACGATTTGAAGRGTGGAATTACTGTGC
15	stSG3492	71	G	0.88	C	0.13	0.22	TAAGGCCATTCTGTGSTATTTTTAAACTT
	stSG3523	33	C	0.63	T	0.38	0.47	TTCTTTTGGGTTTTYGCATATATGTGTGTA
	stSG3536	213	A	0.63	G	0.38	0.47	GCTTGTGCACCATTARTCCTGCTGGGTGTTT
	stSG3583	112	G	0.88	A	0.13	0.22	ACATCCACACAGGCARTAACATACACAGTAC
	stSG3586	60	G	0.94	C	0.06	0.12	ATCAGGTGTGGTGGTSACGCCTGTAGTCCCT
20	stSG3589	101	T	0.13	C	0.88	0.22	CAAAAAACCCCAATGYCCTATTTCCAAGAAT
	stSG3590	70	A	0.81	T	0.19	0.30	GTTCTAAAAAATAAATTTCTCTGATGCTC
	stSG3619	78	A	0.88	C	0.13	0.22	TACGCTTCTGTCAATMAACAAACTTCCAGAG
	stSG3644	40	T	0.94	C	0.06	0.12	CATATTTAGGATGAGYGGATTGAGAGGCATG
	stSG3646	43	A	0.63	T	0.38	0.47	TTGGCAAGAATATATWTGATAACAATAATAT
25	stSG3646	55	A	0.81	G	0.19	0.30	TATGATGATAACAARTATGTCTTACTGGTG
	stSG3646	70	G	0.38	A	0.63	0.47	AATATGTCTTACTGGRATATTAACTTTGATA
	stSG3693	30	C	0.88	T	0.13	0.22	CATCCCCTGGTCTCYTCTGAAAGCCGATGA
	stSG3693	85	A	0.75	C	0.25	0.38	AAATATCCTACGAGGTCGCCCTCCGAGACT
	stSG3698	51	C	0.88	G	0.13	0.22	CCAATCCCAGGGTTSTCTCTGACTTCCACC
30	stSG3698	145	G	0.88	A	0.13	0.22	CTAAGTCTTTATTGGRAGAATACCCACCCAC
	stSG3724	107	C	0.88	T	0.13	0.22	GCTCAGTGTGTGAAACACAGGAGTCCCTC
	stSG3725	104	G	0.56	A	0.44	0.49	CAACAGCAAACAGCCRAGCAGGAATCGGCAC
	stSG3751	128	G	0.56	A	0.44	0.49	GAGAGGATATGGTCCRTTGCTGACTCCATGT
	stSG3787	49	T	0.44	A	0.56	0.49	AGCAGAGGATCTCTTWAAGTTCCCTAAGAC
35	stSG3880	36	G	0.56	C	0.44	0.49	CCAGAGACCAGGGCTSGGCAGCTGGGGTCC
	stSG3880	115	G	0.50	C	0.50	0.50	CTGGGGAGCAGGCTSGGCACGGGAGTGC
	stSG3895	44	A	0.88	G	0.13	0.22	GTATTGTTAGTGTGRTTTTTTTTTCCATTA
	stSG3902	104	T	0.88	C	0.13	0.22	GAACTGCTTTCTTTTYTCAGCTCAATAGCTT
	stSG3935	50	G	0.88	A	0.13	0.22	AACAAGCAATTTGTCRCTAGTGTGCAGGCTC
40	stSG40	25	A	0.75	G	0.25	0.38	GTTGAAGAAGTGCTGRAAATATATTTAAGAT
	stSG4009	32	A	0.69	G	0.31	0.43	GATGAATGGCGCGCTRTACTCTTTACGGTCT
	stSG4033	123	T	0.75	C	0.25	0.38	AGCATAAAGGTACTTYTGTGAACAGGTGGGC
	stSG4038	29	G	0.88	A	0.13	0.22	GTACAGCCACGCCTGRCCAGGCCCCTCTG
	stSG406	53	T	0.88	C	0.13	0.22	AGCTAAACGAACAAYGGTTTTAGTTTTGCT
45	stSG4095	27	A	0.81	C	0.19	0.30	ATTAGTCAAGCAGGTMGATACTATTGTCTGC
	stSG4095	55	G	0.81	T	0.19	0.30	CTGCTAGATGTATTAKATAAAAAAGTTTGCT
	stSG4120	65	G	0.94	A	0.06	0.12	ACTTATGGATAATCARCTTTTTCCCTCAGA
	stSG4128	54	A	0.88	G	0.13	0.22	CTTTGTGTACATTTCTRTATATTATTTACTT
	stSG4209	65	G	0.81	A	0.19	0.30	CATCCACATGGCACARCAGGGCCGGCCACTC
50	stSG4209	128	G	0.88	A	0.13	0.22	GAGGCCGCACTCCCTRGCAGGGGGACCACGG
	stSG4254	31	G	0.56	A	0.44	0.49	CATGGAGGACCAGRCCACGGCCGGGACTC
	stSG4301	81	T	0.38	G	0.63	0.47	ATTTAAGCAAATAAAKAGCTTCTGAGTAGTT
	stSG4331	71	T	0.25	G	0.75	0.38	GTTTTATGACACAGAKTTTTCAAACAAGTTT
	stSG4340	76	G	0.56	A	0.44	0.49	AAAACCACATGTTCTRTAAGTGGGAGATAAA
55	stSG4361	24	T	0.81	C	0.19	0.30	CATTGAGTGACAGAGYCACTCATGCAGAACT
	stSG4361	109	A	0.75	C	0.25	0.38	TAAGTGCATCTTTTGMCTTCCAACTAGAA
	stSG4376	73	A	0.63	G	0.38	0.47	TCCAAGGGGAGAACARCTGGAACCTGCGGCTC
	stSG4381	50	T	0.94	C	0.06	0.12	AACATACGATTTTCTYTCAGTCTTGTAGTAT
	stSG4410	79	A	0.69	G	0.31	0.43	ACCATTCAAACACCCGRTGACAACGAACCCAG
60	stSG443	65	C	0.69	T	0.31	0.43	GGCAGTGAACACATCYGTATGCAATGAGAAA
	stSG4430	54	A	0.94	G	0.06	0.12	AGTAGTTCTATAAGGRATTAACATAGGTAGG
	stSG4448	99	G	0.94	A	0.06	0.12	CCTCTGGGGTCACTGRTGGGTTAGGCCCCCA

	stSG4449	92	T	0.63	C	0.38	0.47	GACAACTTAAAACCTTYTAGTGACATTGCTGT
	stSG4465	60	G	0.94	A	0.06	0.12	CTGCACACTGGAAGGRAAACCTGGGAGAGAG
	stSG4467	42	C	0.94	A	0.06	0.12	CTGGGACAGAGCCTCMAGATGATGTCCATGT
5	stSG4469	74	C	0.63	T	0.38	0.47	GCTTCTTGCCAGGCTYTTAAATTGTGCTGTA
	stSG4475	21	A	0.81	C	0.19	0.30	TCATTTCTGACCAGMTATTAATAGTTTTAT
	stSG4477	32	A	0.94	G	0.06	0.12	GGGGGTGAGACAAACRATGAACCAATAATTA
	stSG4531	79	C	0.94	T	0.06	0.12	GGGACAGCAGGCGTCYGCCACGTCTGGCGT
	stSG4550	85	C	0.56	G	0.44	0.49	AAGAGACAGTGGGCASGCAATTGGAGGGGAA
10	stSG4550	86	G	0.81	A	0.19	0.30	AGAGACAGTGGGCACRCAATTGGAGGGGAAG
	stSG4551	74	C	0.75	T	0.25	0.38	CTCAATGCAATAGAA YTGACATGGGGCCAAA
	stSG4590	47	A	0.94	G	0.06	0.12	AAAAGCTCTTCTG CARATGGGAGGGAGACAC
	stSG4617	125	C	0.75	A	0.25	0.38	GAGATGATCTTCTCMCCCTTCTCTCAGGGT
	stSG4623	22	T	0.56	C	0.44	0.49	TATCACCCAGCGCTGYCAATTGACATAGTAGC
15	stSG4843	102	A	0.94	C	0.06	0.12	CTAAATTTTGTAGTCAMATCAGAAAGCTTCC
	stSG4850	38	C	0.88	T	0.13	0.22	GAGGAGGAAGGGGCTYGTGCACTTGCAGGCC
	stSG4879	86	A	0.38	G	0.63	0.47	CCTGGGACTGGAGCARCTTGGGTGAGCTCTA
	stSG4885	104	G	0.88	A	0.13	0.22	ACGACTACGCTCTGCRGTGGGAAAGCAGAAG
	stSG4896	112	C	0.75	T	0.25	0.38	GCTGGGCACCTTTTCYAGCCACAGGCCCT
20	stSG4932	22	G	0.44	A	0.56	0.49	CCGATGGTTACACAARTGTAAATGTATTTA
	stSG4950	24	A	0.88	G	0.13	0.22	CCCAGGAAAAGGTCCRTTGTAGCTTCCCTCT
	stSG4957	136	G	0.75	A	0.25	0.38	AGGATTCATGAGCCCRGTGACACAGATGGGG
	stSG4961	91	C	0.88	T	0.13	0.22	GATGAAAAGGAAAGTYAGAGAGGGCATTGAG
	stSG4967	72	A	0.06	G	0.94	0.12	TAGGAGTGCAAGGGCRTACCCCGGAGCTAG
25	stSG4997	22	T	0.81	C	0.19	0.30	AGAGTAGGAGCCCA YTTTAAATGGTTTCT
	stSG50	125	C	0.44	G	0.56	0.49	GTTCCGGACCTAGATSTGACGAAGGTAGCAC
	stSG6312	37	C	0.94	T	0.06	0.12	CTTTAGTGCAAAAACYTATGCCATGCGGGAA
	stSG6345	107	G	0.63	A	0.38	0.47	GTGATGTTTTGTCCARATAGTTCAGGCAATT
	stSG6362	88	G	0.94	C	0.06	0.12	ATGAGCACTGTATGTSAGAAAAGGGAGGAG
30	stSG8010	62	G	0.81	T	0.19	0.30	TTTGGGTGTGCACTGKTGTCTTTCAACTGGG
	stSG8022	53	G	0.25	A	0.75	0.38	GCCTGAAATGGACCARGTGGGAGTTATTTAC
	stSG8032	67	G	0.31	C	0.69	0.43	TCAGAAAATTGTGTGSTGGGAGGCAGGGTAG
	stSG8064	23	G	0.94	C	0.06	0.12	TCTTCTTCTGTGCGSTTCGGGAGGCTTCAC
	stSG8064	46	C	0.81	A	0.19	0.30	AGGCTTACGCTCTCMCCGTGGTCCCTGGGT
35	stSG8072	59	A	0.69	G	0.31	0.43	TCTTGCTGTCTTAGGRTGGCAGAGGCAGAAAG
	stSG8100	40	A	0.94	G	0.06	0.12	CTTGATCAAATTCRAAGTGAACATAAAGT
	stSG8102	138	T	0.75	C	0.25	0.38	ATACAAATGTGAAATGYTGCATAAATCAAT
	stSG8105	110	A	0.75	G	0.25	0.38	AGGCCTGAGAAATATRTTTCTAACAAGTTCC
	stSG8130	36	C	0.81	G	0.19	0.30	GGAGGGAAATAAATGSTGGATGGTCTGCTGCT
40	stSG8130	96	T	0.19	C	0.81	0.30	AAGCGGTGCCTGAGCYGTGCCTGTCTTCAGA
	stSG8145	97	C	0.81	T	0.19	0.30	AGAACACAATTGTGAYACAAATCTAAGAAAT
	stSG8145	124	T	0.81	A	0.19	0.30	GAAATGAATGAGATGWCTGAAATCTGATTCA
	stSG8150	36	A	0.94	G	0.06	0.12	GATTTTTCAGAATAGRATAAATAATAACGGG
	stSG8340	30	C	0.81	T	0.19	0.30	AGAGCTGGGCAGGATYCAACATTATGACCCCT
45	stSG8416	65	A	0.63	G	0.38	0.47	CAGGCTGCCTACTCRTGTGGTTTGCTAGCC
	stSG8465	56	A	0.88	G	0.13	0.22	TCATGGGGCAAAGTRCTATGGGGCCAGACT
	stSG8466	111	G	0.94	A	0.06	0.12	GGTATTTGCACTACCRGTGAAGCAGCACAGCA
	stSG8656	44	C	0.94	T	0.06	0.12	ATGACCTTGATGCCGYGGAATTATATTCAGA
	stSG8880	28	C	0.94	T	0.06	0.12	CTGTACCCCGACGTYTCCCTGCTCGGCAC
50	stSG8904	35	G	0.88	A	0.13	0.22	TCACGCTGATCCAGCRGGCACCCTGCTTAAG
	stSG8917	64	G	0.75	A	0.25	0.38	GTAACATGACTAGARAGGCAGAGGAGTGGG
	stSG8944	30	C	0.44	T	0.56	0.49	TTGTAAGGATGTTTCYATAGAAATCACGGAT
	stSG8944	48	T	0.69	C	0.31	0.43	TAGAAATCACGGATAYATCACCAAGTCTACAG
	stSG8944	59	T	0.38	C	0.63	0.47	GATAGTATCACCAAGTYACAGCCACTATCTAT
55	stSG90	40	A	0.25	G	0.75	0.38	CGAGGAGTAGCCAGGRGGCGAGACACAAAAA
	stSG90	69	G	0.25	C	0.75	0.38	AAAGGCCTGGGACAGSTCAGTACAAGTCAGG
	stSG9044	67	G	0.56	A	0.44	0.49	TGACTGTAGAGGATTRATGATCCCTGAATAC
	stSG9062	83	C	0.38	G	0.63	0.47	GTACAGCAGGCTCTASCATTCTCTCTCTCT
	stSG9073	88	G	0.75	A	0.25	0.38	CTGGGCATGGCCGTGRCACCTGTGTGGCGA
60	stSG9075	65	C	0.94	T	0.06	0.12	GATTCTACAGCACCGCYGACACATAACATCA
	stSG9354	41	C	0.25	T	0.75	0.38	CCAGGGAGACACCTCYGTGAGATGACCTGCA
	stSG9535	42	C	0.75	T	0.25	0.38	CTCCCTGCCAGTCTTYCCGTCTAACCTCAG

	stSG9615	38	A	0.56	T	0.44	0.49	GGTTGGAATGTGTATWCAACTTGATGATGAA
	stSG9615	156	A	0.56	G	0.44	0.49	AATAAGTGTGTGAARGATTTTATTATAAAT
	stSG9673	82	A	0.88	G	0.13	0.22	TCCCTCTAATGAAGGRAAGGGTTTTTGAACA
5	stSG9757	195	G	0.94	T	0.06	0.12	TTAATCATTACTATTKCAACTCCGTATTTTC
	stD22S972E	20	A	0.56	G	0.44	0.49	TCCAGGAGCTGTTATRCTACCAGTTTCTGGC
	stSG10082	48	G	0.88	A	0.13	0.22	ACTGGCAGGGATTTGRTATCTAAACATAGAA
	stSG10082	58	A	0.75	C	0.25	0.38	ATTTGCGTATCTAAAMTAGAAAAGGTACAGT
	stSG1398	73	T	0.81	A	0.19	0.30	TTGCTTTTTATAATTWAAAGCAAATAACACA
10	stSG1437	71	G	0.25	T	0.75	0.38	AAGTTTTGACTTTGGKTCAGTTTTTATTAC
	stSG1446	106	T	0.50	A	0.50	0.50	TCAGCGTGAGATGATWTTGATTAACCTTGCT
	stSG1446	147	G	0.75	C	0.25	0.38	TAGTCAAACGTCGAAS TTGCTTGAGATGGCT
	stSG149	107	G	0.19	T	0.81	0.30	TTGGGGAGGAACCATKCTCNCTCTGGGCGCG
	stSG1514	78	T	0.81	G	0.19	0.30	TGGGTTTTCTGTGGGAKCAGCGGGGGCTTCT
15	stSG9800	134	C	0.50	A	0.50	0.50	TTAGTTTTGGATTAATMGACTTAAGAAAACAA
	stSG9828	32	G	0.88	A	0.13	0.22	ATTATGTGTTTCAGARTTATTA AAAAAGGCTA
	stSG9889	128	C	0.94	A	0.06	0.12	AGGAACTGAGAAAGAMCTGCCTAAGCAGCAC
	stSG9950	139	G	0.88	A	0.13	0.22	AAAATACTTGGTTAARTTGAAAGGACCTAGT
	stSG9961	33	A	0.19	G	0.81	0.30	TCTATTAGATAAAATRCAGATAAAGAATCTG
20	stSG9961	45	T	0.63	C	0.38	0.47	AATAACAGATAAAGAYCTGGAGAAAGGCTTF
	stVPREB	30	G	0.94	A	0.06	0.12	ATATTTCTCACAATCRACAAGAGCCAGGGCC
	stSG1615	57	T	0.58	C	0.42	0.49	GAGACATCCAGCCCA YTCTCTGGAACAGGAA
	stSG1615	79	T	0.75	C	0.25	0.38	GGAACAGGAAAGATGYCGGGGAGGGAACACA
	stSG1615	88	G	0.42	A	0.58	0.49	AAGATGATCGGGGAGRAACACAGGTCAGTNT
25	stSG1615	119	G	0.50	A	0.50	0.50	TGGGGACAGGGGTCARGTGGACACGGGGGTG
	stSG1813	41	C	0.50	T	0.50	0.50	GTGAGGGCCCAGGGTYTCCACGGAGAGGACA
	stSG1828	191	G	0.50	A	0.50	0.50	TGCTGTAGCCAAATTRTTGTCATACCAGGAA
	stSG2020	51	C	0.75	T	0.25	0.38	ATTAGAAAAGGACGCYCTGTGGCTGAACAA
	stSG2125	55	A	0.83	G	0.17	0.28	TTTACAAAATTTTCATRGAACGACAAATGTTA
30	stSG2294	139	T	0.92	C	0.08	0.15	AACACTGCAAAAACGYTCAAGTCAATAAAAAG
	stSG2314	89	T	0.75	A	0.25	0.38	ATGTCTTTTCCCAGTWGTCATATTTTTGTCC
	stSG2417	84	T	0.83	C	0.17	0.28	ACTCTTTATGACAAAYAGTGATTGANCTCTA
	stSG2482	121	A	0.08	T	0.92	0.15	TGANGCAGGCTATGGW TAAAAGAAACAACAA
	stSG2623	77	C	0.92	T	0.08	0.15	TTGTCTTTTTTTTTCYGGCAAAC TTTCTGCT
35	stSG2679	39	A	0.58	G	0.42	0.49	TACATTAATTTTCTTRTGAACACAGTAGACA
	stSG2773	49	C	0.83	T	0.17	0.28	ATATACACTGTTTATYTTTTTCTTTTTCACG
	stSG3009	88	C	0.92	T	0.08	0.15	TTACTTTTTATGTAGYTTAAGTGTGTTTATAA
	stSG3094	79	C	0.75	G	0.25	0.38	CTCCCAGAGTAAAASGTTTTTCTCTGGGNCT
	stSG3234	74	C	0.94	G	0.06	0.12	TCTNCAGATTGCACTSTAAGATCCTAGTTAC
40	stSG3248	40	A	0.38	G	0.63	0.47	ACATTCAGAATTATGRAAA CAATTAGTTATA
	stSG3277	43	A	0.75	G	0.25	0.38	TCATTTGCTACATGARCAGAGGCAGAGTATT
	stSG3349	141	T	0.31	A	0.69	0.43	CCTTTTAAAAAATGTWGAAATTTTAAAGTGGG
	stSG3388	28	T	0.94	C	0.06	0.12	AGTGAATTAGGGAGTYCTTGTGACCCCTTT
	stSG3552	40	G	0.56	A	0.44	0.49	AAAACCACATGTNCTRTAAGTGGGAGATAAA
45	stSG3809	87	T	0.44	C	0.56	0.49	AGTTACAGCCCCCTCYCACTCTGTATCTGC
	stSG3809	122	G	0.63	T	0.38	0.47	GGGTGGTGATGTGTTK GCTCCTAGACTCTCT
	stSG3809	123	G	0.88	C	0.13	0.22	GGTGGTGATGTGTTT SCTCCTAGACTCTCTC
	stSG3885	36	G	0.06	C	0.94	0.12	ATTTCTGACATTCATSCCAAAGANGGCAAAG
	stSG3927	84	A	0.94	C	0.06	0.12	ACAAAATAAACCGCTMGTTTTTCTGGCTCCA
50	stSG3927	118	T	0.0000	C	1.00	0.0000	CACGCCATATGAAGCYGCCAATGTCACTTAT
	stSG4025	41	G	0.88	A	0.13	0.22	ATCAACAGCTGCTACRTTCACCCCAGAGGTG
	stSG4044	22	A	0.44	G	0.56	0.49	TAATATGGGGGCTCTRAACACAGCACCCCCA
	stSG4085	30	A	0.94	C	0.06	0.12	GCCCCAGTGATTCTCMTACATTTTTACCTC
	stSG4085	97	C	0.69	T	0.31	0.43	TTTTCTTGCCTGGAGYTTCA TTGTTACCCCT
55	stSG4148	68	T	0.38	A	0.63	0.47	GATAAGCAGATCAGCWGCCAGCCAAGCTCAT
	stSG4389	52	G	0.38	T	0.63	0.47	GGCAGTATTTTAAAAKATTTCTCTAATGTTT
	stSG4494	71	T	0.94	C	0.06	0.12	ATTATTTCAAGTCATCYTAACATGTGACTTTA
	stSG4537	42	G	0.94	T	0.06	0.12	CCTCTGGCGAGCCCTKCGGCTCCACATCCTC
	stSG4702	124	C	0.94	T	0.06	0.12	ACAGTTGCTGACTCCYGGCTGTGGGCAGNC
60	stSG4978	102	C	0.44	G	0.56	0.49	AGAGAGGCATCACTGSGCTGCATCTGCCATG
	stSG6328	117	G	0.50	C	0.50	0.50	GCTTTAACAGAAACTSAACTCTTACAGCTTG

	stSG8971	95	T	0.88	C	0.13	0.22	AGCAATTTAATACAGYGAAAACAAATACAAT
	A002Q12	26	T	0.25	C	0.75	0.38	TCATTATTCTCTCCTYAGATTATTAATATT
	A002Q19	32	C	0.75	G	0.25	0.38	TCCTTCCCCTCCTGCSCCACTGCTGGCCA
5	A002Q20	138	T	0.88	C	0.13	0.22	GCCTGCATTTGGCTTYGTGCCTGAAAAAGAA
	A002S01	83	A	0.69	C	0.31	0.43	AAATGAAGATTTAATMTTCCTAAATTTAAGT
	A002T26	86	C	0.81	T	0.19	0.30	CGTCAAAAAGAAAACCCYCCCGGGACCCACTGT
	A002V42	50	T	0.06	C	0.94	0.12	TTTCATTCAGTTATAYCTTTGGCTCAGCTAG
	A002Y34	89	A	0.88	G	0.13	0.22	TAACAGAAAACGCCTTRGACACTATGTTGGG
10	A002Y45	85	C	0.75	A	0.25	0.38	GTGTGTCAGGATGCAMTGAAAGCCCTCGGCT
	A002Y45	106	G	0.38	C	0.63	0.47	AGCCCTCGGCTCGGTSSTAGCCAATCTTCCT
	A003B21	49	T	0.63	C	0.38	0.47	GACAACTTAAAACTTYTAGTGACATTGCTGT
	A003B21	120	T	0.63	A	0.38	0.47	TTAAAAGAGCAAAGTWCCCTCCCTTTCTTA
	A003B29	68	G	0.88	A	0.13	0.22	TTTTGGCCATAGACARTTATTTTTGATTCTAA
15	stSG9569	191	A	0.19	T	0.81	0.30	ATATGTATATATATAWTTTTTTAATTCCCTC
	stSG9574	43	T	0.81	G	0.19	0.30	TTGGGGCAAAAAGAGTKTCTTCATTATCAATC
	stSG9792	105	G	0.75	T	0.25	0.38	CTGGTGCCTGAGGCKTACACACCCGGCAGAA
	stSG9792	108	C	0.94	T	0.06	0.12	GTGCGCTGAGGCTGYACACCCGGCAGAACAG
	stSG9915	81	T	0.94	C	0.06	0.12	CAAACCCATTTAAGTYGGAATGATTTATATG
20	stSG9997	99	C	0.88	G	0.13	0.22	GCCCTAATAATCCAGSATTCTNACTCTCTT
	A004A22	125	G	0.94	A	0.06	0.12	TATCTGGCGAGGAGRCGGCATGGGATCCAG
	A004A30	135	G	0.94	C	0.06	0.12	GAATTTTAGATGCAGSARTCTTTTATATATA
	A004B17	146	T	0.25	C	0.75	0.38	ATTGCTGGGGCTCTAYTCCACAATTTGTTTT
	A004B36	107	A	0.94	G	0.06	0.12	TTGGAGTGCCTGGCRTCCCTCAGATTTGTC
25	A004B39	58	G	0.94	T	0.06	0.12	CCTCCCTCCAGACCKCTCCTTCTCCCTGCT
	A004F06	71	C	0.94	T	0.06	0.12	ATAATTTATACCACAYCTGAAGAAATTATCT
	A004F17	47	G	0.94	A	0.06	0.12	TATGGACTATGTACARACAATAACAAGAGGCG
	A004G25	85	C	0.94	T	0.06	0.12	AATCATTCTCTCTCYTTACATGGTGTACT
	A004H43	35	C	0.75	T	0.25	0.38	GGACTTCTAGCTTYAGCAAGCTTAGAGGA
30	A004H45	26	T	0.75	G	0.25	0.38	GAAGTTGCTATAGGTKCTTTCTCAAAAAGT
	A004I05	49	G	0.06	A	0.94	0.12	TAATGCCATTTTGATRITTAACATTACGTGTC
	A004I26	62	A	0.94	G	0.06	0.12	AAAAAGATTAAAAACRAAATAATATTAAGG
	A004I35	45	C	0.13	T	0.88	0.22	CCCCAGGTTAAACACYTGTAATTTACCTTGA
	A004I36	173	A	0.63	C	0.38	0.47	AGACAATGTCACTTGMAACACAAGGTATGAA
35	A004I36	190	T	0.31	C	0.69	0.43	AACACAAGGTATGAAAYATAAATAATAGTCAG
	A004M04	188	G	0.88	A	0.13	0.22	CTCCATTTTCCCTAARGCTGCCACTCTTGGG
	A004M43	78	C	0.81	A	0.19	0.30	GCAGTTTACTGTACMAGAAGTGCAATGCTA
	A004N13	110	A	0.63	G	0.38	0.47	TATCCTTTCTCTGCRGGGACTAAACAAGAA
	A004N44	65	G	0.88	A	0.13	0.22	ACTCATTTAGCAAAAARTCCTGAACACAATAT
40	A004P08	105	G	0.94	A	0.06	0.12	GGTACTTCCCTAGAGRGTCCTCCGAGGCTCAGA
	A004Q09	25	T	0.63	G	0.38	0.47	CAATGATAATACACCKTTGGATAAGGGGGAT
	A004Q11	40	T	0.56	C	0.44	0.49	TCAGCACAGGAATTGYAATCTTCTCACTTCA
	A004R33	68	C	0.94	T	0.06	0.12	CCCAACTACGATAAGYCATTGCCGGATGCTG
	A004R38	74	T	0.94	C	0.06	0.12	TTTTTCTGATATACTYCTGAAAATTTTATAA
45	A005C35	158	C	0.94	T	0.06	0.12	GGGGCCTTGTTCCYGCCATCGGACAGCTG
	A006N42	138	G	0.81	A	0.19	0.30	GTAAGTGAAGTGGGARAGGCAAGGTCTGCTA
	A006O23	37	G	0.94	A	0.06	0.12	GGGTGTGAGAAGCACRCAATAGGAAGTCTCT
	A006P16	33	T	0.88	C	0.13	0.22	TTGTTCAGGCTGATCYAAACTCTAGGCTCA
	A006P20	149	A	0.44	G	0.56	0.49	ATCCTTTCCCTGCTARAAAGACAAAACAAAA
50	A006Q32	19	G	0.13	A	0.88	0.22	TTCATTGGCATTAAAGRCATTACAATGCTGT
	A006Q32	84	G	0.81	A	0.19	0.30	TTTTCTTCATCGCTARAAGGAGTAATCCTTT
	A006Q33	86	C	0.94	A	0.06	0.12	TGTCCTTTCTCAATTMACAAATGCTGTATAA
	A006R10	61	T	0.88	C	0.13	0.22	TGTTCTGCTCATAAATYCCAATATGTACCAGA
	A006R44	78	A	0.38	G	0.63	0.47	GCCAACGTGCTGATCRGTGCCTGTCTGGAG
55	A006T39	130	G	0.88	C	0.13	0.22	TTTTTATCTGAAATSTTTTTAGAAGCCCTG
	A006U19	46	G	0.94	A	0.06	0.12	TACTGGATAACACTTRTTGGCCCATGACCTC
	A006U44	237	C	0.75	G	0.25	0.38	AGGACTATTTCCATGSATGTGTTATTGGCAG
	A006X15	172	A	0.81	G	0.19	0.30	GACTGCTGCCCCAGRCAGGCAGGGGGTGTG
	A006Y09	47	C	0.25	T	0.75	0.38	GGCTGAAACAGTGCCYAAGCTGGTCAGAGAT
60	A006Y32	176	G	0.19	A	0.81	0.30	ATTCTTTCTTACCRTAAAGGCTGTTCTTG
	A006Y36	72	T	0.69	C	0.31	0.43	TCCTTAATCTCAAAGYATTTTTAGTAATACA

	A007B18	156	A	0.88	T	0.13	0.22	TCCCACGGTGAATAWTACACACAATTACAC
	A007B24	62	G	0.38	T	0.63	0.47	GGAACAGAATGACAGKGGGATGCTGAGGAGC
	A007C36	22	A	0.94	G	0.06	0.12	TTACTGATATTCATTRATTATCCATAGGAC
5	A007C36	49	T	0.94	A	0.06	0.12	AGGACAGTTGTTTGAWTTGGTGCCACCTTAT
	A007C36	67	G	0.94	T	0.06	0.12	TGGTGCCACCTTATTKCCCCTTTATACAGAT
	A007D14	54	A	0.50	G	0.50	0.50	AAAGTTAAAAGGATARC GGTTACAGGAAAGT
	A007D35	53	G	0.81	C	0.19	0.30	ATGTCTTGAGAACATSAAATGAAATGGACAA
	A007E33	36	T	0.88	A	0.13	0.22	CACCTTCAAAAATTAWTGTGACTTACGGAAA
10	A007G47	40	A	0.94	G	0.06	0.12	TACCAGGCAAATAATRGTTACATCCCCAAACC
	A007H07	180	T	0.94	C	0.06	0.12	TGCCTACCATCTTCAYGGCCTCTGGGCACAA
	A007I32	134	T	0.94	C	0.06	0.12	GGGGCGCTCGGGGAGAYTGTGGACAATACCAA
	A007K44	103	T	0.88	C	0.13	0.22	TTTATTATTATTATTTTGAGATAAGGTCATG
	A007L07	150	T	0.69	G	0.31	0.43	CGCTGGGTGTGGGTTKATTCAGAGGCCACCA
15	A008B14	99	C	0.94	T	0.06	0.12	GATTCTACAGCACGCY GACACTAACACATCA
	A008B43	93	A	0.88	G	0.13	0.22	TGTGCCAACTCAAGGRGCTACCTTGACATTA
	A008C11	110	G	0.94	A	0.06	0.12	GCTCGTTCGTCAGGARTGGTGGTGAAGGCC
	A008C11	213	T	0.13	C	0.88	0.22	ATGGCGGTGGTGGCAYGGGAGCCTATGCCCC
	A008C18	57	A	0.88	G	0.13	0.22	TCTAGAAATGTCTAARAACAACCTTTTTTAT
20	stSG8656	44	C	0.94	T	0.06	0.12	ATGACCTTGATGCCGYGGAATTATATTCAGA
	stSG8880	28	C	0.94	T	0.06	0.12	CTGTACCCCGACGTYTCCCCTGCTCGGCAC
	stSG8904	35	G	0.88	A	0.13	0.22	TCACGCTGATCCAGCRGGCACCCCTGGCTTAAG
	stSG8917	64	G	0.75	A	0.25	0.38	GTAACTATGACTAGARAGGCAGAGGAGTGGG
	stSG8944	30	C	0.44	T	0.56	0.49	TTGTAAGGATGTTTCYATAGAAATCACGGAT
25	stSG8944	48	T	0.69	C	0.31	0.43	TAGAAATCACGGATAYATCACCAGTCTACAG
	stSG8944	59	T	0.38	C	0.63	0.47	GATAGTATCACCAGTYACAGCCACTATCTAT
	stSG90	40	A	0.25	G	0.75	0.38	CGAGGAGTAGCCAGGRGGCGAGACACAAAAA
	stSG90	69	G	0.25	C	0.75	0.38	AAAGGCCTGGGACAGSTCAGTACAAGTCAGG
	stSG9044	67	G	0.56	A	0.44	0.49	TGACTGTAGAGGATTRATGATCCCTGAATAC
	stSG9062	83	C	0.38	G	0.63	0.47	GTACAGCAGGCTCTASCATTCCTGTCTCTTT
30	stSG9073	88	G	0.75	A	0.25	0.38	CTGGGCATGGCCGTGRACCCCTGTCTGGCGA
	stSG9075	65	C	0.94	T	0.06	0.12	GATTCTACAGCACGCY GACACTAACACATCA
	stSG9354	41	C	0.25	T	0.75	0.38	CCAGGGAGACACCTCYGTGAGATGACCTGCA
	stSG9535	42	C	0.75	T	0.25	0.38	CTCCCTGCCAGTCTTYCCGTCTAACCCCTCAG
35	stSG9615	38	A	0.56	T	0.44	0.49	GGTTGGAATGTGTATWCAACTTGATGATGAA
	stSG9615	156	A	0.56	G	0.44	0.49	AATAAGTGTGTGAARGATTTTATTATAAAAT
	stSG9673	82	A	0.88	G	0.13	0.22	TCCCTCTAATGAAGGAAAGGGTTTTTGAACA
	stSG9757	195	G	0.94	T	0.06	0.12	TTAATCATTACTATTKCAACTCCGATTTTTC
	stD22S972E	20	A	0.56	G	0.44	0.49	TCCAGGAGCTGTTATRCTACCAGTTTCTGGC
40	stSG10082	48	G	0.88	A	0.13	0.22	ACTGGCAGGGATTTGRTATCTAAACATAGAA
	stSG10082	58	A	0.75	C	0.25	0.38	ATTTGCGTATCTAAAMTAGAAAAGGTACAGT
	stSG1398	73	T	0.81	A	0.19	0.30	TTGCTTTTATAATTWAAAGCAAATAACACA
	stSG1437	71	G	0.25	T	0.75	0.38	AAGTTTTGACTTTGGKCAAGTTTTTATTAC
	stSG1446	106	T	0.50	A	0.50	0.50	TCAGCGTGAGATGATWTTGATTAAACTTGCT
45	stSG1446	147	G	0.75	C	0.25	0.38	TAGTCAAACGTGCAASTTGCTTGAGATGGCT
	stSG149	107	G	0.19	T	0.81	0.30	TTGGGGAGGAACCATKCTCCNTCTGGGCCGC
	stSG1514	78	T	0.81	G	0.19	0.30	TGGGTTTTCTGTGGGAKCAGCGGGGGCCTCCT
	stSG9800	134	C	0.50	A	0.50	0.50	TTAGTTTTGGATTAATMGACTTAAGAAAACAA
	stSG9828	32	G	0.88	A	0.13	0.22	ATTATGTGTTTCAGARTTATTA AAAAGGCTA
50	stSG9889	128	C	0.94	A	0.06	0.12	AGGAACTGAGAAAAGAMCTGCCTAAGCAGCAC
	stSG9950	139	G	0.88	A	0.13	0.22	AAAATACTTGGTTAARTTGAAAGGACCTAGT
	stSG9961	33	A	0.19	G	0.81	0.30	TCTATTAGATAAAAATRCAGATAAAGAATCTG
	stSG9961	45	T	0.63	C	0.38	0.47	AATAACAGATAAAGAYCTGGAGAAAAGGCTTT
	stVPREB	30	G	0.94	A	0.06	0.12	ATATTTCTACAATCRACAAGAGCCAGGGCC
55	stSG1615	57	T	0.58	C	0.42	0.49	GAGACATCCAGCCAYTCTCTGGAACAGGAA
	stSG1615	79	T	0.75	C	0.25	0.38	GGAACAGGAAAGATGYCGGGGAGGGAACACA
	stSG1615	88	G	0.42	A	0.58	0.49	AAGATGATCGGGGAGRAACACAGGTCAGTNT
	stSG1615	119	G	0.50	A	0.50	0.50	TGGGGACAGGGGTCAARGTGGACACGGGGGTG
	stSG1813	41	C	0.50	T	0.50	0.50	GTGAGGGCCAGGGTYTCCACGGAGAGGACA
60	stSG1828	191	G	0.50	A	0.50	0.50	TGCTGTAGCCAAATTRTTGT CATACCAGGAA
	stSG2020	51	C	0.75	T	0.25	0.38	ATTAGAAAAGGACGCYCTGTTGGCTGAACAA

	stSG2125	55	A	0.83	G	0.17	0.28	TTTACAAAATTTTCATRGAACACTGACAATGTTA
	stSG2294	139	T	0.92	C	0.08	0.15	AACACTGCAAAAACCYTCAAGCATAAAAAAG
	stSG2314	89	T	0.75	A	0.25	0.38	ATGTCTTTTCCCAGTWGTCATATTTTGTCC
5	stSG2417	84	T	0.83	C	0.17	0.28	ACTCTCTTATGACAAYAGTGATTGANCTCTA
	stSG2482	121	A	0.08	T	0.92	0.15	TGANGCAGGCTATGGWTAAGAAAGAAACAACA
	stSG2623	77	C	0.92	T	0.08	0.15	TTGTCTTTTTTTTTTCYGGCAAACCTTCTGT
	stSG2679	39	A	0.58	G	0.42	0.49	TACATTAATTTTCTTRTGAACACAGTAGACA
	stSG2773	49	C	0.83	T	0.17	0.28	ATATACACTGTTTTATYTTTTTCTTTTTTACG
10	stSG3009	88	C	0.92	T	0.08	0.15	TTACTTTTTATGTAGYTAAGTGTGTTTATAA
	stSG3094	79	C	0.75	G	0.25	0.38	CTCCCCAGAGTAAAASGTTTTTCTCTGGNCT
	stSG3234	74	C	0.94	G	0.06	0.12	TCTNCAGATTGCACTSTAAGATCCTAGTTAC
	stSG3248	40	A	0.38	G	0.63	0.47	ACATTCAGAATTATGRAAAACAATTAGTTATA
	stSG3277	43	A	0.75	G	0.25	0.38	TCATTTGCTACATGARCAGGCGAGAGTATT
15	stSG3349	141	T	0.31	A	0.69	0.43	CCTTTTAAAAAATGTWGAATTTTAAGTGGG
	stSG3388	28	T	0.94	C	0.06	0.12	AGTGAATTAGGGAGTYCTTGTTGACCCCTT
	stSG3552	40	G	0.56	A	0.44	0.49	AAAACCACATGTNCTRTAAGTGGGAGATAAA
	stSG3809	87	T	0.44	C	0.56	0.49	AGTTACAGCCCCCTCYCACTCCTGTATCTGC
	stSG3809	122	G	0.63	T	0.38	0.47	GGGTGGTGATGTGTTKGCTCCTAGACTCTCT
20	stSG3809	123	G	0.88	C	0.13	0.22	GGTGGTGATGTGTTTSCCTCTAGACTCTCTC
	stSG3885	36	G	0.06	C	0.94	0.12	ATTTCTGACATTCATSCCAAAGNGCAAAG
	stSG3927	84	A	0.94	C	0.06	0.12	ACAAAATAAACCCGCTMGTTTTTCTGGTCCA
	stSG3927	118	T	0.00	C	1.00	0.00	CACGCCATATGAAGCYGCCAATGTCACTTAT
	stSG4025	41	G	0.88	A	0.13	0.22	ATCAACAGCTGCTACRTTCACCCAGAGGTG
25	stSG4044	22	A	0.44	G	0.56	0.49	TAATATGGGGGGTCTRAACACAGCACCCCA
	stSG4085	30	A	0.94	C	0.06	0.12	GCCCCAGTGATTCTCMTACATTTTTTACCTC
	stSG4085	97	C	0.69	T	0.31	0.43	TTTTCTTGCTGGAGYTCATTGTTACCCCT
	stSG4148	68	T	0.38	A	0.63	0.47	GATAAGCAGATCAGCWGCCAGCCAAGCTCAT
	stSG4389	52	G	0.38	T	0.63	0.47	GGCAGTATTTTAAAAKATTTCTCTAATGTTT
30	stSG4494	71	T	0.94	C	0.06	0.12	ATTATTTCACTCATCYTAACATGTGACTTTA
	stSG4537	42	G	0.94	T	0.06	0.12	CCTCTGGCGAGCCCTKCGGCTCCACATCCTC
	stSG4702	124	C	0.94	T	0.06	0.12	ACAGTTGCTGACTCCYGGCCTGTGGGCAGNC
	stSG4978	102	C	0.44	G	0.56	0.49	AGAGAGGCATCACTGSGCTGCATCTGCCATG
	stSG6328	117	G	0.50	C	0.50	0.50	GCTTTAACAGAAACTSAACTCTTACCGCTTG
35	stSG8971	95	T	0.88	C	0.13	0.22	AGCAATTTAATACAGYGA AAAACAATACAAT
	A002Q12	26	T	0.25	C	0.75	0.38	TCATTATTCTCTCTYAGATTATTTAAATTT
	A002Q19	32	C	0.75	G	0.25	0.38	TCCTTCCCCTCCTGCSACCATCTGGCCA
	A002Q20	138	T	0.88	C	0.13	0.22	GCTTGCATTTGGCTTYGTGCCTGAAAAGAA
	A002S01	83	A	0.69	C	0.31	0.43	AAATGAAGATTTAATMTTCTAAATTTAAGT
40	A002T26	86	C	0.81	T	0.19	0.30	CGTCAAAAAGAAAACCYCCC GGGACCCACTGT
	A002V42	50	T	0.06	C	0.94	0.12	TTTCATTCAGTTATAYCTTTGGCTCAGTAG
	A002Y34	89	A	0.88	G	0.13	0.22	TAACAGAAAACGCTTRGACACTATGTTGGG
	A002Y45	85	C	0.75	A	0.25	0.38	GTGTGTCAGGATGCAMTGAAGCCCTCGGCT
	A002Y45	106	G	0.38	C	0.63	0.47	AGCCCTCGGCTCGGTSTTAGCCAATCTTCT
45	A003B21	49	T	0.63	C	0.38	0.47	GACAACTTAAAACCTTYTAGTGACATTGCTGT
	A003B21	120	T	0.63	A	0.38	0.47	TTAAAAGAGCAAAAGTWCCCTCCCTTTCTTA
	A003B29	68	G	0.88	A	0.13	0.22	TTTTGGCCATAGACARTTATTTTGATTCTAA
	stSG9569	191	A	0.19	T	0.81	0.30	ATATGTATATATATAWTTTTTTTAAATCCTC
	stSG9574	43	T	0.81	G	0.19	0.30	TTGGGGCAAAAAGATKCTTTCATTATCAATC
50	stSG9792	105	G	0.75	T	0.25	0.38	CTGGTGCCTGAGGCKTACACACCCGGCAGAA
	stSG9792	108	C	0.94	T	0.06	0.12	GTGCGCTGAGGCTGYACACCCGGCAGAACAG
	stSG9915	81	T	0.94	C	0.06	0.12	CAAACCCATTTAAGTYGGAATGATTTATATG
	stSG9997	99	C	0.88	G	0.13	0.22	GCCCTAATAATCCAGSATCCTNACTCTCTT
	A004A22	125	G	0.94	A	0.06	0.12	TATCTGGCGAGGAGGRGCGCATGGAGTCCAG
55	A004A30	135	G	0.94	C	0.06	0.12	GAATTTTAGATGCAGSATCATTTTATATATA
	A004B17	146	T	0.25	C	0.75	0.38	ATTGCTGGGGCTCTAYTCCACAATTTGTTTT
	A004B36	107	A	0.94	G	0.06	0.12	TTGGAGTGCACAGGCRCTCCCTCAGATTTGTC
	A004B39	58	G	0.94	T	0.06	0.12	CCTCCCCTCCAGACKCTCCTTCTCCCTGT
	A004F06	71	C	0.94	T	0.06	0.12	ATAATTTATACCACAYCTGAAGAAATATCT
60	A004F17	47	G	0.94	A	0.06	0.12	TATGACTATGTACARACAATAACAAGAGCGG
	A004G25	85	C	0.94	T	0.06	0.12	AATCATTCTCTCTCYTTCACATGGTGTACT
	A004H43	35	C	0.75	T	0.25	0.38	GGACTTCTAGCCTTYAGCAAGCTTAGAGGA

	A004H45	26	T	0.75	G	0.25	0.38	GAAGTTGCTATAGGKCTCTTTCTCAAAGT
	A004I05	49	G	0.06	A	0.94	0.12	TAATGCCATTTTGTATRTTAACATTACGTGTC
	A004I26	62	A	0.94	G	0.06	0.12	AAAAAGATTAAAAACRAAATAATATTAAGG
5	A004I35	45	C	0.13	T	0.88	0.22	CCCCAGGTAAACACYTGTAAATTTACCTTGA
	A004I36	173	A	0.63	C	0.38	0.47	AGACAATGTCACTTGMAACACAAGGTATGAA
	A004I36	190	T	0.31	C	0.69	0.43	AACACAAGGTATGAAAYATAAATAATAGTCAG
	A004M04	188	G	0.88	A	0.13	0.22	CTCCATTTTCCCTAARGCTGCCACTCTTGGG
	A004M43	78	C	0.81	A	0.19	0.30	GCAGTTTTACTGTACMAGAAGTGCAATGCTA
10	A004N13	110	A	0.63	G	0.38	0.47	TATCCTTTCCTCTGCRGGGACTAAACAAGAA
	A004N44	65	G	0.88	A	0.13	0.22	ACTCATTTAGCAAAARTCTGAACACAATAT
	A004P08	105	G	0.94	A	0.06	0.12	GGTACTTCCCTAGARGTCCCAGGACTCAGA
	A004Q09	25	T	0.63	G	0.38	0.47	CAATGATAATACACCKTTGGATAAGGGGGAT
	A004Q11	40	T	0.56	C	0.44	0.49	TCAGCACAGGAATTGYAATCTTCTCACTTCA
15	A004R33	68	C	0.94	T	0.06	0.12	CCCAACTACGATAAGYCATTGCCGGATGCTG
	A004R38	74	T	0.94	C	0.06	0.12	TTTTTCTGATATACTYCTGAAAATTTTATAA
	A005C35	158	C	0.94	T	0.06	0.12	GGGGCCTTGTGTTCCYGCCATCGGACAGCTG
	A006N42	138	G	0.81	A	0.19	0.30	GTACTGGAAGTGGGARAGGAAGGTCTGCTA
	A006O23	37	G	0.94	A	0.06	0.12	GGTGTGAGAAGCACRCAATAGGAAGTCTCT
20	A006P16	33	T	0.88	C	0.13	0.22	TTGTTCAAGGCTGATCYAAACTCCTAGGCTCA
	A006P20	149	A	0.44	G	0.56	0.49	ATCCTTTCCTGCTARAAAGACAAAAA
	A006Q32	19	G	0.13	A	0.88	0.22	TCATTGGCATTAAAGRCATTACAATGCTGT
	A006Q32	84	G	0.81	A	0.19	0.30	TTTTCTTCATCGCTARAAGGAGTAATCCTT
	A006Q33	86	C	0.94	A	0.06	0.12	TGTCCTTCTCAATTMACAAATGCTGTTAAA
25	A006R10	61	T	0.88	C	0.13	0.22	TGTTCTGCTCATAATYCCAATGTACCAGA
	A006R44	78	A	0.38	G	0.63	0.47	GCCAACGTGCTGATCRGCTGCTGCTGGAG
	A006T39	130	G	0.88	C	0.13	0.22	TTTTATCCTGAAATSTTTTTAGAAGCCCTG
	A006U19	46	G	0.94	A	0.06	0.12	TACTGGATAACACTTRTTGGCCCATGACCTC
	A006U44	237	C	0.75	G	0.25	0.38	AGGACTATTTCCATGSATGTGTTATTGGCAG
30	A006X15	172	A	0.81	G	0.19	0.30	GACTGCTGCCCCAGRCAGGCAGGGGGTGTG
	A006Y09	47	C	0.25	T	0.75	0.38	GGTGAAACAGTGCCYAAGCTGGTCAGAGAT
	A006Y32	176	G	0.19	A	0.81	0.30	ATTCTTTCTTCAACRTAAAGGTGTGTTCTG
	A006Y36	72	T	0.69	C	0.31	0.43	TCCTTAATCTCAAAGYATTTTTAGTATACA
	A007B18	156	A	0.88	T	0.13	0.22	TCCCACGGTGGAATAWTACACACAATTACAC
35	A007B24	62	G	0.38	T	0.63	0.47	GGAACAGAATGACAGKGGGATGCTGAGGAGC
	A007C36	22	A	0.94	G	0.06	0.12	TTACTGATATTCATTRATTATTCCATAGGAC
	A007C36	49	T	0.94	A	0.06	0.12	AGGACAGTTGTTGAWTTGGTGCCACCTTAT
	A007C36	67	G	0.94	T	0.06	0.12	TGGTGCCACCTTATTKCCCCTTTATACAGAT
	A007D14	54	A	0.50	G	0.50	0.50	AAAGTTAAAAGGATARCGGTTACAGGAAAGT
40	A007D35	53	G	0.81	C	0.19	0.30	ATGTCTTGAGAACATSAATGAATTGGACAA
	A007E33	36	T	0.88	A	0.13	0.22	CACCTTCAAAAATTAWTGTGACTTACGGAAA
	A007G47	40	A	0.94	G	0.06	0.12	TACCAGGCAAATAATRGATACATCCCCAAACC
	A007H07	180	T	0.94	C	0.06	0.12	TGCCTACCATCTTCAYGGCCTCTGGGCACAA
	A007I32	134	T	0.94	C	0.06	0.12	GGGGCGCTCGGGAGAYTGTGGACAATACCAA
45	A007K44	103	T	0.88	C	0.13	0.22	TTTATTATTATTATTYTGAGATAAGGTCATG
	A007L07	150	T	0.69	G	0.31	0.43	CGCTGGGTGTGGGTTKATTCAGAGGCCACA
	A008B14	99	C	0.94	T	0.06	0.12	GATTCTACAGCACGCGYACACTAACACATCA
	A008B43	93	A	0.88	G	0.13	0.22	TGTGCCAACTCAAGGRGCTACCTTGACATTA
	A008C11	110	G	0.94	A	0.06	0.12	GCTCGTCTGCAGGARTGGTGGTGAAGGCC
50	A008C11	213	T	0.13	C	0.88	0.22	ATGGCGGTGGTGGCAYGGGAGCCTATGCCCC
	A008C18	57	A	0.88	G	0.13	0.22	TCTAGAAATGTCTAARAACAACCTTTTTAT

Fragments prefaced stSG are from the Sanger Centre, UK.

Fragments prefaced with A are from Genethon, France.

Fragments without a prefix are from the Whitehead

55 Institute.

## Analysis of Polymorphisms

### A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. See generally *PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., *Nucleic Acids Res.* 19, 4967 (1991); Eckert et al., *PCR Methods and Applications* 1, 17 (1991); PCR (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202 (each of which is incorporated by reference for all purposes).

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren et al., *Science* 241, 1077 (1988), transcription amplification (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli et al., *Proc. Nat. Acad. Sci. USA*, 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based

on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

5

B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis depending whether a polymorphism in question has already been characterized. The first type of analysis is sometimes referred to as de novo characterization. This analysis compares target sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing a groups of individuals representing the greatest ethnic diversity among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified, and the frequencies of such populations in the population determined. Additional allelic frequencies can be determined for subpopulations characterized by criteria such as geography, race, or gender. The de novo identification of the polymorphisms of the invention is described in the Examples section. The second type of analysis is determining which form(s) of a characterized polymorphism are present in individuals under test. There are a variety of suitable procedures, which are discussed in turn.

10

15

20

25

1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki et al., *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions

30

35

should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles.

5 Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15 mer at the 7 position; in a 16 mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in  
10 hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs  
15 of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

## 2. Tiling Arrays

20 The polymorphisms can also be identified by hybridization to nucleic acid arrays, some example of which are described by WO 95/11995 (incorporated by reference in its entirety for all purposes). One form of such arrays is described in the Examples section in  
25 connection with de novo identification of polymorphisms. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of a variant forms of a precharacterized polymorphism.  
30 Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first reference sequence. The second group of probes is designed by the same principles as described in the Examples except that the probes  
35 exhibit complementarily to the second reference sequence. The inclusion of a second group (or further groups) can be particular useful for analyzing short subsequences of

the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (*i.e.*, two or more mutations within 9 to 21 bases).

5

### 3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers leading to a detectable product signifying the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer. See, *e.g.*, WO 93/22456.

10  
15  
20  
25

### 4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind et al., *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

30  
35

### 5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W.H. Freeman and Co, New York, 1992), Chapter 7.

### 6. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita et al., *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence difference between alleles of target sequences.

### III. Methods of Use

After determining polymorphic form(s) present in an individual at one or more polymorphic sites, this information can be used in a number of methods.

#### A. Forensics

Determination of which polymorphic forms occupy a set of polymorphic sites in an individual identifies a

set of polymorphic forms that distinguishes the individual. See generally National Research Council, *The Evaluation of Forensic DNA Evidence* (Eds. Pollard et al., National Academy Press, DC, 1996). The more sites that are analyzed the lower the probability that the set of polymorphic forms in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked. Thus, polymorphisms of the invention are often used in conjunction with polymorphisms in distal genes. Preferred polymorphisms for use in forensics are diallelic because the population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci.

The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (e.g., by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance.

$p(\text{ID})$  is the probability that two random individuals have the same polymorphic or allelic form at a given polymorphic site. In diallelic loci, four

genotypes are possible: AA, AB, BA, and BB. If alleles A and B occur in a haploid genome of the organism with frequencies  $x$  and  $y$ , the probability of each genotype in a diploid organism are (see WO 95/12607):

- 5  
 Homozygote:  $p(AA) = x^2$   
 Homozygote:  $p(BB) = y^2 = (1-x)^2$   
 Single Heterozygote:  $p(AB) = p(BA) = xy = x(1-x)$   
 Both Heterozygotes:  $p(AB+BA) = 2xy = 2x(1-x)$

10           The probability of identity at one locus (i.e., the probability that two individuals, picked at random from a population will have identical polymorphic forms at a given locus) is given by the equation:

15           
$$p(ID) = (x^2)^2 + (2xy)^2 + (y^2)^2.$$

          These calculations can be extended for any number of polymorphic forms at a given locus. For example, the probability of identity  $p(ID)$  for a 3-allele system where the alleles have the frequencies in the population of  $x$ ,  $y$  and  $z$ , respectively, is equal to the sum of the squares of the genotype frequencies:

20           
$$p(ID) = x^4 + (2xy)^2 + (2yz)^2 + (2xz)^2 + z^4 + y^4$$

          In a locus of  $n$  alleles, the appropriate binomial expansion is used to calculate  $p(ID)$  and  $p(exc)$ .

25           The cumulative probability of identity (cum  $p(ID)$ ) for each of multiple unlinked loci is determined by multiplying the probabilities provided by each locus.

$$\text{cum } p(ID) = p(ID_1)p(ID_2)p(ID_3)\dots p(ID_n)$$

30           The cumulative probability of non-identity for  $n$  loci (i.e. the probability that two random individuals will be different at 1 or more loci) is given by the equation:

$$\text{cum } p(\text{nonID}) = 1 - \text{cum } p(ID).$$

35           If several polymorphic loci are tested, the cumulative probability of non-identity for random individuals becomes very high (e.g., one billion to one). Such probabilities can be taken into account together with other evidence in determining the guilt or innocence of the suspect.

B. Paternity Testing

The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child.

If the set of polymorphisms in the child attributable to the father does not match the putative father, it can be concluded, barring experimental error, that the putative father is not the real father. If the set of polymorphisms in the child attributable to the father does match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match.

The probability of parentage exclusion (representing the probability that a random male will have a polymorphic form at a given polymorphic site that makes him incompatible as the father) is given by the equation (see WO 95/12607):

$$p(\text{exc}) = xy(1-xy)$$

where  $x$  and  $y$  are the population frequencies of alleles A and B of a diallelic polymorphic site.

(At a triallelic site  $p(\text{exc}) = xy(1-xy) + yz(1-yz) + xz(1-xz) + 3xyz(1-xyz)$ ), where  $x$ ,  $y$  and  $z$  are the respective population frequencies of alleles A, B and C).

The probability of non-exclusion is

$$p(\text{non-exc}) = 1-p(\text{exc})$$

The cumulative probability of non-exclusion (representing the value obtained when  $n$  loci are used) is thus:

$$\text{cum } p(\text{non-exc}) = p(\text{non-exc}_1)p(\text{non-exc}_2)p(\text{non-exc}_3)\dots p(\text{non-exc}_n)$$

The cumulative probability of exclusion for  $n$  loci (representing the probability that a random male will be excluded)

$$\text{cum } p(\text{exc}) = 1 - \text{cum } p(\text{non-exc}).$$

5           If several polymorphic loci are included in the analysis, the cumulative probability of exclusion of a random male is very high. This probability can be taken into account in assessing the liability of a putative father whose polymorphic marker set matches the child's  
10 polymorphic marker set attributable to his/her father.

C. Correlation of Polymorphisms with Phenotypic Traits

15           The polymorphisms of the invention may contribute to the phenotype of an organism in different ways. Some polymorphisms occur within a protein coding sequence and contribute to phenotype by affecting protein structure. The effect may be neutral, beneficial or detrimental, or both beneficial and detrimental, depending on the  
20 circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on  
25 replication, transcription, and translation. A single polymorphism may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by polymorphisms in different genes. Further, some polymorphisms predispose an individual to a distinct  
30 mutation that is causally related to a certain phenotype.

Phenotypic traits include diseases that have known but hitherto unmapped genetic components (e.g., agammaglobulinemia, diabetes insipidus, Lesch-Nyhan  
35 syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial hypercholesterolemia, polycystic kidney disease, hereditary spherocytosis, von

Willebrand's disease, tuberous sclerosis, hereditary hemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, and acute intermittent porphyria). Phenotypic traits also include symptoms of, or susceptibility to, multifactorial diseases of which a component is or may be genetic, such as autoimmune diseases, inflammation, cancer, diseases of the nervous system, and infection by pathogenic microorganisms. Some examples of autoimmune diseases include rheumatoid arthritis, multiple sclerosis, diabetes (insulin-dependent and non-independent), systemic lupus erythematosus and Graves disease. Some examples of cancers include cancers of the bladder, brain, breast, colon, esophagus, kidney, leukemia, liver, lung, oral cavity, ovary, pancreas, prostate, skin, stomach and uterus. Phenotypic traits also include characteristics such as longevity, appearance (e.g., baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments.

Correlation is performed for a population of individuals who have been tested for the presence or absence of a phenotypic trait of interest and for polymorphic markers sets. To perform such analysis, the presence or absence of a set of polymorphisms (i.e. a polymorphic set) is determined for a set of the individuals, some of whom exhibit a particular trait, and some of which exhibit lack of the trait. The alleles of each polymorphism of the set are then reviewed to determine whether the presence or absence of a particular allele is associated with the trait of interest. Correlation can be performed by standard statistical methods such as a  $\chi$ -squared test and statistically significant correlations between polymorphic form(s) and phenotypic characteristics are noted. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart disease. As a

further example, it might be found that the combined presence of allele A1 at polymorphism A and allele B1 at polymorphism B correlates with increased milk production of a farm animal.

5           Such correlations can be exploited in several ways. In the case of a strong correlation between a set of one or more polymorphic forms and a disease for which treatment is available, detection of the polymorphic form set in a human or animal patient may justify immediate  
10 administration of treatment, or at least the institution of regular monitoring of the patient. Detection of a polymorphic form correlated with serious disease in a couple contemplating a family may also be valuable to the couple in their reproductive decisions. For example, the  
15 female partner might elect to undergo in vitro fertilization to avoid the possibility of transmitting such a polymorphism from her husband to her offspring. In the case of a weaker, but still statistically significant correlation between a polymorphic set and  
20 human disease, immediate therapeutic intervention or monitoring may not be justified. Nevertheless, the patient can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little cost to the patient but confer potential  
25 benefits in reducing the risk of conditions to which the patient may have increased susceptibility by virtue of variant alleles. Identification of a polymorphic set in a patient correlated with enhanced receptiveness to one of several treatment regimes for a disease indicates that  
30 this treatment regime should be followed.

          For animals and plants, correlations between characteristics and phenotype are useful for breeding for desired characteristics. For example, Beitz et al., US  
5,292,639 discuss use of bovine mitochondrial  
35 polymorphisms in a breeding program to improve milk production in cows. To evaluate the effect of mtDNA D-loop sequence polymorphism on milk production, each cow

was assigned a value of 1 if variant or 0 if wildtype with respect to a prototypical mitochondrial DNA sequence at each of 17 locations considered. Each production trait was analyzed individually with the following animal model:

$$Y_{ijkpn} = \mu + YS_i + P_j + X_k + \beta_1 + \dots + \beta_{17} + PE_n + a_n + e_p$$

where  $Y_{ijkpn}$  is the milk, fat, fat percentage, SNF, SNF percentage, energy concentration, or lactation energy record;  $\mu$  is an overall mean;  $YS_i$  is the effect common to all cows calving in year-season;  $X_k$  is the effect common to cows in either the high or average selection line;  $\beta_1$  to  $\beta_{17}$  are the binomial regressions of production record on mtDNA D-loop sequence polymorphisms;  $PE_n$  is permanent environmental effect common to all records of cow  $n$ ;  $a_n$  is effect of animal  $n$  and is composed of the additive genetic contribution of sire and dam breeding values and a Mendelian sampling effect; and  $e_p$  is a random residual. It was found that eleven of seventeen polymorphisms tested influenced at least one production trait. Bovines having the best polymorphic forms for milk production at these eleven loci are used as parents for breeding the next generation of the herd.

#### D. Genetic Mapping of Phenotypic Traits

The previous section concerns identifying correlations between phenotypic traits and polymorphisms that directly or indirectly contribute to those traits. The present section describes identification of a physical linkage between a genetic locus associated with a trait of interest and polymorphic markers that are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Such analysis is useful for mapping a genetic locus associated with a phenotypic trait to a chromosomal position, and thereby cloning gene(s) responsible for the trait. See Lander et al., *Proc. Natl. Acad. Sci. (USA)* 83, 7353-7357 (1986); Lander et

al., *Proc. Natl. Acad. Sci. (USA)* 84, 2363-2367 (1987); Donis-Keller et al., *Cell* 51, 319-337 (1987); Lander et al., *Genetics* 121, 185-199 (1989)). Genes localized by linkage can be cloned by a process known as directional  
5 cloning. See Wainwright, *Med. J. Australia* 159, 170-174 (1993); Collins, *Nature Genetics* 1, 3-6 (1992) (each of which is incorporated by reference in its entirety for all purposes).

Linkage studies are typically performed on members  
10 of a family. Available members of the family are characterized for the presence or absence of a phenotypic trait and for a set of polymorphic markers. The distribution of polymorphic markers in an informative meiosis is then analyzed to determine which polymorphic  
15 markers co-segregate with a phenotypic trait. See, e.g., Kerem et al., *Science* 245, 1073-1080 (1989); Monaco et al., *Nature* 316, 842 (1985); Yamoka et al., *Neurology* 40, 222-226 (1990); Rossiter et al., *FASEB Journal* 5, 21-27 (1991).

Linkage is analyzed by calculation of LOD (log of  
20 the odds) values. A lod value is the relative likelihood of obtaining observed segregation data for a marker and a genetic locus when the two are located at a recombination fraction  $\theta$ , versus the situation in which the two are not  
25 linked, and thus segregating independently (Thompson & Thompson, *Genetics in Medicine* (5th ed, W.B. Saunders Company, Philadelphia, 1991); Strachan, "Mapping the human genome" in *The Human Genome* (BIOS Scientific Publishers Ltd, Oxford), Chapter 4). A series of  
30 likelihood ratios are calculated at various recombination fractions ( $\theta$ ), ranging from  $\theta = 0.0$  (coincident loci) to  $\theta = 0.50$  (unlinked). Thus, the likelihood at a given value of  $\theta$  is: probability of data if loci linked at  $\theta$  to probability of data if loci unlinked. The computed  
35 likelihoods are usually expressed as the  $\log_{10}$  of this

ratio (i.e., a lod score). For example, a lod score of 3 indicates 1000:1 odds against an apparent observed linkage being a coincidence. The use of logarithms allows data collected from different families to be combined by simple addition. Computer programs are available for the calculation of lod scores for differing values of  $\theta$  (e.g., LIPED, MLINK (Lathrop, *Proc. Nat. Acad. Sci. (USA)* 81, 3443-3446 (1984)). For any particular lod score, a recombination fraction may be determined from mathematical tables. See Smith et al., *Mathematical tables for research workers in human genetics* (Churchill, London, 1961); Smith, *Ann. Hum. Genet.* 32, 127-150 (1968). The value of  $\theta$  at which the lod score is the highest is considered to be the best estimate of the recombination fraction.

Positive lod score values suggest that the two loci are linked, whereas negative values suggest that linkage is less likely (at that value of  $\theta$ ) than the possibility that the two loci are unlinked. By convention, a combined lod score of +3 or greater (equivalent to greater than 1000:1 odds in favor of linkage) is considered definitive evidence that two loci are linked. Similarly, by convention, a negative lod score of -2 or less is taken as definitive evidence against linkage of the two loci being compared. Negative linkage data are useful in excluding a chromosome or a segment thereof from consideration. The search focuses on the remaining non-excluded chromosomal locations.

#### 30 IV. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described in Table 1, column 8, in which the polymorphic position is occupied by one of the alternative bases for that position. Some nucleic acid encode full-length variant forms of

proteins. Similarly, variant proteins have the prototypical amino acid sequences of encoded by nucleic acid sequence shown in Table 1, column 8, (read so as to be in-frame with the full-length coding sequence of which it is a component) except at an amino acid encoded by a codon including one of the polymorphic positions shown in the Table. That position is occupied by the amino acid coded by the corresponding codon in any of the alternative forms shown in the Table.

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example *trp*, *lac*, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, e.g., mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide.

Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

5 The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles and Practice*, 2nd Edition, Springer-Verlag, New York (1987); 10 and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the 15 protein can be isolated from a lysate of the host cells.

The invention further provides transgenic nonhuman animals capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Expression of an exogenous 20 variant gene is usually achieved by operably linking the gene to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. See Hogan et al., "Manipulating the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory. Inactivation of 25 endogenous variant genes can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a positive selection marker. See Capecchi, *Science* 244, 1288-1292 (1989). The transgene is then introduced into an embryonic stem cell, where it 30 undergoes homologous recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

In addition to substantially full-length 35 polypeptides expressed by variant genes, the present

invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any  
5 portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or  
10 large cellular structures.

Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or  
15 other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York  
20 (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays  
25 for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

#### V. Kits

The invention further provides kits comprising at  
30 least one allele-specific oligonucleotide as described above. Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a  
35 substrate. For example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at

least 10, 100 or all of the polymorphisms shown in Table 1. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

#### VI. Computer Systems For Storing Polymorphism Data

Fig. 1A depicts a block diagram of a computer system 10 suitable for implementing the present invention. Computer system 10 includes a bus 12 which interconnects major subsystems such as a central processor 14, a system memory 16 (typically RAM), an input/output (I/O) controller 18, an external device such as a display screen 24 via a display adapter 26, serial ports 28 and 30, a keyboard 32, a fixed disk drive 34 via a storage interface 35 and a floppy disk drive 36 operative to receive a floppy disk 38, and a CD-ROM (or DVD-ROM) device 40 operative to receive a CD-ROM 42. Many other devices can be connected such as a user pointing device, e.g., a mouse 44 connected via serial port 28 and a network interface 46 connected via serial port 30.

Many other devices or subsystems (not shown) may be connected in a similar manner. Also, it is not necessary for all of the devices shown in Fig. 1A to be present to practice the present invention, as discussed below. The devices and subsystems may be interconnected in different ways from that shown in Fig. 1A. The operation of a computer system such as that shown in Fig. 1A is well known. Databases storing polymorphism information according to the present invention can be stored, e.g., in system memory 16 or on storage media

such as fixed disk 34, floppy disk 38, or CD-ROM 42. An application program to access such databases can be operably disposed in system memory 16 or sorted on storage media such as fixed disk 34, floppy disk 38, or CD-ROM 42.

Fig. 1B depicts the interconnection of computer system 10 to remote computers 48, 50, and 52. Fig. 1B depicts a network 54 interconnecting remote servers 48, 50, and 52. Network interface 46 provides the connection from client computer system 10 to network 54. Network 54 can be, e.g., the Internet. Protocols for exchanging data via the Internet and other networks are well known. Information identifying the polymorphisms described herein can be transmitted across network 54 embedded in signals capable of traversing the physical media employed by network 54.

Information identifying polymorphisms shown in Table 1 is represented in records, which optionally, are subdivided into fields. Each record stores information relating to a different polymorphisms in Table 1. Collectively, the records can store information relating to all of the polymorphisms in Table 1, or any subset thereof, such as 5, 10, 50, or 100 polymorphisms from Table 1. In some databases, the information identifies a base occupying a polymorphic position and the location of the polymorphic position. The base can be represented as a single letter code (i.e., A, C, G or T/U) present in a polymorphic form other than that in the reference allele. Alternatively, the base occupying a polymorphic site can be represented in IUPAC ambiguity code as shown in Table 1. The location of a polymorphic site can be identified as its position within one of the sequences shown in Table 1. For example, in the first sequence shown in Table 1, the polymorphic site occupies the 16th base. The position can also be identified by reference to, for example, a chromosome, and distance from known markers within the chromosome. In other databases, information

identifying a polymorphism contains sequences of 10-100  
bases shown in Table 1 or the complements thereof,  
including a polymorphic site. Preferably, such  
information records at least 10, 15, 20, or 30 contiguous  
5 bases of sequences including a polymorphic site.

#### EXAMPLES

The polymorphisms shown in Table 1 were identified  
10 by resequencing of target sequences from eight unrelated  
individuals of diverse ethnic and geographic backgrounds  
by hybridization to probes immobilized to microfabricated  
arrays. The strategy and principles for design and use  
of such arrays are generally described in WO 95/11995.  
15 The strategy provides arrays of probes for analysis of  
target sequences showing a high degree of sequence  
identity to the reference sequences of the fragments  
shown in Table 1, column 1. The reference sequences  
were sequence-tagged sites (STSs) developed in the course  
20 of the Human Genome Project (see, e.g., *Science* 270,  
1945-1954 (1995); *Nature* 380, 152-154 (1996)). Most  
STS's ranged from 100 bp to 300 bp in size.

A typical probe array used in this analysis has  
two groups of four sets of probes that respectively tile  
25 both strands of a reference sequence. A first probe set  
comprises a plurality of probes exhibiting perfect  
complementarity with one of the reference sequences.  
Each probe in the first probe set has an interrogation  
position that corresponds to a nucleotide in the  
30 reference sequence. That is, the interrogation position  
is aligned with the corresponding nucleotide in the  
reference sequence, when the probe and reference sequence  
are aligned to maximize complementarity between the two.  
For each probe in the first set, there are three  
35 corresponding probes from three additional probe sets.  
Thus, there are four probes corresponding to each

nucleotide in the reference sequence. The probes from the three additional probe sets are identical to the corresponding probe from the first probe set except at the interrogation position, which occurs in the same  
5 position in each of the four corresponding probes from the four probe sets, and is occupied by a different nucleotide in the four probe sets. In the present analysis, probes were 25 nucleotides long. Arrays tiled for multiple different references sequences were included  
10 on the same substrate.

Multiple target sequences from an individual were amplified from human genomic DNA using primers for the fragments indicated in the listed Web sites. The amplified target sequences were fluorescently labelled  
15 during or after PCR. The labelled target sequences were hybridized with a substrate bearing immobilized arrays of probes. The amount of label bound to probes was measured. Analysis of the pattern of label revealed the nature and position of differences between the target and  
20 reference sequence. For example, comparison of the intensities of four corresponding probes reveals the identity of a corresponding nucleotide in the target sequences aligned with the interrogation position of the probes. The corresponding nucleotide is the complement  
25 of the nucleotide occupying the interrogation position of the probe showing the highest intensity (see WO 95/11995). The existence of a polymorphism is also manifested by differences in normalized hybridization intensities of probes flanking the polymorphism when the  
30 probes hybridized to corresponding targets from different individuals. For example, relative loss of hybridization intensity in a "footprint" of probes flanking a polymorphism signals a difference between the target and reference (i.e., a polymorphism) (see EP 717,113,  
35 incorporated by reference in its entirety for all purposes). Additionally, hybridization intensities for corresponding targets from different individuals can be

classified into groups or clusters suggested by the data,  
not defined *a priori*, such that isolates in a give  
cluster tend to be similar and isolates in different  
clusters tend to be dissimilar. See WO 97/29212  
5 (incorporated by reference in its entirety for all  
purposes). Hybridizations to samples from different  
individuals were performed separately. Table 1  
summarizes the data obtained for target sequences in  
comparison with a reference sequence for the eight  
10 individuals tested.

From the foregoing, it is apparent that the  
invention includes a number of general uses that can be  
expressed concisely as follows. The invention provides  
for the use of any of the nucleic acid segments described  
15 above in the diagnosis or monitoring of diseases, such as  
cancer, inflammation, heart disease, diseases of the CNS,  
and susceptibility to infection by microorganisms. The  
invention further provides for the use of any of the  
nucleic acid segments in the manufacture of a medicament  
20 for the treatment or prophylaxis of such diseases. The  
invention further provides for the use of any of the DNA  
segments as a pharmaceutical.

All publications and patent applications cited  
above are incorporated by reference in their entirety for  
25 all purposes to the same extent as if each individual  
publication or patent application were specifically and  
individually indicated to be so incorporated by  
reference. Although the present invention has been  
described in some detail by way of illustration and  
30 example for purposes of clarity and understanding, it  
will be apparent that certain changes and modifications  
may be practiced within the scope of the appended claims.

WHAT IS CLAIMED IS:

1           1       A nucleic acid segment of between 10 and 100  
2 bases from a fragment shown in Table 1 including a  
3 polymorphic site, or the complement of the segment.

1           2.       The nucleic acid segment of claim 1 that is  
2 DNA.

1           3.       The nucleic acid segment of claim 1 that is  
2 RNA.

1           4.       The segment of claim 1 that is less than 50  
2 bases.

1           5.       The segment of claim 1 that is less than 20  
2 bases.

1           6.       The segment of claim 1, wherein the fragment  
2 is WI-14263 and the polymorphic site is at position 49.

1           7.       The segment of claim 1, wherein the  
2 polymorphic site is diallelic.

1           8.       The segment of claim 1, wherein the  
2 polymorphic form occupying the polymorphic site is the  
3 reference base for the fragment listed in Table 1, column  
4 3.

1           9.       The segment of claim 1, wherein the  
2 polymorphic form occupying the polymorphic site is an  
3 alternative form for the fragment listed in Table 1,  
4 column 5.

1           10.       An allele-specific oligonucleotide that  
2 hybridizes to a segment of a fragment shown in Table 1,  
3 column 8 or its complement.

1           11. The allele-specific oligonucleotide of claim  
2 10 that is probe.

1           12. The allele-specific oligonucleotide of claim  
2 10, wherein a central position of the probe aligns with  
3 the polymorphic site of the fragment.

1           13. The allele-specific oligonucleotide of claim  
2 10 that is a primer.

1           14. The allele-specific oligonucleotide of claim  
2 13, wherein the 3' end of the primer aligns with the  
3 polymorphic site of the fragment.

1           15. An isolated nucleic acid comprising a  
2 sequence of Table 1, column 8 or the complement thereof,  
3 wherein the polymorphic site within the sequence or  
4 complement is occupied by a base other than the reference  
5 base show in Table 1, column 3.

1           16. A method of analyzing a nucleic acid,  
2 comprising:  
3       obtaining the nucleic acid from an individual; and  
4       determining a base occupying any one of the polymorphic  
5 sites shown in Table 1.

1           17. The method of claim 16, wherein the  
2 determining comprises determining a set of bases  
3 occupying a set of the polymorphic sites shown in Table  
4 1.

1           18. The method of claim 16, wherein the nucleic  
2 acid is obtained from a plurality of individuals, and a  
3 base occupying one of the polymorphic positions is  
4 determined in each of the individuals, and the method  
5 further comprising testing each individual for the

6 presence of a disease phenotype, and correlating the  
7 presence of the disease phenotype with the base.

1           19. A computer-readable storage medium for  
2 storing data for access by an application program being  
3 executed on a data processing system, comprising:  
4           a data structure stored in the computer-  
5 readable storage medium, the data structure including  
6 information resident in a database used by the  
7 application program and including:  
8           a plurality of records, each record of the  
9 plurality comprising information identifying a  
10 polymorphisms shown in Table 1.

1           20. The computer-readable storage medium of claim  
2 19, wherein each record has a field identifying a base  
3 occupying a polymorphic site and a location of the  
4 polymorphic site.

1           21. The computer-readable storage medium of claim  
2 19, wherein each record identifies a nucleic acid  
3 segment of between 10 and 100 bases from a fragment shown  
4 in Table 1 including a polymorphic site, or the  
5 complement of the segment.

1           22. The computer-readable storage medium of claim  
2 19, comprising at least 10 records, each record  
3 comprising information identifying a different  
4 polymorphism shown in Table 1.

1           23. The computer-readable storage medium of claim  
2 19, comprising at least 100 records, each record  
3 comprising information identifying a different  
4 polymorphisms shown in Table 1.

1           24. A signal carrying data for access by an  
2 application program being executed on a data processing  
3 system, comprising:  
4           a data structure encoded in the signal, said data  
5 structure including information resident in a database  
6 used by the application program and including:  
7           a plurality of records, each record of the  
8 plurality comprising information identifying a  
9 polymorphism shown in Table 1.

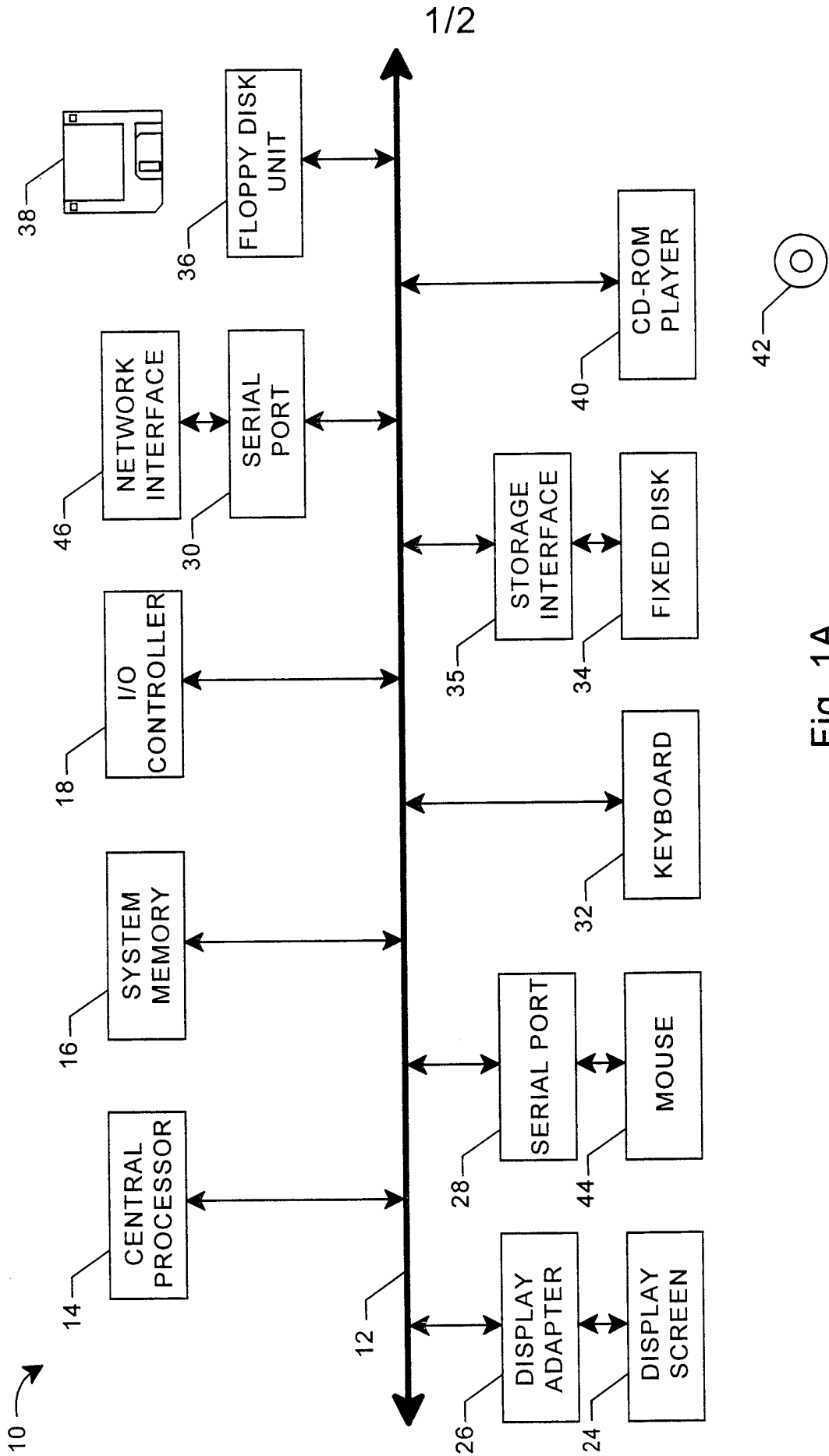


Fig. 1A

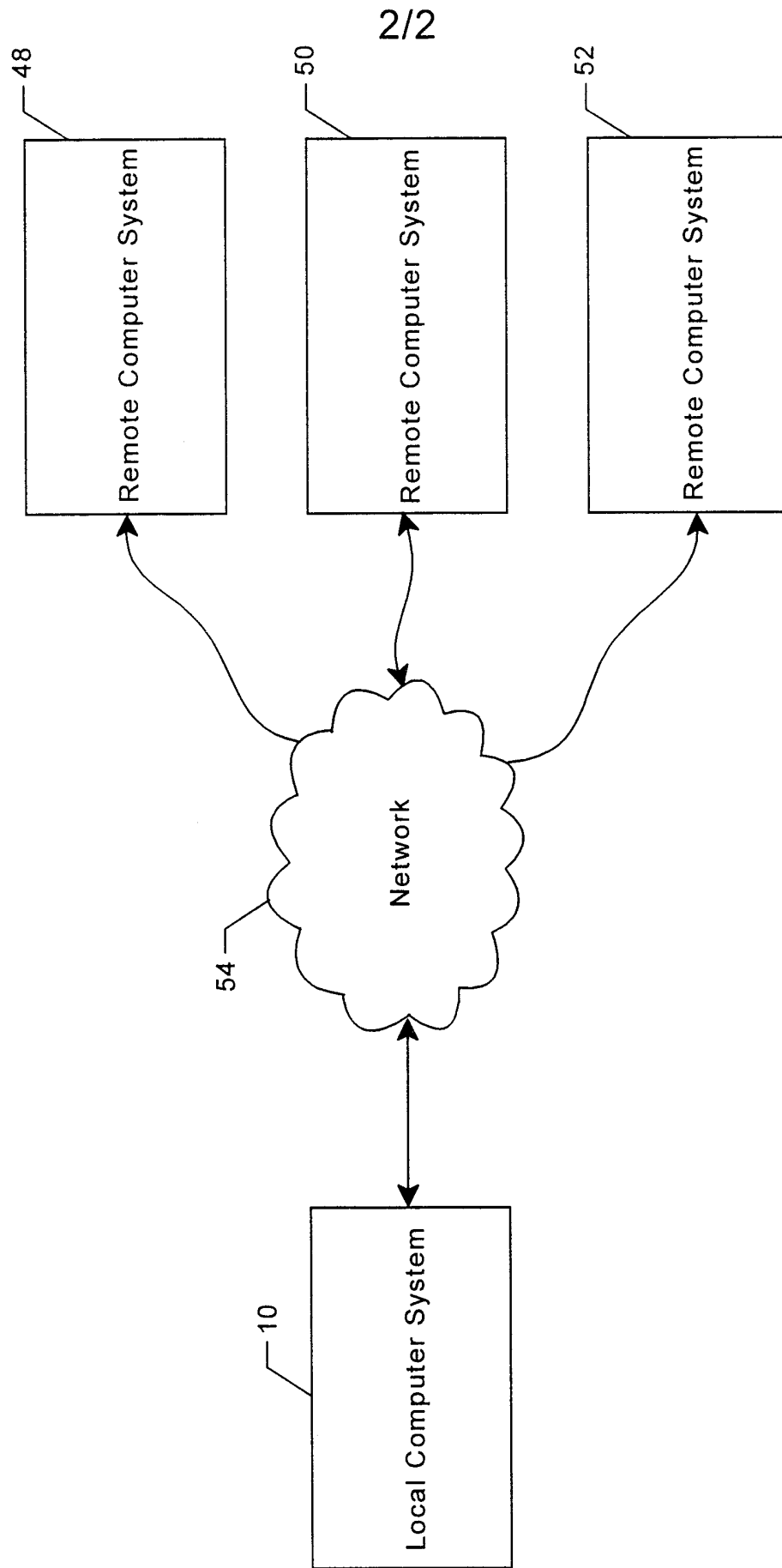


Fig. 1B

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/19325

<p><b>A. CLASSIFICATION OF SUBJECT MATTER</b>                  IPC(6) :C07H 21/00                  US CL :536/23.1                  According to International Patent Classification (IPC) or to both national classification and IPC</p>		
<p><b>B. FIELDS SEARCHED</b>                  Minimum documentation searched (classification system followed by classification symbols)                  U.S. : Please See Extra Sheet.</p>		
<p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched</p>		
<p>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)                  Please See Extra Sheet.</p>		
<p><b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b></p>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MATTHEWS et al. Analytical Strategies for the Use of DNA Probes. Analytical Biochemistry. February 1988, Volume 169, pages 1-25, see the entire document.	1-24
Y	Sigma Chemical Catalog, (Published in 1990 by Sigma Chemical Company, P.O. Box 14508, Saint Louis, Missouri 63178) page 845, see especially Product P 0887 as compared to fragment stSG3590 on page 17 of the instant description.	1,2,4,5, 8,10,11, 13
<p><input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.</p>		
* *A* *E* *L* *O* *P*	Special categories of cited documents: document defining the general state of the art which is not considered to be of particular relevance earlier document published on or after the international filing date document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) document referring to an oral disclosure, use, exhibition or other means document published prior to the international filing date but later than the priority date claimed	*T* *X* *Y* *&*
		later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art document member of the same patent family
Date of the actual completion of the international search 04 JANUARY 1999		Date of mailing of the international search report <b>22 JAN 1999</b>
Name and mailing address of the ISA-US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230		Authorized officer <i>D. Lawrence For</i> ARDIN MARSCHIEL Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/19325

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	YE et al. Progression of Coronary Atherosclerosis Is Associated with a Common Genetic Variant of the Human Stromelysin-1 Promoter Which Results in Reduced Gene Expression. The Journal of Biological Chemistry. 31 May 1996, Volume 271, Number 22, pages 13055-13060, see especially the abstract and page 13056, second column, first full paragraph.	1-24
Y	US 5,639,607 A (DESNICK ET AL.) 17 June 1997, see especially the abstract.	1-24
Y	US 5,449,604 A (SCHELLENBERG ET AL.) 12 September 1995, see especially the abstract and Table 1 in columns 15-18.	1-24
Y	US 4,965,190 A (WOO ET AL.) 23 October 1990, see especially the abstract and Figures 2B and 3.	1-24
Y	US 5,494,794 A (WALLACE) 27 February 1996, see especially the abstract and Figure 8.	1-24
Y	US 5,400,249 A (SOLL ET AL.) 21 March 1995, see the entire disclosure.	19-24

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/19325

B. FIELDS SEARCHED

Minimum documentation searched

Classification System: U.S.

341/1,50,126,137; 360/1,32,40,131,135; 130; 365/49,52; 435/6; 536/23.1,24.1,24.3,24.31,24.32,24.33; 935/77,78

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, MEDLINE, WPI, BIOTECH ABS, EMBASE search terms: nucleic acid, hybridize, polymorphic, probe, pattern, computer, disk, floppy, memory