

ORIGINAL ARTICLE

Aritoshi Iida · Susumu Saito · Akihiro Sekine
Takuya Kitamoto · Yuri Kitamura · Chihiro Mishima
Saori Osawa · Kimie Kondo · Satoko Harigae
Yusuke Nakamura

Catalog of 434 single-nucleotide polymorphisms (SNPs) in genes of the alcohol dehydrogenase, glutathione S-transferase, and nicotinamide adenine dinucleotide, reduced (NADH) ubiquinone oxidoreductase families

Received: March 15, 2001 / Accepted: April 6, 2001

Abstract An approach based on development of a large archive of single-nucleotide polymorphisms (SNPs) throughout the human genome is expected to facilitate large-scale studies to identify genes associated with drug efficacy and side effects, or susceptibility to common diseases. We have already described collections of SNPs present among various genes encoding drug-metabolizing enzymes. Here we report SNPs for such enzymes at additional loci, including 8 alcohol dehydrogenases, 12 glutathione S-transferases, and 18 belonging to the NADH-ubiquinone oxidoreductase family. Among DNA samples from 48 Japanese volunteers, we identified a total of 434 SNPs at these 38 loci: 27 within coding elements, 52 in 5' flanking regions, five in 5' untranslated regions, 293 in introns, 20 in 3' untranslated regions, and 37 in 3' flanking regions. The ratio of transitions to transversions was approximately 2.1 to 1. Among the 27 coding SNPs, 13 were nonsynonymous changes that resulted in amino acid substitutions. Our collection of SNPs derived from this study should prove useful for investigations designed to detect associations between genetic variations and common diseases or responsiveness to drug therapy.

Key words Single-nucleotide polymorphism (SNP) · Alcohol dehydrogenase · Glutathione S-transferase · NADH ubiquinone oxidoreductase · High-dense SNP map · Nonsynonymous substitution · Japanese population

Introduction

Human genetic variations result from a dynamic process over time that includes sudden mutations, random genetic drift, and, in some cases, founder effects. Variations at a single gene locus are useful as markers of individual risk for adverse drug reactions or susceptibility to complex diseases (for reviews, see Risch and Merikangas 1996; Kruglyak 1997; McCarthy and Hilfiker 2000). Common types of sequence variation in the human genome include single-nucleotide polymorphisms (SNPs), insertion/deletion polymorphisms, and variations in the number of repeats of certain motifs (e.g., microsatellites and variable number of tandem repeat loci). Among those variations, SNPs are the most abundant, stable, and widely distributed across the genome; moreover, they lend themselves to automated analysis on a very large scale with high-throughput typing technologies (for reviews, see Wang et al. 1998; Halushka et al. 1999).

Projects to establish large collections of SNPs at candidate loci and to develop detailed SNP maps across the human genome have been under way for some time in various laboratories (Sherry et al. 2000, 2001). Construction of an SNP database for the Japanese population began in 2000, and already around 34,000 SNPs have been archived (<http://snps.ims.u-tokyo.ac.jp>). Simultaneously, we have been focusing on identifying SNPs within genes encoding drug-metabolizing enzymes, drug receptors, and transporters. Local, highly dense SNP maps have been completed for four gene families related to drug metabolism, including 17 members of the S-transferase family, three quinone oxidoreductases, two epoxide hydrolases, and five N-acetyltransferases (Iida et al. 2001; Saito et al. 2001; Sekine et al. 2001). Such genes are likely to have significant roles in responsiveness to drugs and/or in susceptibility to common diseases such as arteriosclerosis, diabetes mellitus, and cancer.

In this study we report a total of 434 SNPs that we identified in a Japanese population sample by direct sequencing of 38 gene loci encoding alcohol

A. Iida · S. Saito · A. Sekine · T. Kitamoto · Y. Kitamura ·
C. Mishima · S. Osawa · K. Kondo · S. Harigae · Y. Nakamura
Laboratory for Genotyping, The SNP Research Center, Institute of
Physical and Chemical Research (RIKEN), Tokyo, Japan

Y. Nakamura (✉)
Laboratory of Molecular Medicine, Human Genome Center,
Institute of Medical Science, The University of Tokyo, 4-6-1
Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
Tel. +81-3-5449-5372; Fax +81-3-5449-5433
e-mail: yusuke@ims.u-tokyo.ac.jp

dehydrogenases, glutathione S-transferases, and nicotinamide adenine dinucleotide, reduced (NADH) ubiquinone oxidoreductases.

Subjects and methods

DNA samples and genes encoding drug-metabolizing enzymes

Blood samples were obtained, with informed consent, from 48 healthy Japanese individuals for this study, which was approved by the SNP Research Center, RIKEN. Genomic DNA was extracted from each sample according to standard protocols. On the basis of DNA sequences released from the GenBank database, we predicted structures and designed primers to amplify each gene in its entirety as well as up to 2kb upstream from the first exon (putative promoter region) and downstream from the last exon. Accession numbers are shown at the top of each gene map in Fig. 1. Regions corresponding to repetitive sequences were predicted by the RepeatMasker program with the option "Do not mask simple repeats and low-complexity DNA" selected (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>).

Polymerase chain reaction (PCR), direct sequencing, and detection of SNPs

PCR experiments and DNA sequencing were performed according to methods described previously (Ohnishi et al. 2000). Each PCR experiment was carried out using 20ng of DNA pooled from three individuals. All SNPs detected by the PolyPhred computer program (Nickerson et al. 1997) were confirmed by sequencing both strands of each PCR product.

Results and discussion

We present here a series of highly dense SNP maps that include a total of 434 SNPs among 38 genes belonging to three families of enzymes associated with drug metabolism. The genomic structures and locations of identified SNPs are illustrated schematically in Fig. 1. Eight of the genes studied here belong to the alcohol dehydrogenase family, 12 to the glutathione S-transferase family, and 18 to the NADH ubiquinone oxidoreductase family. Of the SNPs identified, 334 (77%) were not previously included in the National Center for Biotechnology Information (U.S.) dbSNP database.

Alcohol dehydrogenase family

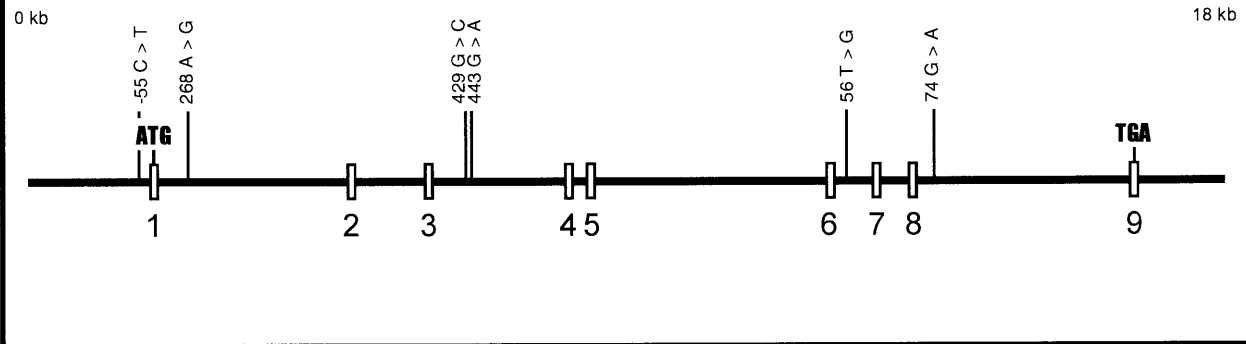
The family of alcohol dehydrogenases (EC1.1.1.1) in vertebrates consists of several enzymes that are able to catalyze reversible oxidation of a wide variety of endogenous and

xenobiotic primary and secondary alcohols, to produce the corresponding aldehydes and ketones. Several distinct classes of vertebrate alcohol dehydrogenase, encoded by nonorthologous genes, have been defined on the basis of sequence homologies and/or unique catalytic properties or gene expression patterns (for reviews, see Duester et al. 1999; Ashmarin et al. 2000). We analyzed here six of the seven known *ADH* gene loci and two genes that had shown sequence similarities to *ADH* by direct sequencing (Fig. 1, a1–a8). A total of 66 SNPs were identified from these eight loci: eight were located in coding elements, eight in 5' flanking regions, one in a 5' untranslated region, 37 in introns, eight in 3' untranslated regions, and four in 3' flanking regions. Among these 66 SNPs, 43 (65%) were novel and seven were located within regions corresponding to repetitive sequences predicted by the RepeatMasker program. Among the eight SNPs identified within coding elements, three were nonsynonymous and one, an Ala97Thr substitution (G289A) within the *HEP27* gene, had not been reported before, while the remaining two (an Arg48His within *ADH2* gene and an Arg272Gln with *ADH3* gene) were identical to SNPs deposited in the dbSNP database (ID: rs1789884 and ID: rs1789912, respectively). With regard to the distribution of SNPs in each gene, we found 27 within the 27-kb genomic region containing the *ADH7* gene but detected only three in the 20-kb *ADH5* genomic region.

Glutathione S-transferase family

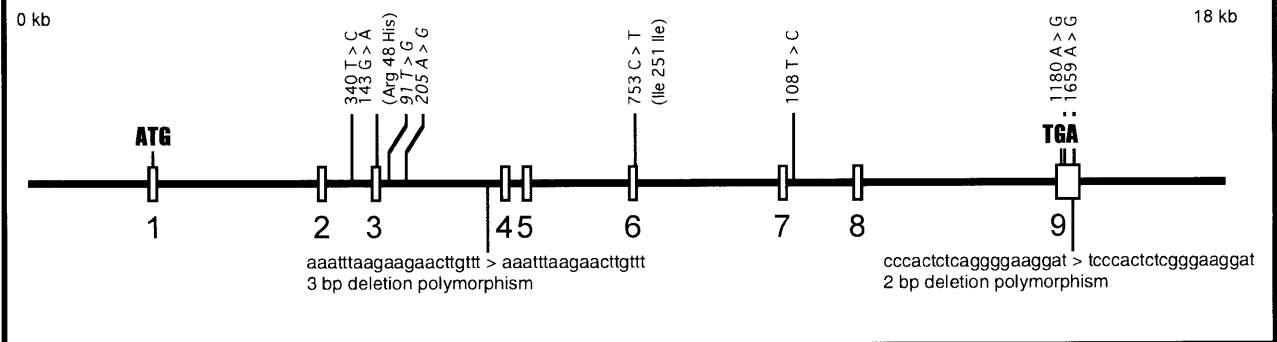
Glutathione S-transferases (EC2.5.1.18) are a unique group of multifunctional isozymes that play important roles in the conjugation and detoxification of various xenobiotics such as aflatoxins and polycyclic aromatic hydrocarbons. On the basis of similarities in amino acid sequences and cross-reactivity of antibodies, mammalian cryptozoic glutathione S-transferases are divided into six classes: alpha, mu, kappa, theta, pi, and sigma. In addition, there is a class of microsomal glutathione S-transferases (for reviews, see Hayes and Pulford 1995; Strange et al. 2000). We analyzed a total of 12 glutathione S-transferase genes by direct sequencing (Fig. 1, g1–g12), and identified 127 SNPs: eight in coding elements, 22 in 5' flanking regions, none in 5' untranslated regions, 86 in introns, seven in 3' untranslated regions, and four in 3' flanking regions. Among these 127 SNPs, 104 (82%) were considered to be novel because they did not overlap with SNPs in the dbSNP database, and 15 were located within regions corresponding to repetitive sequences predicted by the RepeatMasker program. Of the eight SNPs identified within coding regions, six were nonsynonymous and two, a Thr163Ala substitution (A487G) in the *GSTA4* gene and a Val224Ile substitution (G670A) in the *GSTM3* gene, were novel. Regarding distribution, SNPs were most frequent at the *GSTM1* gene locus (21 SNPs were present in the 12-kb region sequenced; i.e., one in every 480bp), while only three SNPs were detected at the 36.5-kb *MGST2* locus (one in every 3000bp).

a1

Alcohol dehydrogenase 1 (*ADH1*): AP002027.1

| | |
|-------------------------------------|---|
| 1 (5' flanking region -55) | atcatgtgtggaactggaat C/T ggggtgttattcaagcaaaaa |
| 2 (intron 1 268) | acatttgccggtaaagcgata A/G tttattccaagctaatcatg |
| 3 (intron 3 429) dbSNP ID:rs1693437 | |
| 4 (intron 3 443) | tgctgagttcacaaaaatgga G/C gctacatggctat G/A gctgaatgagcatgaccttt |
| 5 (intron 6 56) | tacaacttggaggatgcatt T/G aggctgcagaatatatgttt |
| 6 (intron 8 74) | gtctagcagaaaaatgaaaag G/A tggaaggatgagaaaaatta |

a2

Alcohol dehydrogenase 2 (*ADH2*): AP002027.1

| | |
|--|---|
| 1 (intron 2 340) | ctatttttttaaaagcgtgcat T/C cttacataagacttaaatat |
| 2 (coding region 143) dbSNP ID:rs1789884 | ggtggctgtaggaatctgtc G/A (Arg 48 His) cacagatgaccacgtggta |
| 3 (intron 3 91) | aaggcaatgagagacgaaag T/G gcttgcacaaggtcaccgcg |
| 4 (intron 3 205) | atgtattgtacccttcaacc A/G ttatgtaccgagtatctact |
| 5 (coding region 753) dbSNP ID:rs1789882 | caagactacaagaaacccat C/T (Ile 251 Ile) caggaagtgcataaggaat |
| 6 (intron 7 108) | acaattgacaaggcaagatt T/C tgaaaacaatatcaaaaataa |
| 7 (3' untranslated region 1180) dbSNP ID:rs17033 | ttcagcctcctctaccctac A/G agatctggagcaacagctag |
| 8 (3' untranslated region 1659) dbSNP ID:rs1042026 | tttacaagtagtgaaggtcc A/G agagttctaaatacaggaaa |

Fig. 1. Genomic organizations and locations of single-nucleotide polymorphisms (SNPs) in 8 alcohol dehydrogenase genes (**a1**–**a8**), 12 glutathione S-transferase genes (**g1**–**g12**), and 18 NADH ubiquinone oxidoreductase genes (**n1**–**n18**). Exons are represented by *open rectangles* and introns by *horizontal lines*. SNPs are indicated above each gene, and microsatellite sequences, and insertion–deletion polymorphisms are listed below. For cDNA and coding sequences in genomic DNA, the adenine of the initiator Met codon is denoted nucleotide +1. The first nucleotide of the 5' untranslated region is

numbered –1. Likewise, intronic DNA positive numbers start from the guanine of the donor-site invariant GT, and negative numbers start from the guanine of the acceptor-site invariant AG. *Italic type* indicates SNPs located within repetitive sequences predicted by the RepeatMasker program. The complete genomic structure of the *NDUFA2* gene, determined for the first time in this study, has been submitted to the DNA Data bank of Japan under accession number AB054976.1

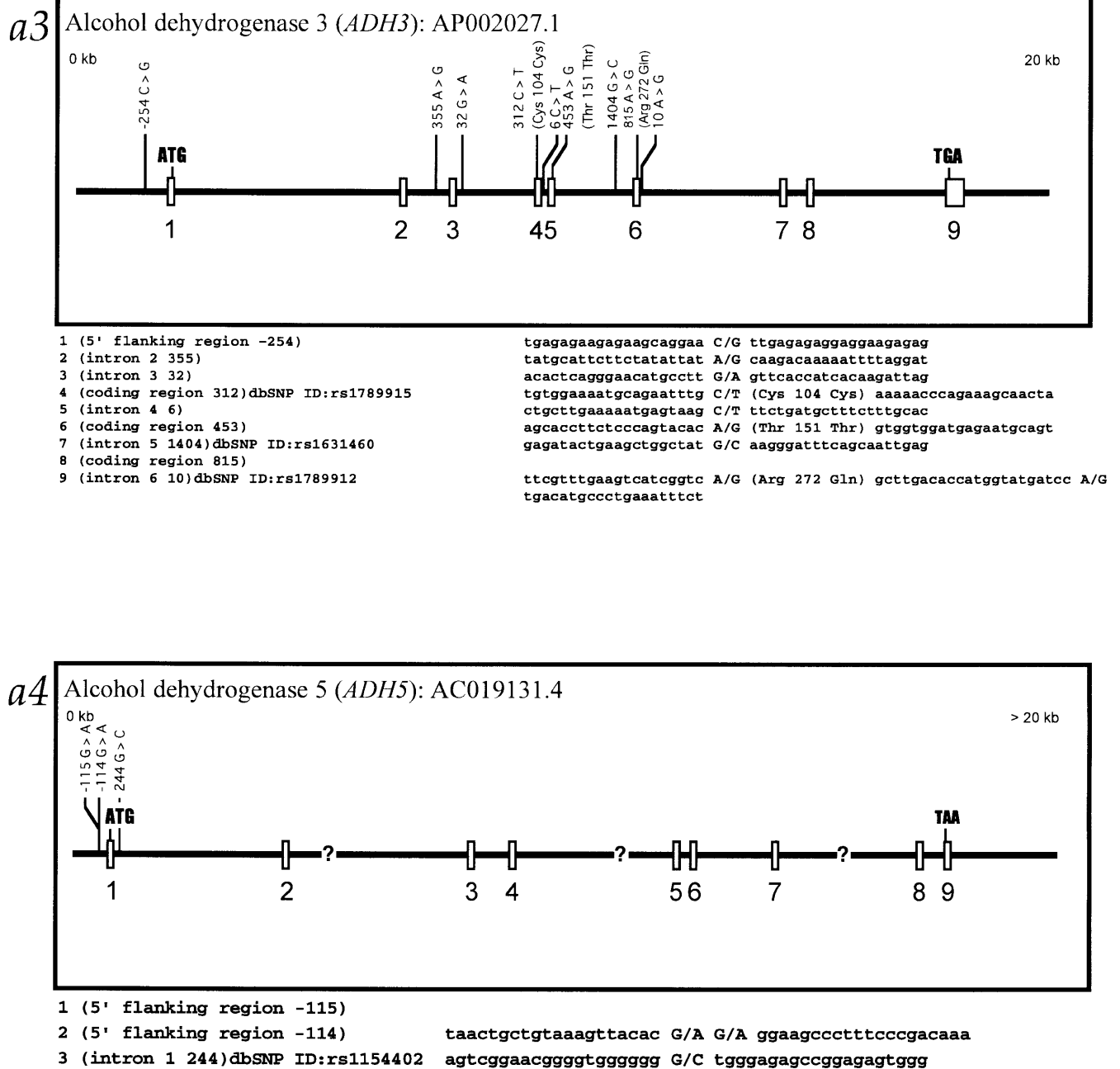
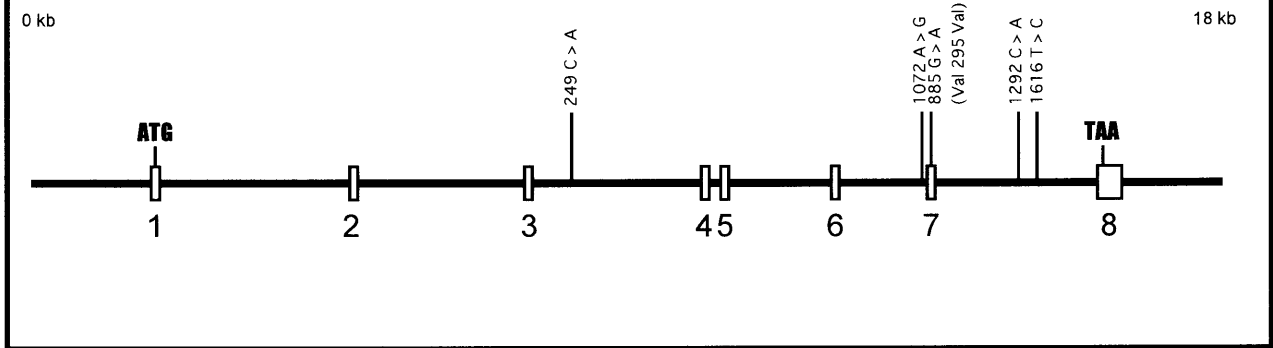


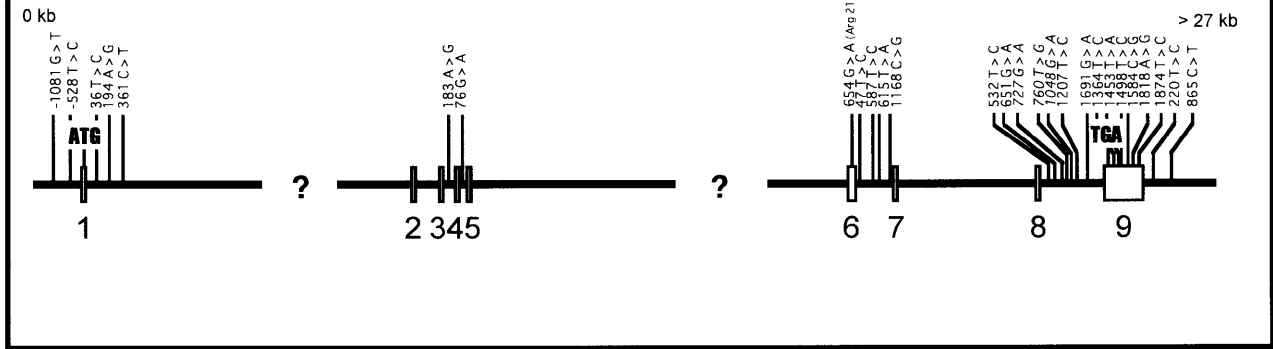
Fig. 1. Continued

a5 Alcohol dehydrogenase 6 (*ADH6*): AP002026.1



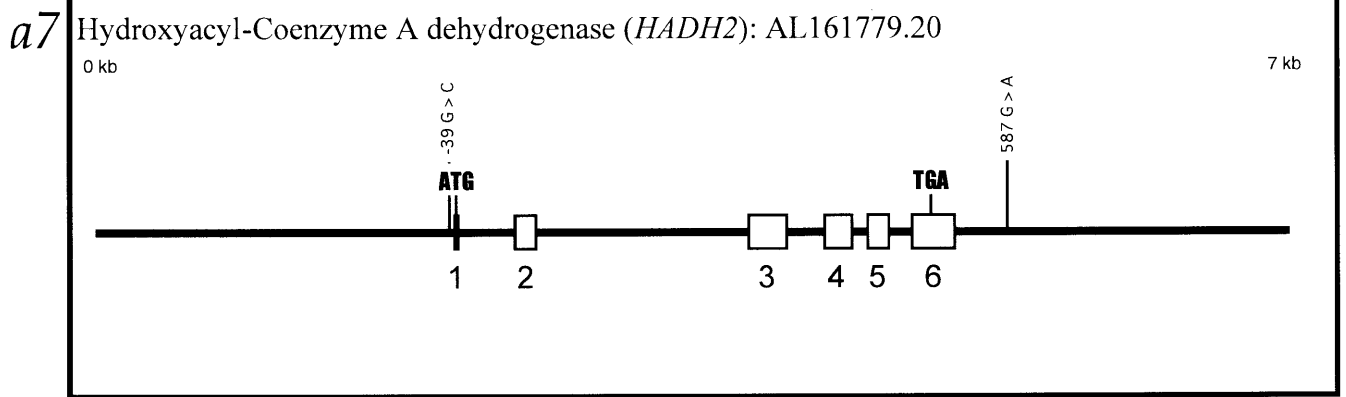
| | | |
|-----------------------|--------------------------------------|----------------------|
| 1 (intron 3 249) | tgaaactggacttgaaagta C/A | aatgagacaaaatttatg |
| 2 (intron 6 1072) | taaccctatactgtattgc A/G | tcactttctaacaggcagct |
| 3 (coding region 885) | gtctgtgtgggtgtgggg G/A (Val 295 Val) | ttgctgccagtggttcaact |
| 4 (intron 7 1292) | gttgagaaacactgcctagt C/A | ccgtctgtggcctagaatt |
| 5 (intron 7 1616) | ctatcacagaataatccgca T/C | agaacactaagcagattacg |

a6 Alcohol dehydrogenase 7 (*ADH7*): AC027065.3

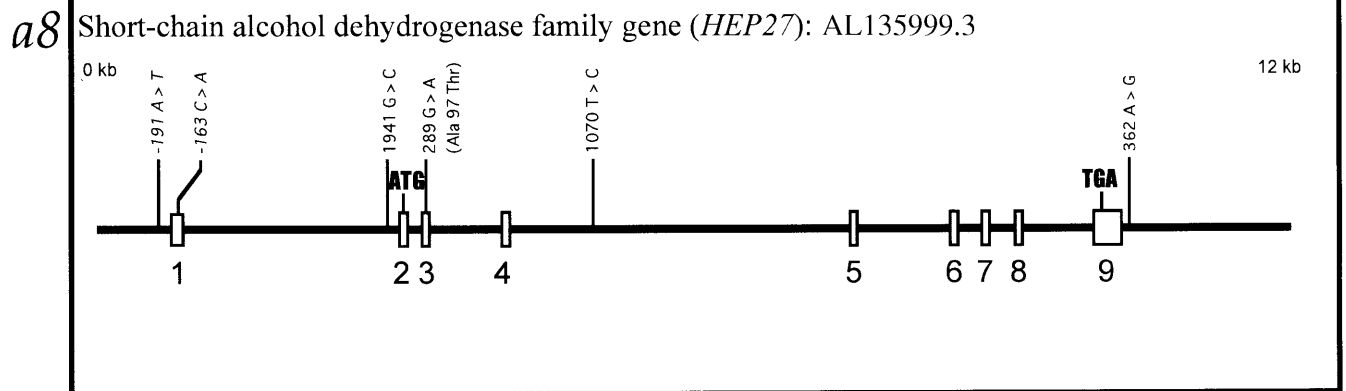


| | | |
|---|---|--------------------------------------|
| 1 (5' flanking region -1081)dbSNP ID:rs1699172 | tgtatacttatttttgaaaa G/T | aaactttgtttgtatctga |
| 2 (5' flanking region -528) | tggtcagacacagaaagttt T/C | acttaactttctacacctaa |
| 3 (intron 1 36)dbSNP ID:rs1789867 | ctgtccctcctcatcatga T/C | ctaagtgtgaggctgata |
| 4 (intron 1 194)dbSNP ID:rs1662029 | gtcagtggaagcaggggtaa A/G | ttggaatagtatccaat |
| 5 (intron 1 361) | tcagtagcatgtgctgcact C/T | gctgcagtagtcaatggga |
| 6 (intron 3 183) | aacctcaacctttagaaggc A/G | aaccttacgggtgttataaa |
| 7 (intron 4 76) | tgaattgaattaattaatac G/A | tgtatttgatgtatcaaaaa |
| 8 (coding region 654)dbSNP ID:rs971074 | aaagtcagctggtgcacttag G/A (Arg 218 Arg) | atcattgggattgacctcaa |
| 9 (intron 6 47)dbSNP ID:rs1789864 | aaaatacccaagttattaga T/C | atgcacaaggtaaatat |
| 10 (intron 6 587)dbSNP ID:rs1617779 | agattcctggaactagaag T/C | tccaacctggcatagcgtaaagagact T/A |
| 11 (intron 6 615) | ggaaaaatggaataaaagcca | aatttccttgtattttagtt C/G |
| 12 (intron 6 1168)dbSNP ID:rs1699175 | aaattcccttgtattttagtt C/G | catggtacactaaaatgccc |
| 13 (intron 8 532) | aagtctaaccatatacacc T/C | ttagtatgccattgtactat |
| 14 (intron 8 651) | gctgctattttattcaagta G/A | gccacaaaatttccttattt |
| 15 (intron 8 727) | ttcagatccctgtaagccag G/A | tattatttttaccatttttagatgaagaccaa T/G |
| 16 (intron 8 760)dbSNP ID:rs284784 | gttgtgaaagcaataaata | agatgaaatgactttcactc G/A |
| 17 (intron 8 1048)dbSNP ID:rs284785 | agatgaaatgactttcactc G/A | ttgggcctataaataattat |
| 18 (intron 8 1207) | tctccacatttggtctagcc T/C | acaggatcatcatattatga |
| 19 (intron 8 1691) | tcctcatctcattgcccac G/A | ctcattgtcttaattcagtc |
| 20 (3' untranslated region 1364) | attacattttgtaaggcta T/C | aattgtatccttttaagaaaa |
| 21 (3' untranslated region 1453)dbSNP ID:rs284786 | ccagctgctgcagatatata T/A | ctcaaaacagatatagcgtga |
| 22 (3' untranslated region 1498) | gatataagtaaatgcactccc T/C | agagtaaatattcacttaaca |
| 23 (3' untranslated region 1584) | aaacactgttatgagttaa C/G | ttggattacattttgaaatc |
| 24 (3' untranslated region 1818) | aatataaacatagagctaga A/G | tcatattatcacttatca |
| 25 (3' untranslated region 1874)dbSNP ID:rs284787 | tacagtgaattgcaagtcc T/C | taagtcctattcactgtgc |
| 26 (3' flanking region 220)dbSNP ID:rs729147 | agaaattgaggcttccaatg T/C | tagggcataatatagattgg |
| 27 (3' flanking region 865) | tacatcaaaagaataaatc C/T | aagaaggaataaacacattt |

Fig. 1. Continued



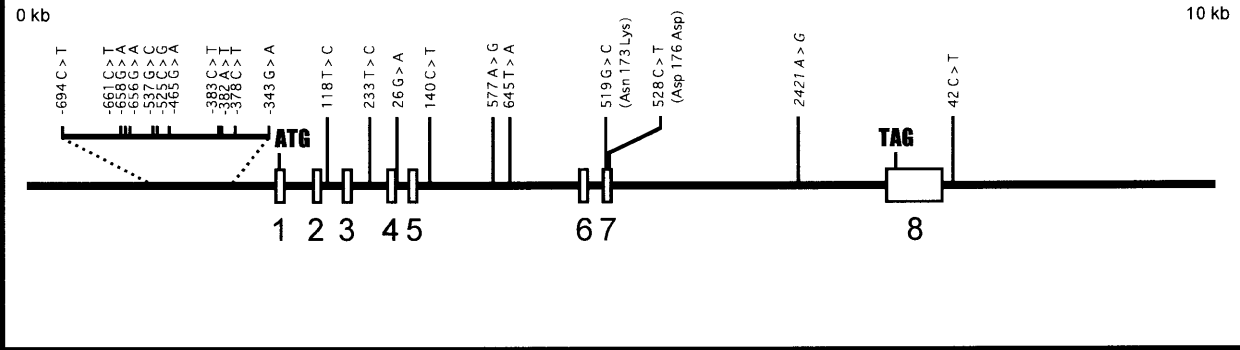
1 (5' flanking region -39)dbSNP ID:rs1264014 cgcatccgcccccttcgcgcg G/C ttgcctcggccaatcaacg
 2 (3' flanking region 587)dbSNP ID:rs1264013 actgcctgggctgagctcat G/A gtctggcttttccattcagt



1 (5' flanking region -191) tcagcactctgtgtctagct A/T aaggtttgtaaatgcaccaa
 2 (5' untranslated region -163) gaacctcatcattccgtaca C/A attttggtgactttgaagag
 3 (intron 1 1941) aaatttacctaaccagcct G/C actctctgccactttctgtt
 4 (coding region 289) ttgtgtgccacgtgggaag G/A (Ala 97 Thr) ctgaggaccgggagcagctg
 5 (intron 4 1070) tgtctcagttcacaggatca T/C gactcttttctcgaaactg
 6 (3' flanking region 362) ggctttgtgtgtctccatt A/G tctgaactgggctgctggg

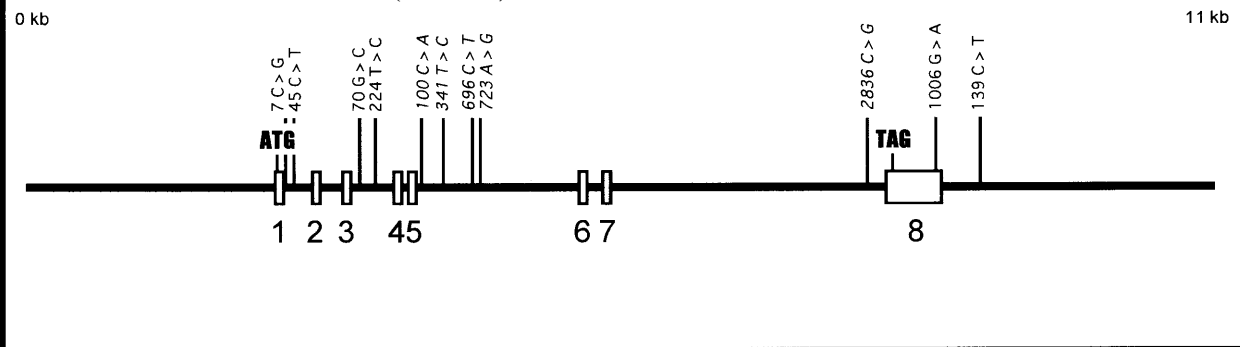
Fig. 1. Continued

g1

Glutathione S-transferase M1 (*GSTM1*): AC000032.7

1 (5' flanking region -694) tacgaagtggctaatttaca C/T agtacttagccagatgaccg
 2 (5' flanking region -661) gatgaccgaaggactcagta C/T ccgaggccctaacagaaa
 3 (5' flanking region -658) gaccgaaggactcagtacc G/A agggccctaacagaaaaca
 4 (5' flanking region -656) ccgaaggactcagtaccga G/A ggccctaacagaaaacaca
 5 (5' flanking region -653) tagaggggagactaagccct G/C ggagtagctttcggatcaga
 6 (5' flanking region -525) taagccctgggagtagcttt C/G ggatcagaggaaatcctgct
 7 (5' flanking region -465) aattaattcccaggttggg G/A ccaccacttttagtctgac
 8 (5' flanking region -383) gcggagagaaggctgagggg C/T accgcggcaggaggagaaa
 9 (5' flanking region -382) cggagagaaggctgagggac A/T ccgcggcaggaggagaaa
 10 (5' flanking region -378) gagaaggctgagggacaccg C/T gggcaggaggagaaaggag
 11 (5' flanking region -343) agggagaagagcttctgctcc G/A ttaggatctggctgtctct
 12 (intron 2 118) tctggagctgcaggctgtc T/C ctccctgagcccoggtgag
 13 (intron 3 233) agtgagtccccggctctctc T/C ctgctcttcttagggag
 14 (intron 4 26) tgtgggtggctgcaatgtgt G/A ggggaaagtgccctctcc
 15 (intron 5 140) actatcagcagttattctca C/T gactccaatgtcatgtcaac
 16 (intron 5 577) ctgccacccattagaagga A/G ctttctacttccctgagct
 17 (intron 5 645) gctggcttggatccagaggg T/A gccaggtgcttgggcctcc
 18 (coding region 519) caccgtatatttgacccaa G/C (Asn 173 Lys) tgettggacgcttcccaaa
 19 (coding region 528) tttgagcccaagtgttga C/T (Asp 176 Asp) gcctcccaaatctgaagga
 20 (intron 7 2421) cagcaccgtgtagaattctc A/G taagtgttagctgttactgt
 21 (3' flanking region 42) atttgcctctggccatctac C/T cagactgtctgtctgtctgt

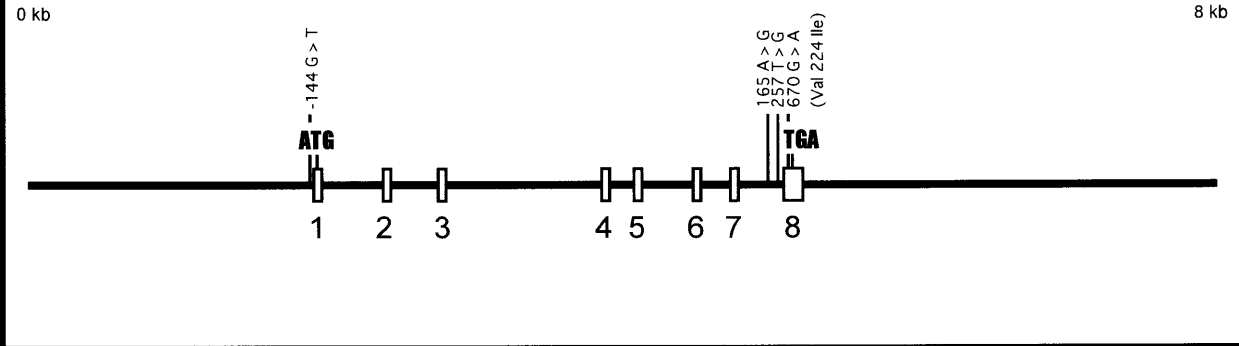
g2

Glutathione S-transferase M2 (*GSTM2*): AC000031.5

1 (intron 1 7) ggaacatccgccccgtgagc C/G agggcccgctgggcggtggg
 2 (intron 1 45) gggacgggggtgcgtgggg C/T ggggaagtgtggagcagctg
 3 (intron 3 70) gactgcattctctctcccca G/C cttagaggtgtaagatcag
 4 (intron 3 224) agcagggccctggctctctct T/C tgccttgcataatgggaaag
 5 (intron 5 100) ttgattccttctggtagtt C/A ttggctctgtgactctaag
 6 (intron 5 341) tcctcttgggtgggtcatgg T/C ctggctggcttcaggagtga
 7 (intron 5 696) accttagctagacacagag C/T gctgatttgtgcatttcaa
 8 (intron 5 723) ttgtgcatttacaatccttt A/G gctaggcagaaaagttctcc
 9 (intron 7 2836) dbSNP ID:rs1149179 aactctgtttcccacatga C/G aaatgggtgataatagattca
 10 (3' untranslated region 1006) ctcagccccagctgtcccc G/A tgttgcataagagcagca
 11 (3' flanking region 139) ttctgtgggcatagtaagg C/T gcttgagaattcttctctcc

Fig. 1. Continued

g3

Glutathione S-transferase M3 (*GSTM3*): AF043105.1

1 (5' flanking region -144)

ccaacgccggcattagtcgc G/T cctgcgcacggccctgtgga

2 (intron 7 165)

agcctaacttctataccttg A/G aggcactgtctacaaaaaaa

3 (intron 7 257)

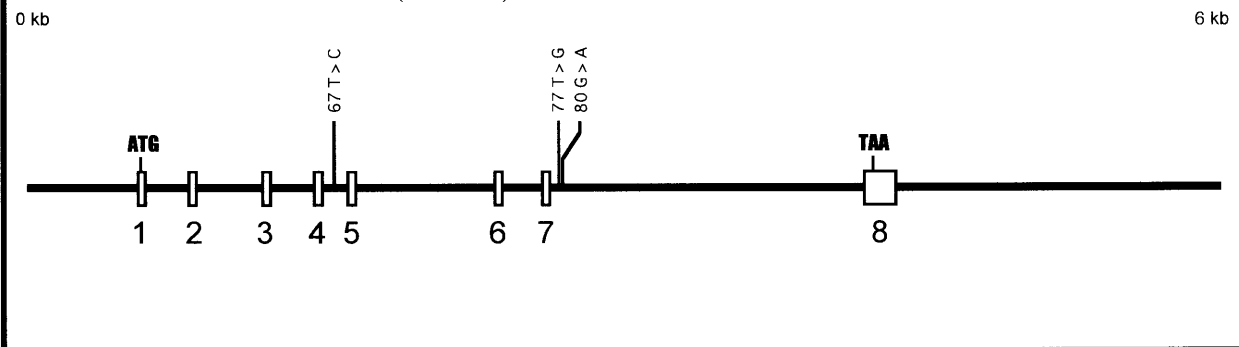
ctgttgactgggtggggctc T/G ttataagattgggtgtatfff

4 (coding region 670)

cccagtggggcaacaagcct G/A (Val 224 Ile)

tatgctgagcaggaggcaga

g4

Glutathione S-transferase M4 (*GSTM4*): M96233.1

1 (intron 4 67)

ttggctggattgggggtgcta T/C gctcagagtgagtctgtgtt

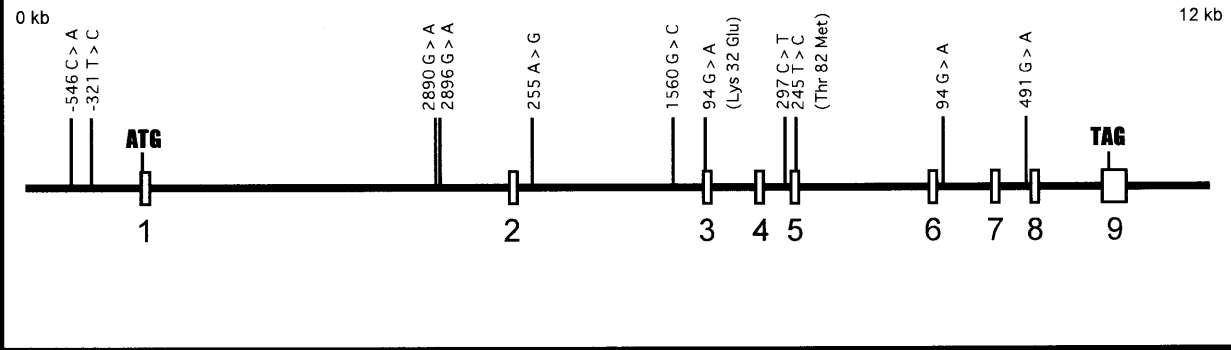
2 (intron 7 77)

gatgctttcccagtcctgga T/G ct G/A cataaagaataacttgcatt

3 (intron 7 80)

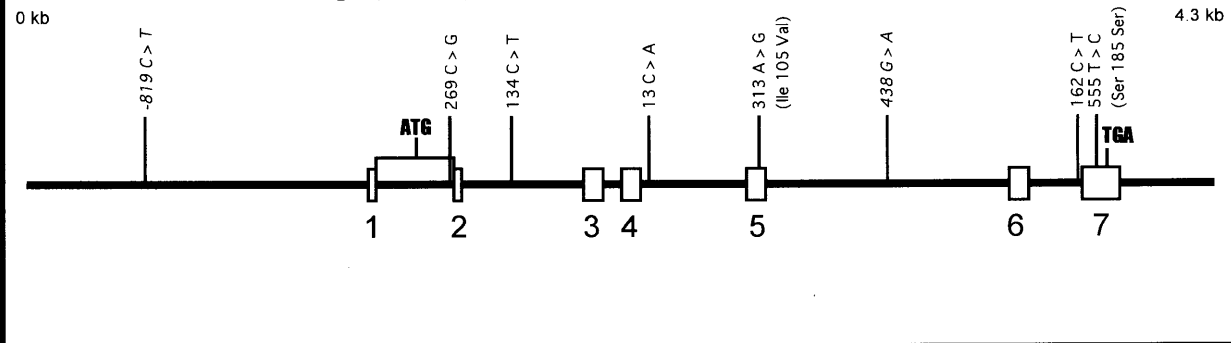
Fig. 1. Continued

g5

Glutathione S-transferase zeta 1 (*GSTZ1*): AC007954.7

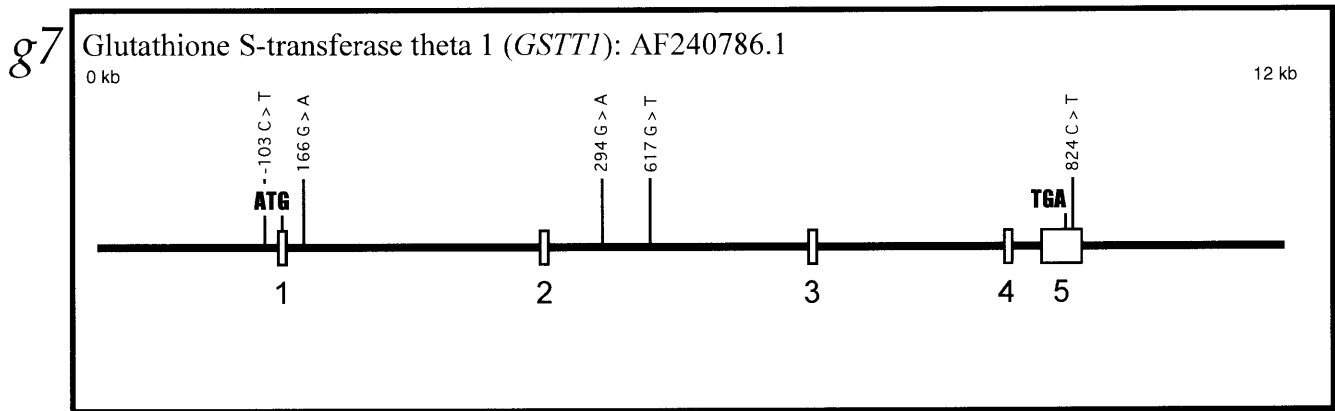
| | | | |
|---|-----------------------|-----|-----------------------------------|
| 1 (5' flanking region -546) | agcaggggcccaccagccgac | C/A | gcctcgaagcgcctgagcc |
| 2 (5' flanking region -321) | tgtctgaccagccgccccgc | T/C | aaggagtcacaagagggcag |
| 3 (intron 1 2890) | aaaatactgcatcaaaacca | G/A | gccacgctctgttggggga |
| 4 (intron 1 2896) | ctgcatcaaaaccaggccac | G/A | ctctgttgggggacaccaa |
| 5 (intron 2 255) | tctccaacactgctctcca | A/G | agcccccttggcaacctgtt |
| 6 (intron 2 1560) | caccactgtttaaggccctg | G/C | gggggcagagttaaacacaa |
| 7 (coding region 94) | ccttgaaggcatcgactac | G/A | (Lys 32 Glu) agacggtgcccatcaatctc |
| 8 (intron 4 297) | agaaggaggagtgttctggc | C/T | ctgtcccctctggtccaggg |
| 9 (coding region 245)dbSNP ID:rs1046428 | cattgagtatctagaggaga | T/C | (Thr 82 Met) gcgtcccactccgcgacttc |
| 10 (intron 6 94) | tatctgaaccagcctcccag | G/A | ctgcttggcctgacagtt |
| 11 (intron 7 491)dbSNP ID:rs731346 | ccctgatgggaaccaccgg | G/A | acctcttccaggtgaccatg |

g6

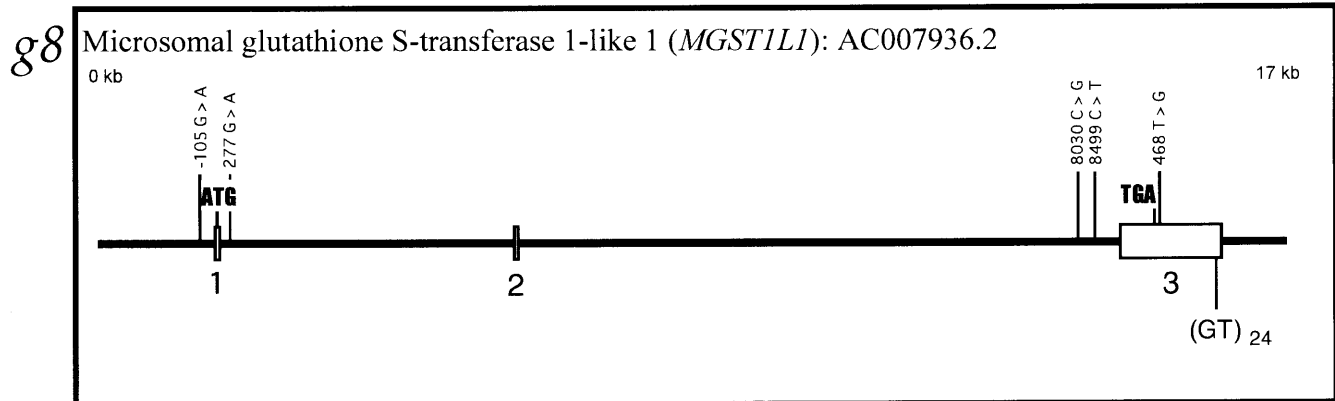
Glutathione S-transferase pi (*GSTPi*): X08058.1

| | | | |
|--|-----------------------|-----|----------------------------------|
| 1 (5' flanking region -819)dbSNP ID:rs643317 | taacaaacctgcacatcctg | C/T | acatgtaccccagaacttca |
| 2 (intron 1 269) | ctccccgggctcccagcaaa | C/G | ttttctttgttcgctgcagt |
| 3 (intron 2 134) | ccccgggctccttctctgtt | C/T | ccgcctctcccgccatgcc |
| 4 (intron 4 13)dbSNP ID:rs762803 | cacccttgggtgagtctttaa | C/A | ctccaagtcaggggcaggca |
| 5 (coding region 313)dbSNP ID:rs947894 | aggacctccgctgcaaatac | A/G | (Ile 105 Val) tctcctcatctacaccaa |
| 6 (intron 5 438) | gtgtgtgcgctgcgtgtgc | G/A | tgtgtgtgcgtgtgtgtg |
| 7 (intron 6 162) | cccgtggtgagtccttag | C/T | cccctgcctgcagatctc |
| 8 (coding region 555)dbSNP ID:rs4891 | gcatatgtggggcgctcag | T/C | (Ser 185 Ser) gccggcccaagctcaagg |

Fig. 1. Continued



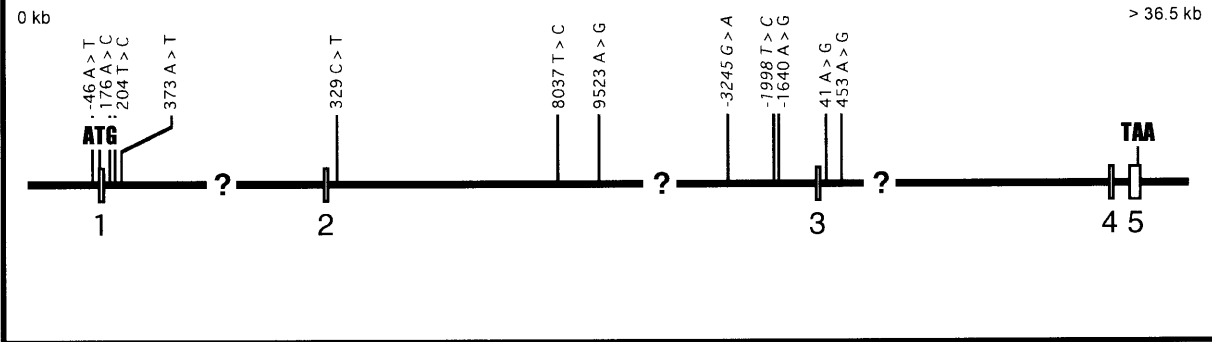
| | | | |
|---|----------------------|-----|----------------------|
| 1 (5' flanking region -103) | taaagagtgtcccagggcgc | C/T | gtgccgcccaatggggcaca |
| 2 (intron 1 166)dbSNP ID:rs140313 | agcagggagatccaagagtc | G/A | gggctcccaaaactctgct |
| 3 (intron 2 294)dbSNP ID:rs140310 | cgatggcagctttgccacc | G/A | tggggcaggcctctggccaa |
| 4 (intron 2 617)dbSNP ID:rs140309 | ctgacattctgccagggccc | G/T | tcttctcctctgggtcccca |
| 5 (3' untranslated region 824)dbSNP ID:rs4630 | agaaagcaggaatggctgc | C/T | taagacttgcccaagtccca |



| | | | |
|--------------------------------|----------------------|-----|----------------------|
| 1 (5' flanking region -105) | tgctgccgctgccgtggggc | G/A | gggcgtgggctgctggct |
| 2 (intron 1 277) | agtgtctgtgagagaagcag | G/A | ttctggagggtggagtgtgg |
| 3 (intron 2 8030) | ggggttatacagagcccctc | C/G | gccccaccacacatatgca |
| 4 (intron 2 8499) | gtatggcaggagtgggtcc | C/T | ggcaagccatagaggtatgg |
| 5 (3' untranslated region 468) | cgccacctgtgaccagcagc | T/G | gatgcctccttggccaccag |

Fig. 1. Continued

g9

Microsomal glutathione S-transferase 2 (*MGST2*): AC019049.4

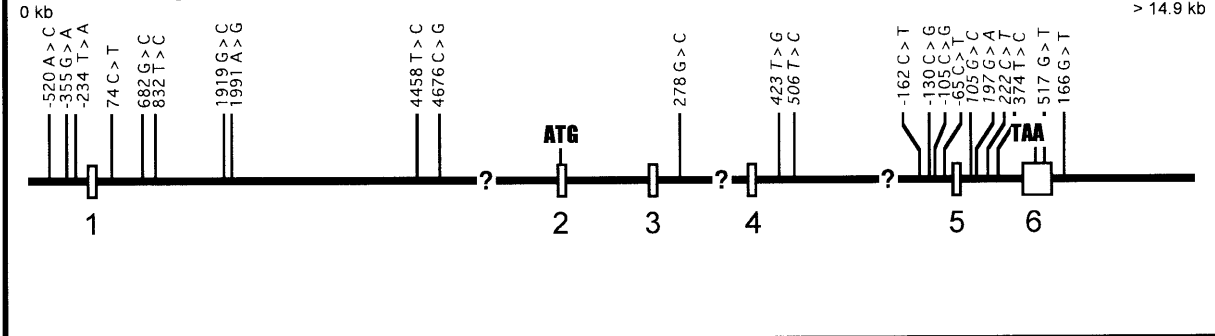
1 (5' flanking region -46)
 2 (intron 1 176)
 3 (intron 1 204)
 4 (intron 1 373)
 5 (intron 2 329) dbSNP ID: rs795589
 6 (intron 2 8037) dbSNP ID: rs1453135
 7 (intron 2 9523) dbSNP ID: rs1377382
 8 (intron 2 -3245)
 9 (intron 2 -1998)
 10 (intron 2 -1640)
 11 (intron 3 41)
 12 (intron 3 453)

```

ggtcagcattcaagtcaag A/T agcgccatttatcttccogt
ggtcaccatgccgcctgct A/C cctccttcccaggggcaag
tcccaggggcaagcagagac T/C gagaacattccagagattag
ttacaagtgttccaaggaa A/T cgtgctcttctaaccctg
agaccacaatagtctgaagt C/T ggattagttatgctacaatt
aagctacactcttctgctc T/C tgtaatgatgtaaaatttat
ctgaaagaccattattaagc A/G tcttctctctctggtattt
cctcgtgatttggccacctc G/A gcctccaaagtgtgggatt
aggccgaggtgggggatca T/C gaggtcaggagatcgagacc
tgtttattccttgcatagcc A/G taatataaagtatgaattt
actgtgttctaataatgact A/G tgatgcttaaacgattaagg
atcagagtgtctatgttgcag A/G tatatgaactttggcttcat

```

g10

Microsomal glutathione S-transferase 3 (*MGST3*): AC064827.2

1 (5' flanking region -520)
 2 (5' flanking region -355)
 3 (5' flanking region -234)
 4 (intron 1 74)
 5 (intron 1 682)
 6 (intron 1 832)
 7 (intron 1 1919)
 8 (intron 1 1991)
 9 (intron 1 4458)
 10 (intron 1 4676)
 11 (intron 3 278)
 12 (intron 4 423)
 13 (intron 4 506)
 14 (intron 4 -162)
 15 (intron 4 -130)
 16 (intron 4 -105)
 17 (intron 4 -65)
 18 (intron 5 105)
 19 (intron 5 197)
 20 (intron 5 222)
 21 (intron 5 374)
 22 (3' untranslated region 517)
 23 (3' flanking region 166)

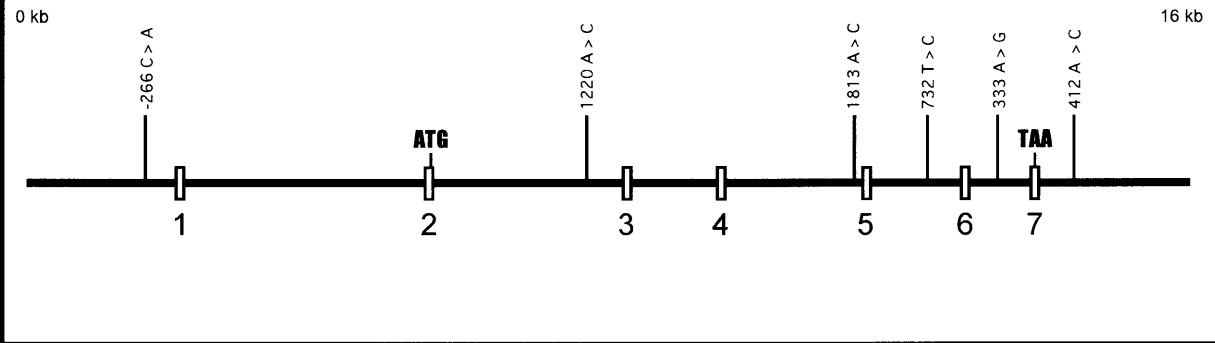
```

acaaaaggccctaacagcg A/C taaatcattcacttcggga
cgctaaaaccgctacggcg G/A ctctgctggggacaaattat
ctgggggagtagatatatgt T/A tttgagaatgagaggagtaa
agcctttgcgaggcactcc C/T atatttcagcctatgagcagc
agaaaatgccctctctttat G/C tggggtggcagcacggagcc
cgagtttacaagctacataa T/C agcgtcgggggcaagtaagt
aataaaaatcctgagtttct G/C tcactcgctcttacagtacc
tgtaattaggcaacagggaaa A/G ttgactactcttcaaatgc
tcttccatctcctaacata T/C agttagcttccactctccaa
tgaatgcaatgcaattgt C/G gggggatagttacttttcat
cagcatgaccatctaaacc G/C atgttgactctcccagcct
cttgcccttttgtgtgggg T/G ggggggtggtcacagagaaag
gtcagagagaagaaaacaaag T/C ggggaaaggtggaaggggat
tcacagatattttatttcc C/T gactgaaactaacttaattc
acttaattctacctaatttg C/G gtggggagtagttggccaaa
ggagtagttggccaaatcat C/G aaattgttaacttttgcta
aacatatgtgtaatacaacc C/T taggtgttaaaaaaggtttg
atcccagcactttgggagcg G/C aaggcaggcagattgcttga
aaaaaatcaaaaattagcc G/A gatgtgggtgacacactg
tggtggtgcacacctgtagt C/T ccagctacttgggaggttga
tcttatgctactatattttt T/C ttcttgggaatttgagaaaa
atgacttaccttatttcca G/T ttacatttttttctaaata
agctctgattgtgggtgatga G/T gtatagctatgccacagtga

```

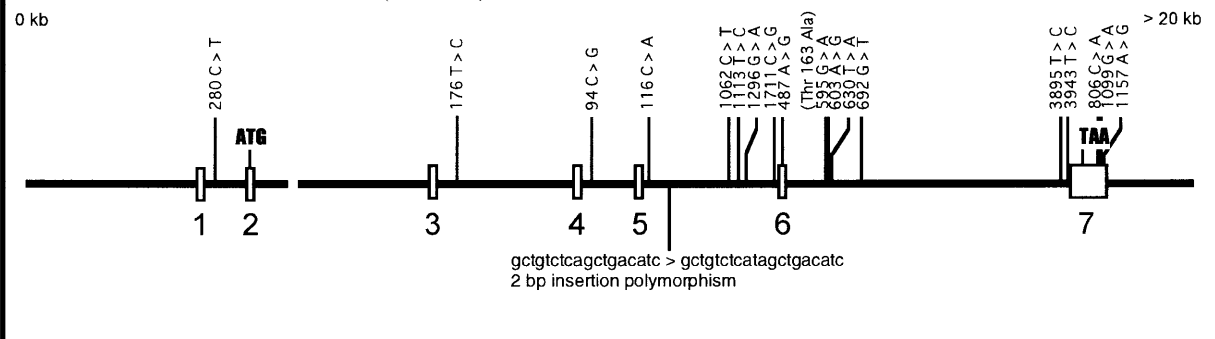
Fig. 1. Continued

g11

Glutathione S-transferase A1 (*GSTA1*): AC021133.4

| | | |
|-----------------------------|--------------------------|-----------------------|
| 1 (5' flanking region -266) | ttgcaaaaagagcaaaatct C/A | ggtgaaatgtattgtgtaaa |
| 2 (intron 2 1220) | gagacacaggctttcctaag A/C | tatgacaacaccataactag |
| 3 (intron 4 1813) | aaaggcaccactggagggtg A/C | attatthttgccatcacctga |
| 4 (intron 5 732) | gaagagtgttgcacgaagg T/C | ggagtcactgcccaggagg |
| 5 (intron 6 333) | ttatcccatatgtgcccaca A/G | tgagccggtctgagcagagc |
| 6 (3' flanking region 412) | ctttcttatgcatttgcaaa A/C | caatgattctgtctgctgtg |

g12

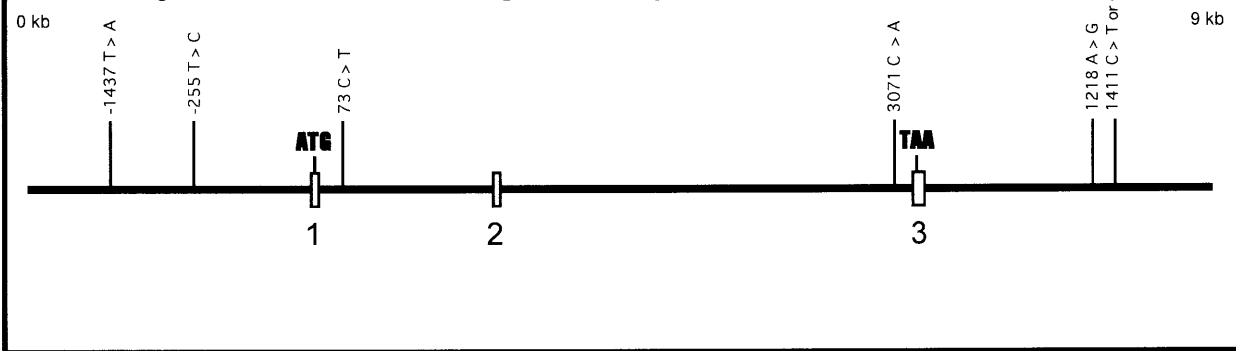
Glutathione S-transferase A4 (*GSTA4*): AC025085.4

| | | |
|--|---------------------------|-----------------------------------|
| 1 (intron 1 280) | gcattggtggaaggtgggct C/T | ggatcgtccccggcctggc |
| 2 (intron 3 176) | ggaaatcacttcttattcaa T/C | agttccataaaagctggccg |
| 3 (intron 4 94) | acaccacattacttattgt C/G | ttacatagtttagtgagatca |
| 4 (intron 5 116) dbSNP ID:rs316128 | ctgagcccagttgaactaac C/A | taagcgctaattggactccta |
| 5 (intron 5 1062) | cacacttgtgcacatgcaga C/T | accatgggcatccaagagt |
| 6 (intron 5 1113) dbSNP ID:rs316131 | ggggatctcccccttgctac T/C | actgctcttgctctccct |
| 7 (intron 5 1296) dbSNP ID:rs316132 | cacctcttgatgttctagca G/A | cagaactgggaaataagagt |
| 8 (intron 5 1711) dbSNP ID:rs783679 | ctgattcaatagcctcatca C/G | tctctccatgtggaatttg |
| 9 (coding region 487) | cagatgtgatttactccaa A/G | (Thr 163 Ala) ccatttagctctagaagag |
| 10 (intron 6 595) | | |
| 11 (intron 6 603) dbSNP ID:rs316136 | tgagctctgagagcaaatga G/A | agatggtt A/G gcaccctaacaatcacccca |
| 12 (intron 6 630) | taaacatcaccccaaggat T/A | cctaccattctctctgtgag |
| 13 (intron 6 692) dbSNP ID:rs316137 | tggttcaaatatccctcag G/T | gttattctggcaggattgat |
| 14 (intron 6 3895) dbSNP ID:rs375887 | acaagtatttattgaacaaa T/C | tttggggggccattttgcat |
| 15 (intron 6 3943) | tcttcgtagatctcaataacc T/C | ttttgttagccttaagatt |
| 16 (3' untranslated region 806) dbSNP ID:rs367836 | agttgggtctttttgtgtcaa C/A | gagatcatctctctagaaa |
| 17 (3' untranslated region 1099) | taatacaaacgaaatgtcta G/A | taaatgactctctctgagc |
| 18 (3' untranslated region 1157) dbSNP ID:rs405729 | tagtaatgaatgcaatcagt A/G | ttagccaaaataaagaattt |

Fig. 1. Continued

n1

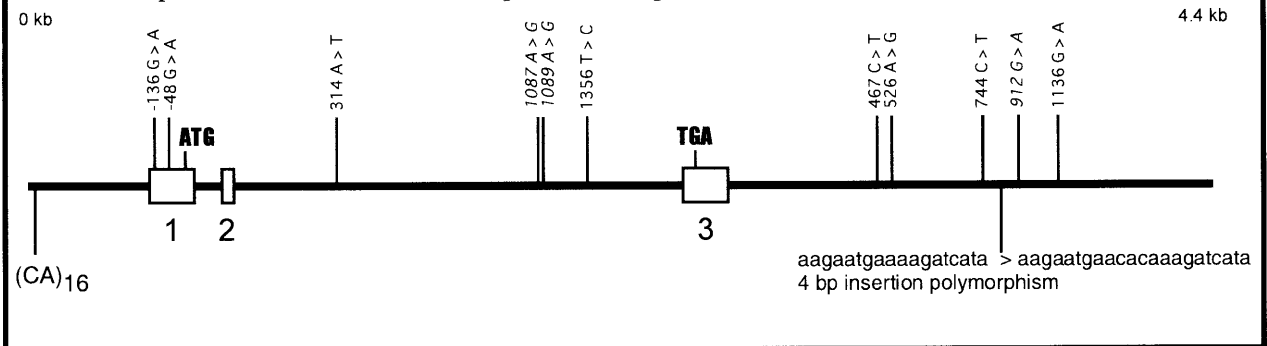
NADH ubiquinone oxidoreductase 1 alpha subcomplex 1 (*NDUFA1*): AC002477.1



- 1 (5' flanking region -1437) agggctaaaaatcctgatta T/A acctaccttgaagcttttaa
 2 (5' flanking region -255)dbSNP ID:rs1800823 gctctgccatttttagagtcc T/C gagctccgaaacagccgtgg
 3 (intron 1 73)dbSNP ID:rs708463 gtgggtgggcagggagaccgt C/T agcctgcgaaccctctctcc
 4 (intron 2 3071) aataaaagtacatggcatat C/A tttgatgggaacagacttgt
 5 (3' flanking region 1218) aactccatgtgtataaagca A/G caccacagatgacacttcca
- tri-allelic
 6,7 (3' flanking region 1411) ggattgtgccatcccttgat C/T/G ggcaatgaccttttactttt

n2

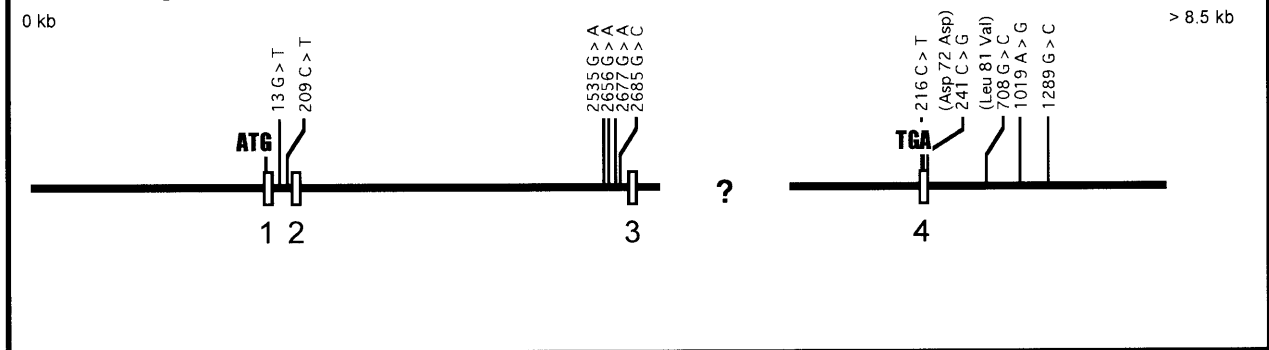
NADH ubiquinone oxidoreductase 1 alpha subcomplex 2 (*NDUFA2*): AB054976.1



- 1 (5' untranslated region -136)dbSNP ID:rs778594 cctcgcagcggctcctacaat G/A cttttatatcattggccaag
 2 (5' untranslated region -48)dbSNP ID:rs778593 tccgcggttgggtcagaccgg G/A gcacttggcctgaagacctg
 3 (intron 2 314)dbSNP ID:rs778592 tccaaataacgtgtagccag A/T gtcagtccaggtagtggtg
 4 (intron 2 1087) aacatacaaaaattagccgg A/G t A/G tggtggcgggacactgtaat
 5 (intron 2 1089) ttccttgaacaaccattg T/C ggcatccagaatcagccaa
 6 (intron 2 1356) cacagcctcatgggtcagcc C/T actccagaggggtgattccc
 7 (3' flanking region 467) tctgacatggaacctatata A/G gttagccaggcccagagaaa
 8 (3' flanking region 526)dbSNP ID:rs702399 ggaagcaggggcccctggcca C/T agccgctggcagtaagcagg
 9 (3' flanking region 744) catgccaagctcttctacat G/A catgttctcatactcctaca
 10 (3' flanking region 912)dbSNP ID:rs778590 actgctgtccatacaatcc G/A actaccagatcttggattca
 11 (3' flanking region 1136)dbSNP ID:rs753279

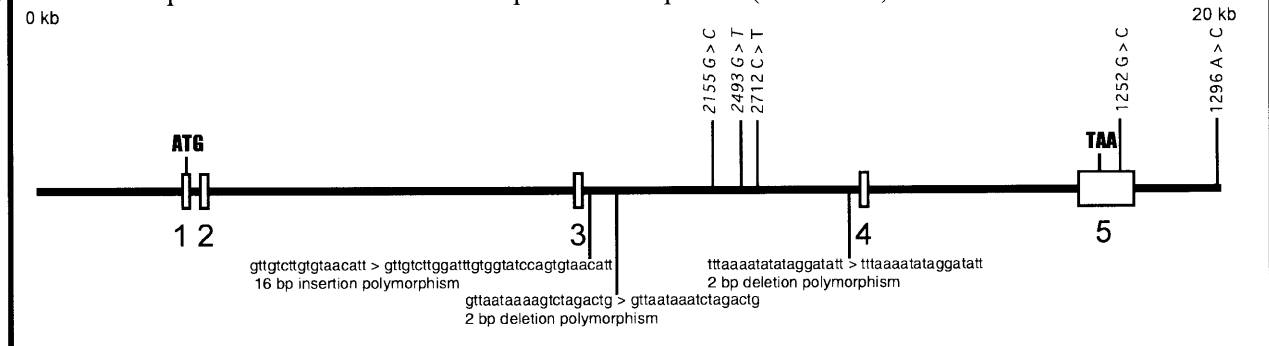
Fig. 1. Continued

n3

NADH ubiquinone oxidoreductase 1 alpha subcomplex 3 (*NDUFA3*): AC009968.6

| | |
|--|---|
| 1 (intron 1 13)dbSNP ID:rs254260 | ggctgcgagtaagtgcaggt G/T cgggtggcgcacggggctcg |
| 2 (intron 1 209)dbSNP ID:rs254259 | ctacttcaggggtgacgct C/T cttgccacccttcaggagt |
| 3 (intron 2 2535)dbSNP ID:rs254257 | atctgtgagtgtaggctcc G/A accctaccctccacttaacc |
| 4 (intron 2 2656) | tcctctgtgcctccctgc G/A cactttatcttcccttggc |
| 5 (intron 2 2677)dbSNP ID:rs254256 | |
| 6 (intron 2 2685)dbSNP ID:rs254255 | cactttatcttcccttggc G/A aggctca G/C cttctcttcc |
| | cctctcttca |
| 7 (coding region 216)dbSNP ID:rs1061333 | gtgccagccacccccagga C/T (Asp 72 Asp) cctcagggcc |
| | ccagcctgga |
| 8 (coding region 241) | agggcccagcctggagtgg C/G (Leu 81 Val) tgaagaact |
| | gtgagcact |
| 9 (3' flanking region 708)dbSNP ID:rs254264 | cccacactggccttccctt G/C cactcgtcttctctgatcc |
| 10 (3' flanking region 1019) | tccttacctgcactggcacc A/G gctctggagccccagtcct |
| 11 (3' flanking region 1289)dbSNP ID:rs1193923 | gtgcctcatcctggcagg G/C aggagggggaggtagtgat |

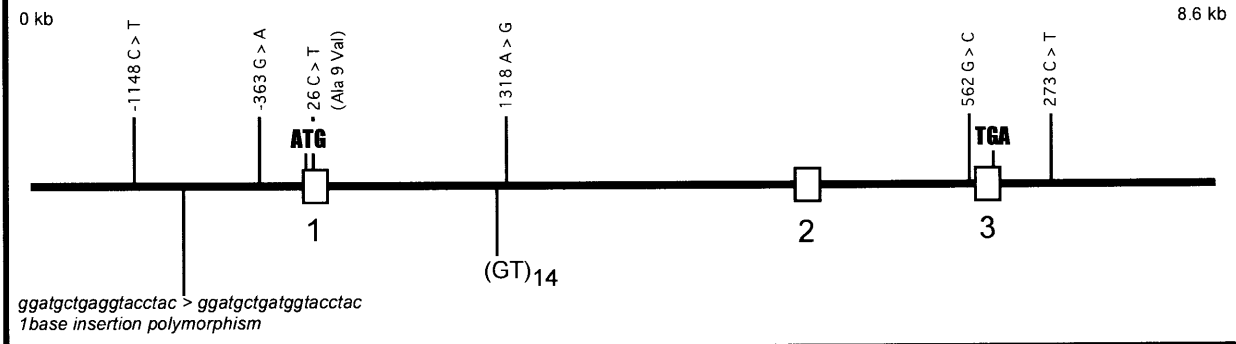
n4

NADH ubiquinone oxidoreductase 1 alpha subcomplex 5 (*NDUFA5*): AC073323.5

| | |
|--|--|
| 1 (intron 3 2155) | agactctagcatggtacctg G/C aacataaggttccttagaaa |
| 2 (intron 3 2493) | ggcatattgctagttttctc G/T gtctcaatttcatcatctat |
| 3 (intron 3 2712) | acaaatgttgaactgttcac C/T taacacaggctttttctgaa |
| 4 (3' untranslated region 1252)dbSNP ID:rs6737 | |
| | gaaatttataatctggcatac G/C gataggcttatattcaaaac |
| 5 (3' flanking region 1296) | aggtatctaaaaggatttgc A/C atttggtcattggttcttctc |

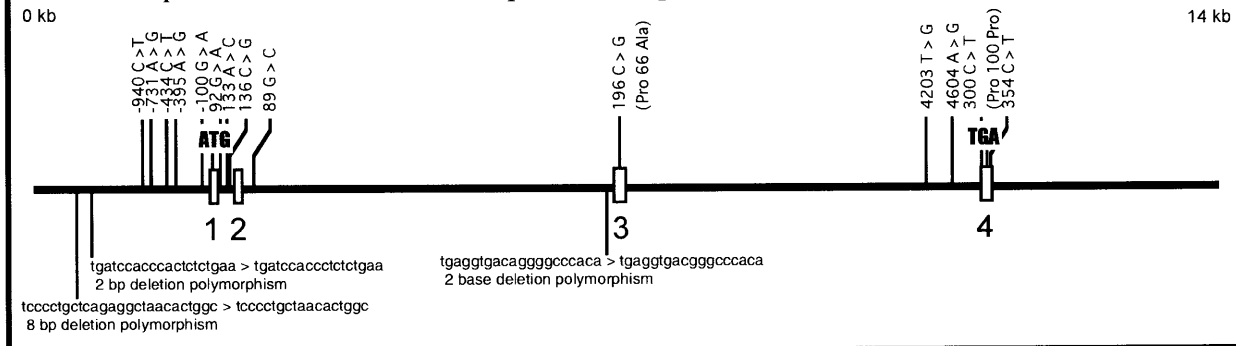
Fig. 1. Continued

n5

NADH ubiquinone oxidoreductase 1 alpha subcomplex 6 (*NDUFA6*): AL021878.1

- | | | |
|--|---------------------------------------|----------------------|
| 1 (5' flanking region -1148) | tttataatttatatatgtta C/T | gtgctttcttttgtatagct |
| 2 (5' flanking region -363) | actaccaaggagcgcggcg G/A | cagccggatagcaggacgct |
| 3 (coding region 26) | ggggagcggcgctccgccaag C/T (Ala 9 Val) | tacttctaccgccagcact |
| 4 (intron 1 1318) | attcagcagtttgaaaacat A/G | atgtttgcctggcagaatac |
| 5 (intron 2 562) | agttaaagaatctgaaaagt G/C | tcagaaatgattaccctga |
| 6 (3' flanking region 273) dbSNP ID:rs742085 | gccctatgattagacactt C/T | agtgtcagggcacagattat |

n6

NADH ubiquinone oxidoreductase 1 alpha subcomplex 7 (*NDUFA7*): AC010323.6

- | | | |
|--|--|----------------------------|
| 1 (5' flanking region -940) dbSNP ID:rs1057232 | ggacaaaaaaaaaggagcaga C/T | tgtgacagaccattcccac |
| 2 (5' flanking region -731) | accaaccaaaggctcatcaa A/G | ggggtgtcctctttgcacc |
| 3 (5' flanking region -434) | aaaggaaccatcagaaccc C/T | gtgatgaaatgagaatcggc |
| 4 (5' flanking region -395) | gctcccgattccggctggc A/G | ggggttagggcagggtagag |
| 5 (5' flanking region -100) | agaggagtcacgtgcttcgg G/A | gagagcctttataggacgtt |
| 6 (intron 1 92) | tcacctccctcctaagccgg G/A | acccttcgctctccccgaat |
| 7 (intron 1 133) | | |
| 8 (intron 1 136) | ctccctgggaacccccagct A/C | gt C/G accccttcagccccggacc |
| 9 (intron 2 89) | tccttagaccctgaaacg G/C | agggtgacatcctgccacc |
| 10 (coding region 196) | gccgcccgggaatctgtgcc C/G (Pro 66 Ala) | cttccatcatcatgtcgtcg |
| 11 (intron 3 4203) | gcctccaccctggggcgcc T/G | cctccatcaccaccctcc |
| 12 (intron 3 4604) | gggccttggtgtacgctggag A/G | ccaaaagtgggaagggagga |
| 13 (coding region 300) dbSNP ID:rs1045629 | gcggtgactccagctcctcc C/T (Pro 100 Pro) | ataaagaggtgggagctgtc |
| 14 (3' untranslated region 354) dbSNP ID:rs561 | tacctgtgacactgcaccct C/T | acggccaccgactactttg |

Fig. 1. Continued

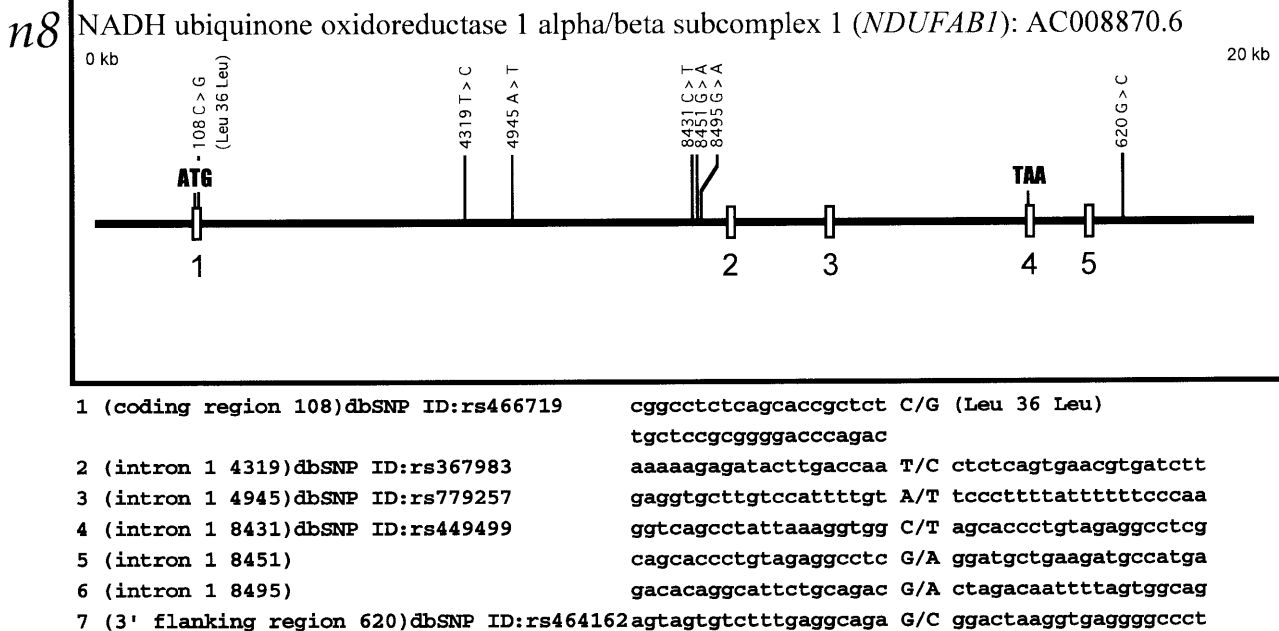
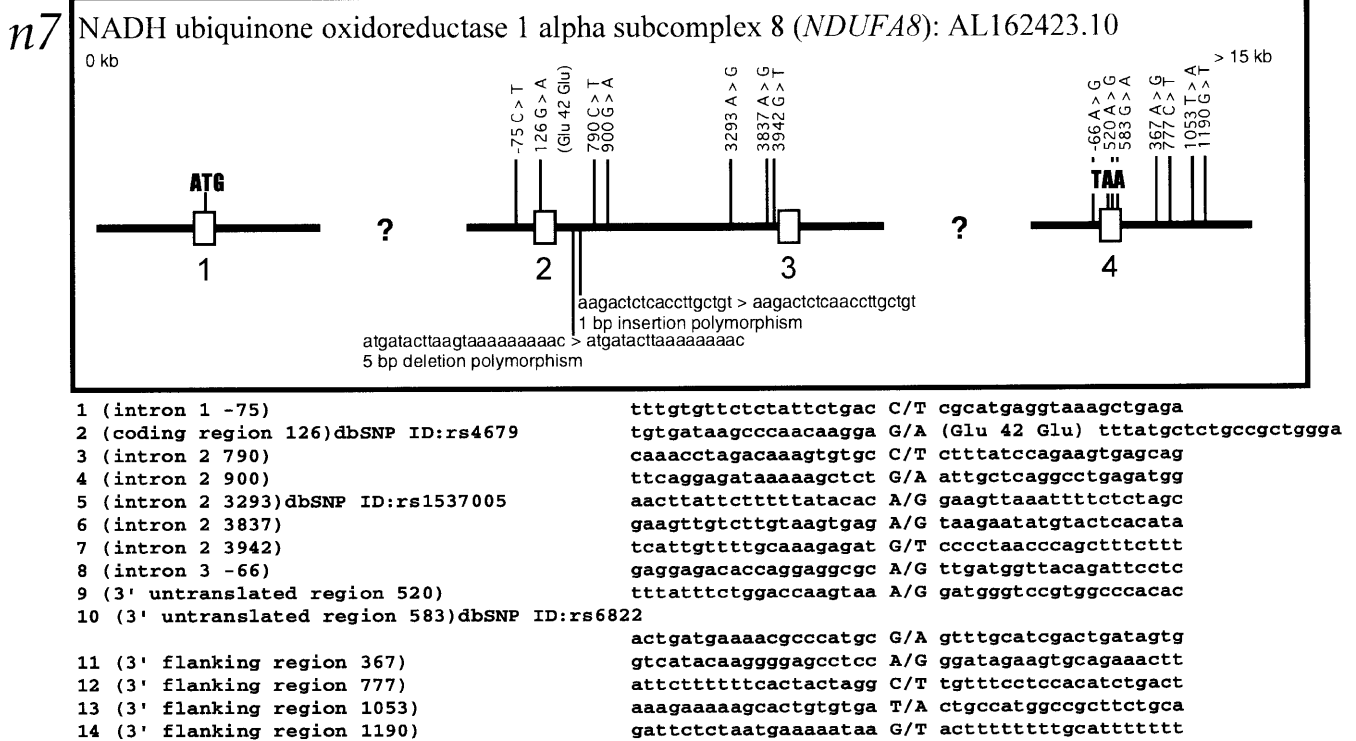
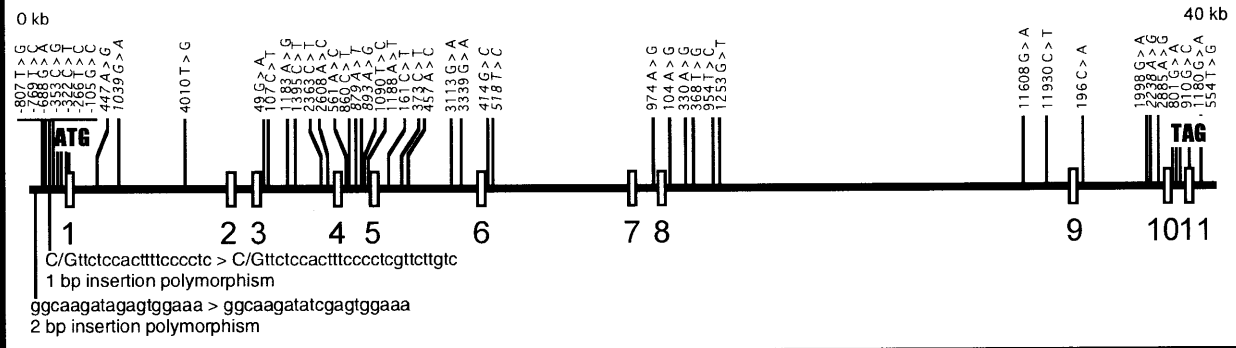


Fig. 1. Continued

n9

NADH ubiquinone oxidoreductase 1 alpha subcomplex 9 (*NDUFA9*): AC005832.1

1 (5' flanking region -807) gatggctctttgtagaacaa T/G gcagattctcaaggtgacc
 2 (5' flanking region -769) accacagttaaagaaaaaat T/C acaagccattgcgctagaga
 3 (5' flanking region -688) dbSNP ID:rs979878 cacacatggccactaaat G/A cataatatttcaggaagttc
 4 (5' flanking region -353) cacaccctattttggtttct C/G ttctccactttcccctcgt
 5 (5' flanking region -322) ttcccctcgttcttctgtccc C/T ctttctctctcctgggcc
 6 (5' flanking region -266) dbSNP ID:rs979877 cggagggtacgcctcctc T/C tcctccggactcaccagggc
 7 (5' untranslated region -105) dbSNP ID:rs714621 acaatctaccgaaaccgtc G/C tttctgggttagcgctgcy
 8 (intron 1 447) attcatatgagcacaatgga A/G atgataatattacaatacca
 9 (intron 1 1039) ggcttgatgttcagcctgag G/A caagaattaggagtgttag
 10 (intron 1 4010) aatgtatccaaaagagattc T/G cattcctgccatgaagaa
 11 (intron 3 49) gacaaatataaataactaag G/A tcatttttaggagtgatagg
 12 (intron 3 107) aatttcttcccagaatggac C/T aaaggcatcctctgttccca
 13 (intron 3 1183) atctctgtaaatatcatac A/G gattatgttaacccctta
 14 (intron 3 1395) atcctagttctttgtccct C/T aagttgttggtcacctgt
 15 (intron 3 2363) agaaaatagtcataatggc C/T ccaactaacactagctctta
 16 (intron 3 2608) gtcatttgattacctgagta A/C agtgactgttacctgtttg
 17 (intron 4 561) attttataaattctttgatg A/C cttgggggtcttattcaact
 18 (intron 4 860) attgtgtagagtaatgacag C/T agagctgtcaacttttttaa
 19 (intron 4 879) gcagagctgtcaactttttt A/T aaaaaataattttagcttaa
 20 (intron 4 893) ttttttaaaaaataatttt A/G gcttaaaaaaattaaaaatt
 21 (intron 4 1090) atcattgctgtttaaaagtt T/C aagtagtgtgaatttcagta
 22 (intron 4 1188) aaccaatccttttattttt A/T tcttccagaaactttgatt
 23 (intron 5 161) ggggtgtgtgatgttttga C/T gttttgattgattgccttct
 24 (intron 5 373) ctttctcacccttgcaactg C/T agtggttttgtgccactctt
 25 (intron 5 457) gccagggaaagatgcctattc A/C cacagtgcctatgctcctt
 26 (intron 5 3113) gatttttctccttcttcaat G/A taagcttcccttaaaataaa
 27 (intron 5 3339) tctaaactcaaacagggtt G/A tttggttattgtttaggctg
 28 (intron 6 414) tatagttttgcctttccag G/C atattacatatatgggtaga
 29 (intron 6 518) ctttcatttcttttcatagc T/C tgatagctcatttcttata
 30 (intron 7 974) ggattatgcgtacttgga A/G tacttgatagcgggtgatta
 31 (intron 8 104) dbSNP ID:rs1009633 gttgcactttacaatatca A/G aatctcctcattgcttgcac
 32 (intron 8 330) dbSNP ID:rs1009634 acaaaacacttttagggagt A/G agtatggttataaaaccaca
 33 (intron 8 368) acattaattttgatggagta T/G cacaatgcctccagaggctg
 34 (intron 8 954) gcatgcaatcagttatata T/C ctagataagaattacaatc
 35 (intron 8 1253) tcctctgaaattgtagata G/T gtactctacatttctcctc
 36 (intron 8 11608) gaaaagatagatgtataaat G/A accaaaaatcgtgaagaaa
 37 (intron 8 11930) ctacaaatatattctaaatg C/T gtaatcctggataagtaaaa
 38 (intron 9 196) dbSNP ID:rs1029767 atattggaacattaatcct C/A ccaattcccaaaagtgaatc
 39 (intron 9 1998) tgttttcaagcctttaaac G/A gctgtggaacctgtgctca
 40 (intron 9 2238) ccagctactggaggctga A/G gtggaggactcacttgagcc
 41 (intron 9 2885) acagcggctgtcttctctg C/A gttctcataggctagcttac
 42 (intron 10 801) tacactaaagtgtctcttac G/A ttatacttgagaaagtgtt
 43 (intron 10 910) tgcagacttccaggtgggta G/C gatgagggattgctgctgct
 44 (intron 10 1180) aaaactgagtcagaacgccc G/A tgctcagaaaacaggggctg
 45 (3' flanking region 554) gtgccagcacttaggaatta T/G gaccttctaataagattctt

Fig. 1. Continued

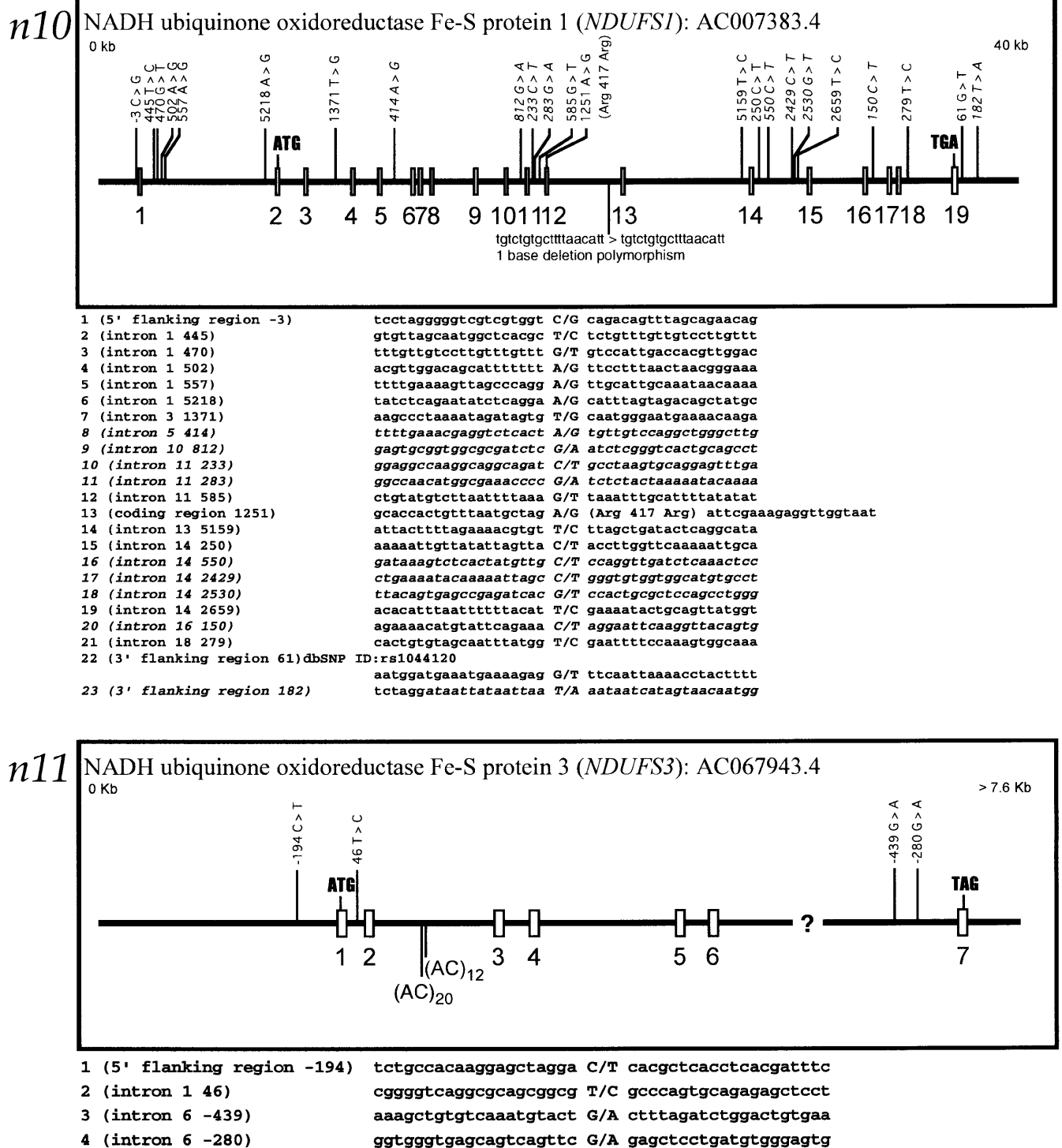
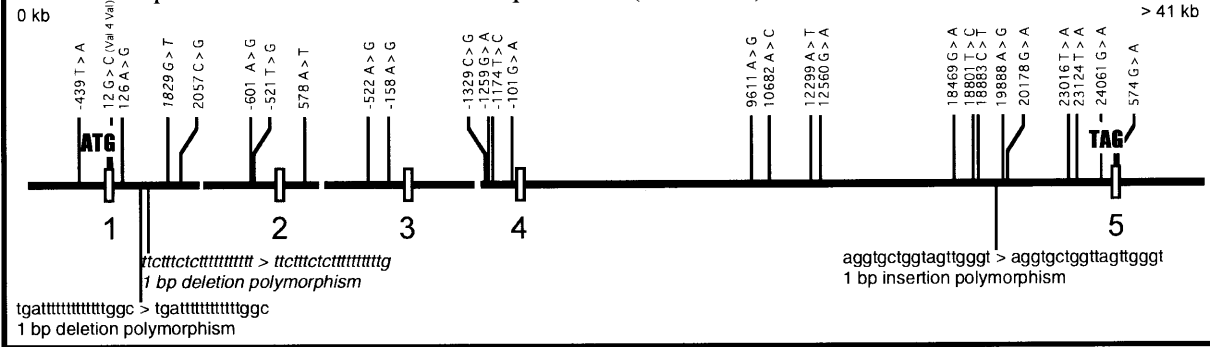


Fig. 1. Continued

n12

NADH ubiquinone oxidoreductase Fe-S protein 4 (*NDUFS4*): AC024569.3



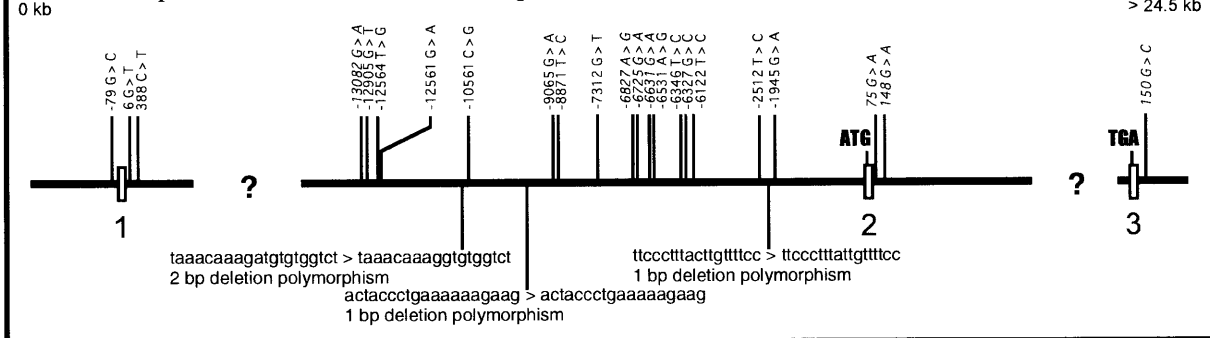
- 1 (5' flanking region -439)
- 2 (coding region 12) dbSNP ID:rs1127795
- 3 (intron 1 126) dbSNP ID:rs1532163
- 4 (intron 1 1829)
- 5 (intron 1 2057)
- 6 (intron 1 -601) dbSNP ID:rs365358
- 7 (intron 1 -521)
- 8 (intron 2 578) dbSNP ID:rs409313
- 9 (intron 2 -522) dbSNP ID:rs810858
- 10 (intron 2 -158) dbSNP ID:rs31305
- 11 (intron 3 -1329) dbSNP ID:rs451591
- 12 (intron 3 -1259)
- 13 (intron 3 -1174)
- 14 (intron 3 -101) dbSNP ID:rs1490769
- 15 (intron 4 9611) dbSNP ID:rs189849
- 16 (intron 4 10682)
- 17 (intron 4 12299)
- 18 (intron 4 12560)
- 19 (intron 4 18469) dbSNP ID:rs256094
- 20 (intron 4 18801)
- 21 (intron 4 18883) dbSNP ID:rs256093
- 22 (intron 4 19888)
- 23 (intron 4 20178)
- 24 (intron 4 23016)
- 25 (intron 4 23124)
- 26 (intron 4 24061) dbSNP ID:rs439552
- 27 (3' untranslated region 574) dbSNP ID:rs567

```

aactgaatacacagccctgtcc T/A gaggcttgcaaatgaatc
tcgacgaagatggcggcggt G/C (Val 4 Val) tcaatgctcagtggtactgag
ttgacctttcttcgggaga A/G gtcgggaggactagggaactg
gaaaaaaaaatcttaatgcca G/T ggaagcgtttttaaatac
attaatgggaaatctacat C/G taaaattctttttatgttaa
taagtaattggagtgaagt A/G ggaataatgctaataaaagg
ttcatttaactaattttat T/G tctccattttgtgaatggg
gcagccatcactataaaat A/T ttttagagtgggtctgtta
ttaaagtgttaaatcca A/G ctttcaaaaaggctcaggt
gtatcctaaatagtaattac A/G ttgtaaagatcagaatggt
tctctgtctcagtatatata C/G ttaactatcaactggaaaa
ataaaattatgatattatta G/A tactaatatagccagccata
aatatataaattatagaa T/C ctcagatagcaaccatggt
taaaatttccatcaataat G/A tactatttggatattatag
tcaaatcaaatgtttatca A/G ttgtattgagatacagttg
cacaatatagggcaaaact A/C ctaccaagcactaacaagt
tttactatagatatatgg A/T atagatagagtatctct
accaaataaggtattatgca G/A gctcatctttttatataaga
tgtgttctaaatttcaatg G/A ctgatttctcagataactt
ggaaagacttgccttggcag T/C gtatccgaaacctctgttat
cattattctgaaccatttca C/T ccctactgtacccaccccc
tcgcacagctgagaagagca A/G gggctggttttcagtaacc
agaaaagatgagataatc G/A tctaacttaccattcttaa
ctactctgtgaaagtaagg T/A atgttgacaagtaaataa
actttcttggagatggagt T/A ccagcagttgggaatgtaat
atacgtttacatattttct G/A tatcccttggagagatcc
gactgtgaaataagtcagct G/A tgcagatttatagtcctag
    
```

n13

NADH ubiquinone oxidoreductase Fe-S protein 5 (*NDUFS5*): AL139015.5



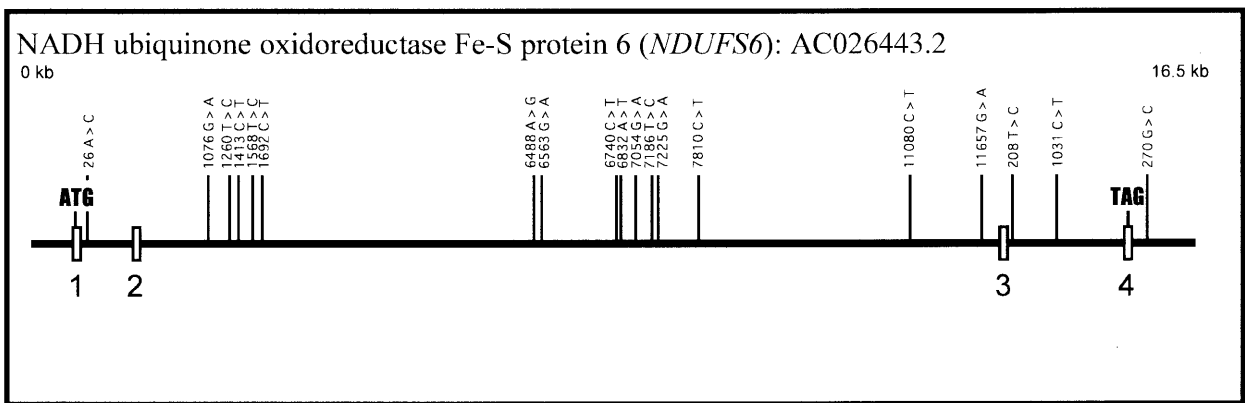
- 1 (5' flanking region -79) dbSNP ID:rs1122471
- 2 (intron 1 6) dbSNP ID:rs1122472
- 3 (intron 1 388)
- 4 (intron 1 -13082)
- 5 (intron 1 -12905)
- 6 (intron 1 -12564)
- 7 (intron 1 -12561)
- 8 (intron 1 -10561)
- 9 (intron 1 -9065)
- 10 (intron 1 -8871)
- 11 (intron 1 -7312)
- 12 (intron 1 -6827)
- 13 (intron 1 -6725)
- 14 (intron 1 -6631)
- 15 (intron 1 -6531)
- 16 (intron 1 -6346)
- 17 (intron 1 -6327)
- 18 (intron 1 -6122)
- 19 (intron 1 -2512)
- 20 (intron 1 -1945)
- 21 (intron 2 75)
- 22 (intron 2 148)
- 23 (3' flanking region 150)

```

ggctcgggtttctatggtttt G/C tggggagacggctgagggaa
acgagcatcgggttaggtagg G/T gctgcttatgtccgcgcggc
ccaaacatagccagcacttc C/T ggctgtaactccgggctggt
agttagccgagattgcacca G/A tgcatccagcctgggcaac
gttttcaacaaggactcca G/T agtagtagagaagtctctgt
atttcatcacacctcaact T/G aaggataaacagccttaaga
tcatcacacctcaactaa G/A gataaacagccttaagaatg
aacaatgtggtatagtggtgg C/G ggggtggtgagcaggtgtcat
cctgatgctcctggctccag G/A gtgaccttttccctttaga
tcaccacgtgtctgtagata T/C aggcaccgagaccttcgctt
aaatccttggcttctagaat G/T ggtcactgatggtatataat
aacctctgcctccccgattc A/G cgccattctcctgcctcagc
agttagagacgggtttcacc G/A tgttagccagcatgggtctcg
aggcgtgagccactgcggcc G/A gcctagaccttcttctata
cccaacagctcccaatgtaa A/G acagatctattaatattctg
gcaacagatcttgacctata T/C cccatagggtacagctgagg
atccatagggtacagctga G/C gactttaatcagaaaaggag
tagccttgcctttactctac T/C gttcctcccaaatcacaccc
acaaactctaatgcaaat T/C tgcagatcaaagtggtgctta
tttaactcctttaaatttc G/A caatttcacaacctagggta
ttttttttttttttggagac G/A aagtctactctgtccccct
ctgtagcctctgcctccag G/A ttcaggcgattcgcgtacct
cagatcgaagtgttctcct G/C cctcagcctcccaagtagct
    
```

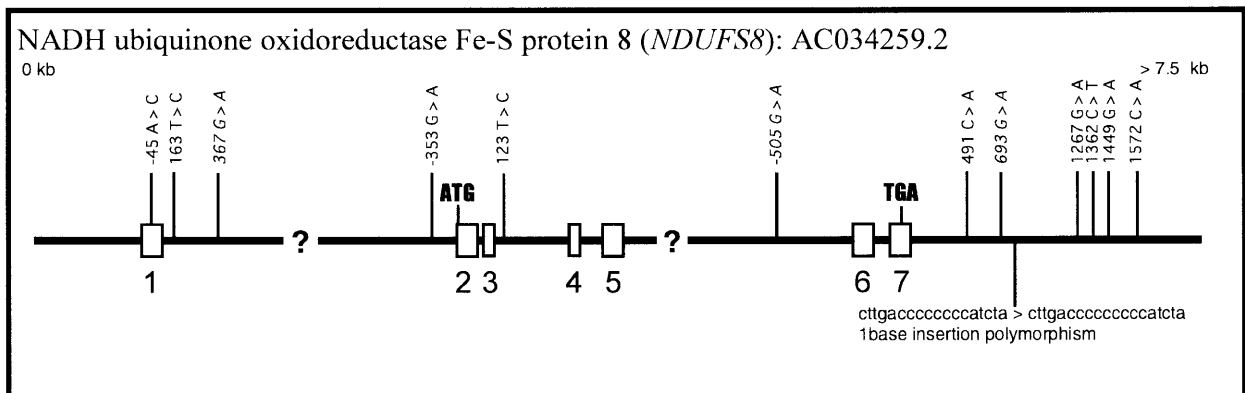
Fig. 1. Continued

n14 NADH ubiquinone oxidoreductase Fe-S protein 6 (*NDUFS6*): AC026443.2



| | | | |
|-----------------------------|-----------------------|-----|----------------------|
| 1 (intron 1 26) | ggccgctgggtacaggatgc | A/C | ccttcctccagccgcacctc |
| 2 (intron 2 1076) | ggatcatggtgggtggagagg | G/A | gcttgtgtctggtgggtttg |
| 3 (intron 2 1260) | cagttgtcagtaagtgtgtg | T/C | atagggtaatgctctcttct |
| 4 (intron 2 1413) | caaaggagctcatggcattg | C/T | gaatgggacatttcttccgt |
| 5 (intron 2 1568) | tggagaaggggaggtttctc | T/C | tagtgtggatgcggtatggt |
| 6 (intron 2 1692) | gaccgtggtgacggagggtt | C/T | ctgggcatcgatgggtggtt |
| 7 (intron 2 6488) | tagcttaaataaattattggc | A/G | tccatgttcagaatgcttga |
| 8 (intron 2 6563) | tttaaaacttttattttaaat | G/A | tccatgaatggggtcggtat |
| 9 (intron 2 6740) | aaagatttaaacctacatat | C/T | tttatgcccaatcatttgat |
| 10 (intron 2 6832) | gcgagggactcattttacag | A/T | ggttggacacttcactgtgt |
| 11 (intron 2 7054) | ttcactgcccggagcttggcc | G/A | tgtgaaccggagccgggct |
| 12 (intron 2 7186) | ggtcagggtcacccttgagc | T/C | gcgcacactaaatgacggga |
| 13 (intron 2 7225) | gagggcatcccgcgtcagtc | G/A | ccagtgtcgaggcgtcagca |
| 14 (intron 2 7810) | cttcactctggggcgggga | C/T | gctgtagaaggagcacaag |
| 15 (intron 2 11080) | gtaactgttcagtgtttct | C/T | ctttggatttcatgtaaac |
| 16 (intron 2 11657) | gggacagaacgatgtggtgg | G/A | gagaagagggcgtggcagag |
| 17 (intron 3 208) | cgaaaaaccctttcaactg | T/C | gaagtgtgggcggcatggt |
| 18 (intron 3 1031) | ctagagtgggactgggcacc | C/T | ggcatgtcccctcctgggct |
| 19 (3' flanking region 270) | gcttcagagagccaaggtgg | G/C | tcttgaggtgcatagtgaag |

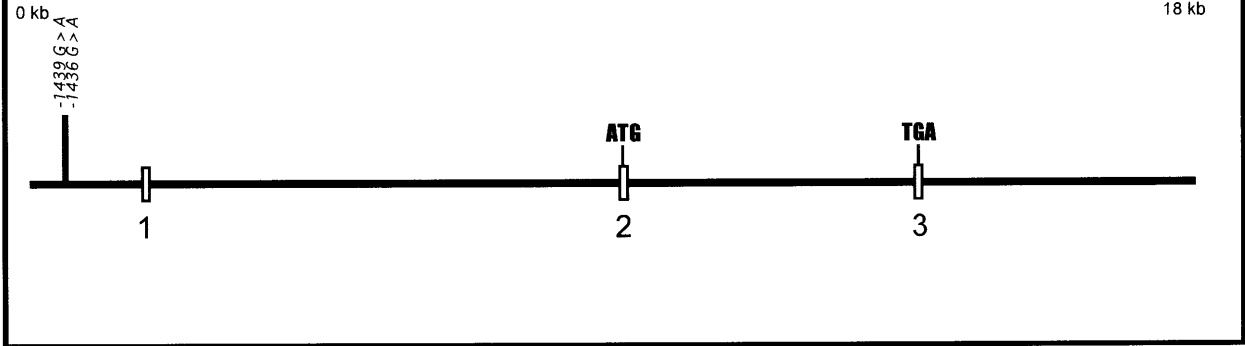
n15 NADH ubiquinone oxidoreductase Fe-S protein 8 (*NDUFS8*): AC034259.2



| | | | |
|-------------------------------------|-----------------------|-----|----------------------|
| 1 (5' untranslated region -45) | agtgtagcctccgcctcccg | A/C | ttgactggcctgcttggcaa |
| 2 (intron 1 163) | agggtcagcggggagccggc | T/C | ctcagggcgcgatgcgccgc |
| 3 (intron 1 367) dbSNP ID:rs1122731 | ctctgggggtgcccttggcat | G/A | tacgggctcccctaattccc |
| 4 (intron 1 -353) dbSNP ID:rs999571 | cgctctcagtcgatgtggc | G/A | ggggctggtaatgttgtttt |
| 5 (intron 3 123) | tctctgagcctgtttccact | T/C | ttaaaatgattatggtgatg |
| 6 (intron 6 -505) | aggcaagggcagggcggcac | G/A | gtggctcacgcttgaatcc |
| 7 (3' flanking region 491) | ggcctgagctggcctgcgt | C/A | cagccacatcctcttctctg |
| 8 (3' flanking region 693) | ttcacttcatttgagtgag | G/A | aaaccagctccgagaggtga |
| 9 (3' flanking region 1267) | ttttccagacgtaaccgcc | G/A | tcagagcgtggcatggagcc |
| 10 (3' flanking region 1362) | cgctgggttctttcccttac | C/T | gtggctcccaggcacttac |
| 11 (3' flanking region 1449) | tgtcagaacaggcctatggc | G/A | cccaccacaagtcccccaa |
| 12 (3' flanking region 1572) | cagccccacaggcctgtgct | C/A | gctgtgtgggcttagggat |

Fig. 1. Continued

n16

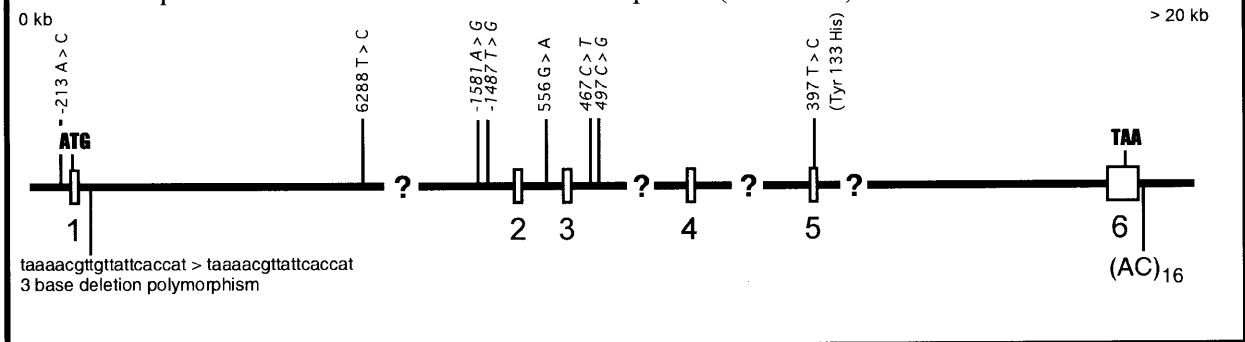
NADH ubiquinone oxidoreductase 1 beta subcomplex 3 (*NDUFB3*): AC007272.3

1 (5' flanking region -1439)

2 (5' flanking region -1436)

ttaaagttgacttttttct G/A cc G/A ggcacggtggctcacgcctg

n17

NADH ubiquinone oxidoreductase 1 beta subcomplex 5 (*NDUFB5*): AC068361.2

1 (5' flanking region -213)

2 (intron 1 6288)

3 (intron 1 -1581)

4 (intron 1 -1487)

5 (intron 2 556)

6 (intron 3 467)

7 (intron 3 497)

8 (coding region 397)

ggcggatgaaactctcctac A/C aagaagggccaaccggccg

ggggatgttgattacctagg T/C cagtaaagtaagaaggcat

cttctgggccactgtatcct A/G tttctttcccttgttaccct

ccctcttagaccgtatatag T/G tctagcataggatctgcaca

ttgtctggaccatctgccac G/A gtagataaagctctgaatca

ggcggccatcgccactccagcc C/T gggcaacagagtgagactct

agtgagactctgtcccccc C/G caaaaaaaaaactataatcct

atgatagtcctgaaagata T/C (Tyr 133 His) atgaaagaacaatggccgtc

Fig. 1. Continued

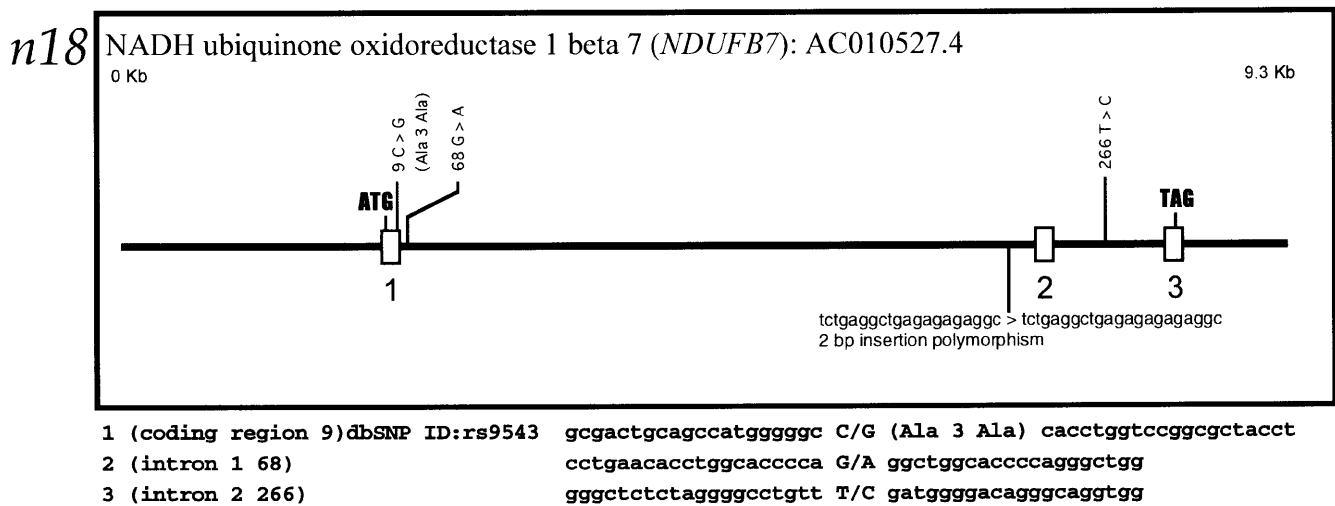


Fig. 1. Continued

NADH ubiquinone oxidoreductase family

Respiratory NADH ubiquinone oxidoreductases (also known as complex I, which is embedded in the inner mitochondrial membrane; EC1.6.5.3) serve to dehydrogenate NADH and to shuttle electrons to coenzyme Q. This electron transport generates a proton gradient across the inner mitochondrial membrane, to provide the proton-motive force for ATP synthesis (Guenebaut et al. 1998). The total human complex consists of 42 or more subunits, seven of which are encoded by the mitochondrial genome and the remainder by the nuclear genome. So far, complete gene structures have been reported for six of the 35 nuclear members of this family, and chromosomal localization is known for 26 of the 35 (for reviews see Ton et al. 1997; Loeffen et al. 1998; Smeitink and van den Heuvel 1999). We analyzed 18 complex I genes by direct sequencing, based on their cDNA sequences (Fig. 1. n1–n18). A total of 241 SNPs were identified from the 18 loci: 11 in coding elements, 22 in 5' flanking regions, four in a 5' untranslated region, 170 in introns, five in 3' untranslated regions, and 29 in 3' flanking regions. Of these, 187 (78%) were novel and 38 were located within regions corresponding to repetitive sequences predicted by the RepeatMasker program. Of the 11 SNPs identified within coding elements, four (Leu81Val in the *NDUFA3* gene, Ala9Val in the *NDUFA6* gene, Pro66Ala in the *NDUFA7* gene, and Tyr133His in the *NDUFB5* gene) were nonsynonymous and had not been reported previously. With regard to distribution, we found 11 SNPs within the 4.4-kb *NDUFA2* genomic locus, while only two were detected in the 18-kb genomic sequence containing the *NDUFB3* gene.

The SNP collection reported here provides genetic data that should be helpful for personalized medical service and also for identifying genes associated with drug efficacy and/or side effects. We believe that, in the near future, using an SNP-based approach to predict individual differences in drug efficacy or toxicity on the basis of genetic factors will

be a realistic approach to treatment of human diseases or toxic conditions. Finally, we hope that the virtual experiments made possible by our catalog will accelerate certain aspects of human genetic research.

References

- Ashmarin IP, Danilova RA, Obukhova MF, Moskvitina TA, Prozorovsky VN (2000) Main ethanol metabolizing alcohol dehydrogenases (ADH I and ADH IV): biochemical functions and physiological manifestations. *FEBS Lett* 486:49–51
- Duester G, Farres J, Felder MR, Holmes RS, Hoog JO, Pares X, Plapp BV, Yin SJ, Jornvall H (1999) Recommended nomenclature for the vertebrate alcohol dehydrogenase gene family. *Biochem Pharmacol* 5:389–395
- Guenebaut V, Schlitt A, Weiss H, Leonard K, Friedrich T (1998) Consistent structure between bacterial and mitochondrial NADH:ubiquinone oxidoreductase (complex I). *J Mol Biol* 276:105–112
- Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 22:239–247
- Hayes JD, Pulford DJ (1995) The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit Rev Biochem Mol Bio* 30:445–600
- Iida A, Sekine A, Saito A, Kitamura Y, Kitamoto T, Osawa S, Mishima C, Nakamura Y (2001) Catalog of 320 single nucleotide polymorphisms (SNPs) in 20 quinone oxidoreductase and sulfotransferase genes. *J Hum Genet* 46:225–240
- Kruglyak L (1997) The use of a genetic map of biallelic markers in linkage studies. *Nat Genet* 17:21–24
- Loeffen JL, Triepels RH, van den Heuvel LP, Schuelke M, Buskens CA, Smeets RJ, Trijbels JM, Smeitink JA (1998) cDNA of eight nuclear encoded subunits of NADH:ubiquinone oxidoreductase: human complex I cDNA characterization completed. *Biochem Biophys Res Commun* 253:415–422
- McCarthy JJ, Hilfiker R (2000) The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nat Biotechnol* 18:505–508
- Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 25:2745–2751
- Ohnishi Y, Tanaka T, Yamada R, Suematsu K, Minami M, Fujii K, Hoki N, Kodama K, Nagata S, Hayashi T, Kinoshita N, Sato H, Sato

- H, Kuzuya T, Takeda H, Hori M, Nakamura Y (2000) Identification of 187 single nucleotide polymorphisms (SNPs) among 41 candidate genes for ischemic heart disease in the Japanese population. *Hum Genet* 106:288–292
- Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516–1517
- Saito S, Iida A, Sekine A, Eguchi C, Miura Y, Nakamura Y (2001) Seventy genetic variations in human microsomal and soluble epoxide hydrolase (EPHX1 and EPHX2) genes in the Japanese population. *J Hum Genet* 46:325–329
- Sekine A, Saito S, Iida A, Mitsunobu Y, Higuchi S, Harigae S, Nakamura Y (2001) Identification of single nucleotide polymorphisms (SNPs) of human N-acetyltransferase genes NAT1, NAT2, AANAT, ARD1 and L1CAM in the Japanese population. *J Hum Genet* 46:314–319
- Sherry ST, Ward M, Sirotkin K (2000) Use of molecular variation in the NCBI dbSNP database. *Hum Mut* 15:68–75
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29:308–311
- Smeitink J, van den Heuvel L (1999) Human mitochondrial complex I in health and disease. *Am J Hum Genet* 64:1505–1510
- Strange RC, Jones PW, Fryer AA (2000) Glutathione S-transferase: genetics and role in toxicology. *Toxicol Lett* 112–113:357–363
- Ton C, Hwang DM, Dempsey AA, Liew CC (1997) Identification and primary structure of five human NADH-ubiquinone oxidoreductase subunits. *Biochem Biophys Res Commun* 241:589–594
- Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077–1082