Carcinoembryonic Antigens: Alternative Splicing Accounts for the Multiple mRNAs that Code for Novel Members of the Carcinoembryonic Antigen Family

Thomas R. Barnett,* Axel Kretschmer,[‡] Douglas A. Austen,[‡] Scott J. Goebel,* John T. Hart,[‡] James J. Elting,* and Michael E. Kamarck[‡]

*Molecular Diagnostics, Inc. and ‡Molecular Therapeutics, Inc., Miles Research Center, West Haven, Connecticut 06516

Abstract. The recent cloning of complete cDNAs encoding carcinoembryonic antigen (CEA) and nonspecific cross-reacting antigen has revealed the existence of a new gene family belonging to the immunoglobulin gene superfamily. We have reported the isolation of a partial CEA cDNA and of L-cell transfectant cell lines that express human antigens cross-reactive with commercial antibodies directed to native CEA (Kamarck, M., J. Elting, J. Hart, S. Goebel, P. M. M. Rae, J. Nedwin, and T. Barnett. 1987. *Proc. Natl. Acad. Sci. USA.* 84:5350–5354). In this study, we describe the identification and cloning of 3.9-, 3.7-, 2.2-, and 1.8-kb cDNAs and a 23-kb genomic transcription unit, which code for new members of the CEA gene family. DNA sequence analysis

ARCINOEMBRYONIC antigen (CEA)¹ is a 180-kD glycoprotein that has been used extensively as a serum marker in some cancers, particularly those of colorectal, breast, and lung origin. Despite its initial promise as a tumor-specific marker, CEA has been shown to be a member of a family of 8-10 cross-reactive isoantigens which can be detected in a variety of normal and tumor tissue types (Shively and Beatty, 1985). CEA and nonspecific crossreacting antigen (NCA) have recently been shown to have considerable structural and sequence similarity (Engvall et al., 1978; Paxton et al., 1987; Thompson et al., 1987), thus explaining their observed immunological cross-reactivity among anti-CEA antibodies (von Kleist et al., 1972; Darcy et al., 1973). This cross-reactivity extends to all members of the CEA isoantigen family and has made interpretation of diagnostic immunoassays for CEA both difficult and inaccurate.

of these cloned DNAs establishes the existence of a set of four alternatively spliced mRNAs which are expressed in several tumor cell lines, in human fetal liver, and in L-cell transfectants. Deduced amino acid sequences of the encoded isoantigens show extensive similarity to CEA and nonspecific cross-reacting antigens, but in addition demonstrate transmembrane and cytoplasmic domains. We designate members of this antigen family transmembrane CEAs. The transmembrane CEA isoantigens share general structural characteristics with members of the immunoglobulin gene superfamily and can be specifically compared to the cell adhesion molecules, N-CAM (neural cell adhesion molecule) and MAG (myelin-associated glycoprotein).

Recently, the complete cDNAs for both CEA and NCA have been isolated and sequenced (Beauchemin et al., 1987; Oikawa et al., 1987; Barnett et al., 1988; Neumaier et al., 1988; Tawaragi et al., 1988). These antigens display a high degree of sequence similarity and also maintain a repeating series of disulfide loops with features of immunoglobulin superfamily members (Hunkapiller and Hood, 1986; Williams and Barclay, 1988). Both CEA and NCA have short hydrophobic peptide segments at their carboxyl termini, but appear to be anchored in the cell membrane by a glycolipid moiety (Takami et al., 1988).

We have previously isolated an L-cell transfectant line, 23.4 11+, and a bacteriophage λ gtl1 partial cDNA clone (λ cLV7) that express epitopes of CEA (Kamarck et al., 1987). The partial cDNA clone encompasses the second and part of the third loop domains of CEA (nucleotides 1,084–1,916 in Barnett et al., 1988) and identifies several major Eco RI and Bam HI segments in genomic DNA. All of these genomic segments map to human chromosome 19, suggesting the existence of a CEA gene locus consisting of an extensive family of related members. In contrast, cDNA from λ cLV7 hybridized to a single genomic segment in L-cells that had been transfected with human genomic DNA and selected for CEA expression using commercial antibod-

Axel Kretschmer's present address is BAYER AG, ZF-F Biotechnologie Q18, D-5090 Leverkusen, FRG.

^{1.} Abbreviations used in this paper: CEA, carcinoembryonic antigen; MAG, myelin-associated glycoprotein; NCA, nonspecific cross-reacting antigen; N-CAM, neural cell adhesion molecule; TM-CEA, transmembrane carcinoembryonic antigen.

ies and the fluorescence-activated cell sorter. These transfectants expressed CEA family members distinguishable from both CEA and NCA based on the molecular sizes of their antigens and their messenger RNAs (Kamarck et al., 1987; Barnett et al., 1988).

In this study, we describe the cloning of a genomic transcription unit and of cDNAs for alternatively spliced mRNAs that code for four previously undescribed CEA isoantigens expressed by a number of tumor cell lines and by normal fetal liver. Despite their extensive amino acid sequence similarities, the four protein species differ from CEA and NCA in that they contain transmembrane and cytoplasmic domains. For this reason we have designated these molecules transmembrane CEAs (TM-CEAs). With the identification of this new class of CEA glycoproteins, we can now account for six of the 8–10 members of the CEA isoantigen family.

Materials and Methods

Cells and Cell Lines

Human tumor cell lines were obtained from the American Type Culture Collection (Rockville, MD) and grown according to specifications. Line designations are LoVo (CCL 229, colon adenocarcinoma), HT-29 (HTB 38, colon adenocarcinoma), BeWo (CCL 98, choriocarcinoma), SW403 (CCL 230, colon adenocarcinoma), KG-1 (CCL 246, acute myelogenous leukemia), SCaBER (HTB 3, squamous bladder carcinoma), and MIA PaCa-2 (CRL 1420, human pancreatic carcinoma). The L-cell transfectant line, 23.4 11+, has been previously described (Kamarck et al., 1987).

DNA Preparation and Analysis

DNA was extracted from cells essentially as described by DiLella and Woo (1987). DNA (10 μ g) was digested at 37°C for 6-8 h with restriction endonucleases (Amersham Corp., Arlington Heights, IL) and subjected to electrophoresis through horizontal 0.8% agarose gels before transfer to ny-Ion membranes. For the preparation of genomic DNA for library construction or subcloning, 100-200 μ g of DNA was digested either with excess Eco RI or with a predetermined amount of Mbo I (mean DNA size produced = 12-15 kb) and subjected to electrophoresis overnight through horizontal 0.5% low melting temperature agarose gels for optimal separation. Appropriate size classes were excised with a razor blade, gel slices were melted in 0.3 M NaOAc, pH 5, and 0.1 M Tris-HCl, pH 7.8, DNA was purified by phenol extraction, precipitated with 95% ethanol, and resuspended in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA.

RNA Preparation and Analysis

Cell pellets stored at -80° C were thawed in the presence of 0.14 M NaCl, 1.5 mM MgCl₂, 10 mM Tris-HCl, pH 7.5, 0.5% NP-40 (Sigma Chemical Co., St. Louis, MO), 4 mM DTT, and 10 u/ml of human placental ribonuclease inhibitor (Stratagene Inc., San Diego, CA). Cells were fractionated by 20 strokes of a low clearance Dounce-type homogenizer and in-cubated in 0.1% Na deoxycholate for 10 min on ice. Cytoplasmic and nuclear fractions were separated by centrifugation for 20 min at 12,000 g, and the supernatant was added to an equal volume of 200 mM Tris-HCl, pH 7.5, 25 mM EDTA, 300 mM NaCl, 2% SDS, and 400 μ g/ml of proteinase K. After incubation at 45°C, nucleic acids were extracted with phenol/chloroform (1:1 [vol/vol]), and ethanol precipitated in the presence of added NaOAc to 0.15 M. Poly A+ RNA was prepared by oligo dT column chromatography as described by Aviv and Leder (1972). For Northern analysis, nucleic acids were separated on 2.2 M formaldehyde–0.8% agarose gels as described by Lehrach et al. (1977), then transferred to nylon membranes.

Probe Preparation and Nucleic Acid Blots

Mini-preparations of plasmid DNA (Maniatis et al., 1982) were digested with restriction enzymes and fractionated by size on 1% agarose gels. Appropriate segments were excised from the gel and used directly for random priming (Feinberg and Vogelstein, 1984) in the presence of α^{-32} P-dCTP (6,000 Ci/mmol; Amersham Corp). Unincorporated nucleotide was removed by spun column chromatography over Sephadex G-50 (Pharmacia Fine Chemicals, Piscataway, NJ). Probes were used at 10^6 cpm/ml in $2\times$ SSPE, $5\times$ Denhardt's, 6% SDS at 68°C for 24-48 h. Posthybridization washes were in 0.2× SSPE, 0.25% SDS at 68°C.

cDNA Library Construction

Poly A+ RNAs from 23.411+ transfectant cells and from HT-29 cells were converted to double-stranded DNA essentially by the method of Gubler and Hoffman (1983). After Eco RI methylase treatment, DNA was fractionated on 0.8% LGT agarose gels, size selected based on identification of RNA from Northern blots, and eluted by melting (see above). Eco RI linkers were added and cDNAs were inserted into Eco RI-cleaved arms of ligated \laglet10 DNA. Phage with inserts were selected on the Hfl⁺ host, NM514. Screening of phage was by the method of Benton and Davis (1977) using radiolabeled DNA probes. Purified positive phage DNAs were cleaved with Eco RI, then inserts were fractionated on 1% LGT agarose gels and ligated to 5 ng of Eco RI-cleaved plasmid Bluescript KS+ (Stratagene). Plasmid minipreps were used as source of DNA for deriving nested deletion sets of cDNA inserts by exonuclease III/mung bean nuclease treatment (Henikoff, 1984) and/or for primer or oligonucleotide directed dideoxy sequencing on double-stranded DNA by the method of Sanger et al. (1977). Computer-aided analysis of DNA sequence information was performed using the Pustell program (IBI) or the series of programs published by the Genetics Computer Group (Madison, WI).

Genomic Library Construction

DNA segments of 12–15 kb from Mbo I partial digests were inserted into left and right arms of Bam HI-cleaved λJam phage DNA. This Aam Bam bacteriophage vector was constructed by substituting the 5.0-kb cosL-Ava I segment of λ Charon 4A phage DNA for the equivalent segment of λ Il DNA (Mullins et al., 1984) and selecting phage on su⁻/su⁺ hosts. A library of >10⁶ independent recombinants was obtained after packaging the ligated λJam and transfectant DNAs in vitro and plating on *Escherichia coli* host LE392. Positive phage were purified, DNAs were prepared, and restriction segments were separated on 1% agarose gels before blotting on nitrocellulose membranes according to Southern (1975). Appropriate segments were subcloned into Bluescribe (–) or Bluescript KS+ or SK+ plasmids for double-stranded DNA sequencing.

Western Blot Analysis

Peptide/N-glycanase was purified from culture supernatants of Flavobacterium meningosepticum by the method of Tarentino et al. (1985). No endoglycosidase F or endogenous protease activity could be demonstrated in this preparation. Transfectant 23.4 11+ or normal mouse L cells were lysed in phosphate buffer containing Triton X-100 (PBS-T; 10 mM NaPO₄, pH 7.3, 140 mM NaCl, 10 mM EDTA, 10 mM benzamidine, 1 mM difluorophosphate, and 1% Triton X-100) by sonication at 4°C. Aliquots of lysate containing up to 100 μ g of total protein were brought to 1% SDS and 1 mM β-mercaptoethanol and heated to 100°C for 3 min. After cooling, the denatured lysates were mixed with an equal volume of 2× endo F digestion buffer ($1 \times = 0.2$ M NaPO₄, pH 8.7, 0.75% NP-40, 5 mM EDTA, 5 mM o-phenanthrolene, and 0.5 mM PMSF). Peptide/N-glycanase (0.4 µg) was added and digestion was performed for 3 h at 37°C. An additional 0.1 μ g of enzyme was added and digestion was continued for an additional hour. Controls for enzyme reactivity included the complete deglycosylation of an α_1 -acid glycoprotein. Reactions were terminated by the addition of 1/3 vol of 4× SDS-PAGE sample buffer, then heated for 7 min at 100°C. Digested samples were analyzed by SDS-PAGE 10-20% gradient gels (Laemmli, 1970). Proteins were transferred to nitrocellulose by the method of Towbin et al. (1979). Membranes were blocked for 1 h in PBS-T containing 5% BSA. Rabbit anti-CEA Ig (DAKOPATTS, Copenhagen, Denmark) was used at 2 µg/ml and goat anti-rabbit Ig alkaline phosphatase conjugate (Promega Biotec, Madison, WI) was used at a dilution of 1:7,500. Detection was by incubation of the membrane in 100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM MgCl₂ containing 66 mg/ml of nitro blue tetrazolium and 33 mg/ml of 5-bromo-4-chloro-3-indolyl phosphate.

Isolation of TM-CEA cDNAs

Northern blot analysis was performed on poly A+ RNA from the L-cell transfectant line, 23.4 11+ that had been selected for the expression of CEA isoantigens (Kamarck et al., 1987). Use of the CEA partial cDNA insert, LV7, as probe identified multiple hybridizing RNA bands with approximate molecular sizes of 3.9, 3.7, 2.2, and 1.8 kb (Fig. 1, lane a). This pattern also recurs in the colon adenocarcinoma cell line HT-29 (Fig. 1, lane b), which in addition expresses mRNAs for CEA (4.0 and 3.6 kb) and NCA (3.0 kb) (Barnett et al., 1988). To clone the unidentified mRNAs, 3-4-kb double-stranded cDNA from 23.4 11+ poly A+ RNA was size selected and inserted into λ gt10. A library of 8×10^{5} pfu was screened by LV7 hybridization. 13 phage were isolated, and the subcloned inserts from three of these, tc16-19 (3.6 kb), tc19-22 (3.5 kb), and tcE22 (3.3 kb), were analyzed by restriction endonuclease mapping. Insert tcE22 differed from the others by an apparent internal deletion of \sim 300 bp. Additional cDNAs were prepared from an HT-29 cDNA library that was size selected in the 1-3-kb region.

Based on the observation by Gold and Freedman (1965) that CEA is expressed during fetal development, we also screened a cDNA library (2×10^5 pfu; Clontech Laboratories Inc., Palo Alto, CA) prepared from normal human fetal liver poly A⁺ RNA with the CEA probe, LV7. Five positive recombinant phage were isolated and the inserts were subcloned into plasmid vectors. Two overlapping inserts, cFL4 and cFL5, together comprised a 3.6-kb fetal liver cDNA identical in restriction map to tcl6–19 and tcl9–22.

Fig. 2 shows the compilation of DNA sequences and deduced amino acid sequences from the inserts of plasmids ptc16–19, ptc19–22, ptcE22, pcFL4, and pcFL5. These inserts together provided the longest open reading frame (nucleotides 73 to 1656) and the longest cDNA (3464 nucleotides). The compiled sequence codes for an apoprotein of 526 amino acids (M_r of 58 kD) with 19 potential extracellular N-linked glycosylation sites (Asn-X-Thr/Ser).

The coding sequence of the proposed apoprotein (designated TM1-CEA) demonstrates marked similarity to sequences found in CEA and NCA cDNA clones. This is indi-



Figure 1. Expression of CEA-related poly A+ RNAs. One-tenth microgram of CEA-transfectant 23.4 11+ poly A+ RNA (lane a) and two micrograms of colon carcinoma cell line HT-29 poly A+ RNAs (lane b) were subjected to electrophoresis on a 2.2 M formaldehyde-0.8% agarose gel, transferred to nylon membrane, and hybridized with ³²P-labeled LV7 DNA. cated by the proposed domain structure illustrated in Fig. 3 and by the sequence relationships detailed in Table I. As is typical for other CEA isoantigens, the TM1-CEA polypeptide initiates with a 34-amino acid leader sequence that ends at a short side-chain amino acid, alanine. The proposed NH₂-terminus of TM1-CEA contains glutamine as the first residue that initiates at 108-amino acid NH₂-terminal domain. This domain is identical in length and highly similar in sequence to the comparable domain of CEA and NCA (Fig. 3 and Table I). This is followed by a disulfide-loop domain (LD) that is a characteristic feature of CEA and is ~178 amino acids in length (Oikawa et al., 1987).

This domain differs from the single disulfide loop domains of immunoglobulins in that each one contains four cysteine residues capable of forming 47- and 39-amino acid disulfidelinked loops. Loop domains are repeated three times in CEA and once in NCA (Fig. 3). The disulfide-loop domain size for TM1-CEA is similar to those of CEA and NCA, and the position of its cysteines is strictly conserved. In contrast to the domains of CEA and NCA, LD I of TM1-CEA has a fifth cysteine (residue 308; Fig. 2) that may not be involved in the formation of an intrachain disulfide loop.

An interesting structural region follows the conserved 177 amino acid loop-domain of TM1-CEA. This region, extending from nucleotides 1,030 to 1,340, codes for \sim 100 amino acids and has the appearance of a "half" disulfide-loop domain (*LD IIa* of Fig. 3). This is indicated by comparison of amino acids that initiate loop domain regions in CEA, NCA, and TM1-CEA and by the characteristic positioning of its only two cysteine residues to form a 47-amino acid disulfide loop. While each of the disulfide-loop domains of the three CEA classes is highly conserved (\sim 70% amino acid similarity overall), LD IIa of TM1-CEA retains \sim 44% amino acid similarity compared to TM1-CEA LD I (Table I).

A 32-amino acid hydrophobic region comes after LD IIa and bears sequence similarity to the short hydrophobic tail that terminates CEA and NCA (Table I). Hydropathy plot analysis supports the inference that this is a membranespanning region (data not shown). By contrast, the carboxylterminal region of TMI-CEA is composed of a hydrophilic stretch of 71 amino acids that terminates at residue 526 (nucleotide 1667).

Identification of TM-CEA cDNAs Derived from Alternatively Spliced mRNAs

Four different molecular sizes of CEA-related mRNA are detected in transfectant cell line 23.4 11+ and are observed in tumor cell line HT-29 (Fig. 1). Additional cDNAs isolated from 23.4 11+ and HT-29 cDNA libraries established that the observed mRNA species are derived by alternative RNA splicing.

DNA Sequence of TM2-CEA. Insert tcE22, isolated from the 23.4 11+ cDNA library, presented a restriction map indistinguishable from TM1-CEA clones when a 300-bp gap was positioned within the molecule. DNA sequence analysis showed that tcE22 codes for a second TM-CEA species (TM2-CEA), which is identical to the TM1-CEA apoprotein with the exception of a 100-amino acid gap that corresponds exactly to LD IIa (Figs. 2 and 3). We conclude that this cDNA accounts for the 3.7-kb mRNA observed in Fig. 1 and that it is derived by alternate splicing of the transcript that

| CAG | ÇCG | TGC | TCG | λλG
 | CGT | TCC | tgg | AGC
 | CCA | AGC | тст | сст | CCA
 | CAG | GTG | λλG | усу | GGG
 | CCA | GCA | GGA | GAC
 | ACC | 72 |
--	--	---	---
---	---	--	--
---	--	---	--
---	--	--	---
---	--	---	--
ATG Met -34	ecc Gly	CAC Ris	CTC Leu
 | GCC
Ala | CCA
Pro | CTT
Leu | CAC
Bis
 | λGλ
λrg | GTG
Val | CGT
Arg | GTA
Val | CCC
Pro
 | TGG
Trp | CAG
Gln | 666
61 y | CTT
Leu | CTG
Leu
 | CTC
Leu | ACA
Thr | GCC
Ala | TCA
Ser
 | CTT
Leu | 144
24 |
| CTA
Leu | ACC
Thr | TTC
Phe | TGG
Trp | AAC
Asn
 | CCG
Pro | CCC
Pro | ACC
Thr | ACT
Thr
 | GCC
Ala
-1 | CAG
Gln
+1 | CTC
Leu | ACT
Thr | ACT
Thr
 | GAA
Glu | TCC
Ser | ATG
Met | CCA
Pro | TTC
Phe
 | AAT
Asn | GTT
Val | GCA
Ala | GAG
Glu
 | GGG
Gly | 216
48 |
| AAG
Lys | GAG
Glu | GTT
Val | CTT
Leu | CTC
Leu
 | CTT
Leu | GTC
Val | CAC
His | እእፕ
እял
 | CTG
Leu | CCC
Pro | CAG
Gln | CAA
Gln | CTT
Leu
 | TTT
Ph e | GGC
Gly | TAC
Tyr | AGC
Ser | TGG
Trp
 | TAC
Tyr | ала
Lys | GGG
Gly | GAA
Glu
 | λGλ
λrg | 288
72 |
| GTG
Val | GАТ
Азр | GGC
Gly | እእር
እ#n | CGT
Arg
 | CAA
Gln | ATT
Ile | GTA
Val | GGA
Gly
 | TAT
Tyr | GCA
Ala | ATA
Ile | GGA
Gly | А СТ
Thr
 | C AA
Gln | CAA
Gln | GCT
Ala | ACC
Thr | CCA
Pro
 | GGG
Gly | CCC
Pro | GC A
Ala | AAC
As n
 | NGC
Ser | 3 60
96 |
| GCT
Gly | CGA
Arg | GAG
Glu | ACA
Thr | ATA
Ile
 | ТАС
Туг | CCC
Pro | λλΤ
λsn | GCA
Ala
 | TCC
Ser | CTG
Leu | CTG
Leu | ATC
110 | CAG
Gln
 | AAC
As n | GTC
Val | ACC
Thr | CAG
Gln | እእፕ
እደ በ
 | GAC
Авр | ACA
Thr | GGA
Gly | TTC
Phe
 | TAC
Tyr | 432
120 |
| ACC
Thr | CTA
Leu | CAA
Gln | GTC
Val | ATA
Ile
 | AAG
Lys | TCA
Ser | сат
Азр | CTT
Leu
 | GTG
Val | እእፕ
እያ በ | GAA
Glu | GAA
Glu | GCA
Ala
 | ACT
Thr | GGA
Gly | CAG
Gln | TTC
Phe | CAT
Bis
 | GTA
Val | TAC
Tyr | ccc
Pro | GAG
Glu
 | CTG
Leu | 504
144 |
| CCC
Pro | AAG
Lys | CCC
Pro | TCC
Ser | ATC
Ile
 | TCC
Ser | AGC
Ser | እእር
አደበ | λλC
λsn
 | TCC
Ser | AAC
Asn | CCT
Pro | GTG
Val | GAG
Glu
 | GλС
Азр | AAG
Lys | сат
Азр | GCT
Ala | GTG
Val
 | GCC
Ala | TTC
Phe | ACC
Thr | тст
Суз
 | GAA
Glu | 576
168 |
| CCT
Pro | GAG
Glu | ACT
Thr | CAG
Gln | GАС
Авр
 | ACA
Thr | ACC
Thr | TAC
Tyr | CTG
Leu
 | TGG
Trp | TGG
Trp | ATA
Ile | λλC
λ∎n | እእፕ
እደበ
 | CAG
Gln | AGC
Ser | CTC
Leu | CCG
Pro | GTC
Val
 | AGT
Ser | CCC
Pro | NGG
Arg | CTG
Leu
 | CAG
Gln | 648
192 |
| CTG
Leu | TCC
Ser | AAT
Asn | GGC
Gly | AAC
As n
 | AGG
Arg | ACC
Thr | CTC
Leu | ACT
Thr
 | CTA
Leu | CTC
Leu | AGT
Ser | GTC
Val | ACA
Thr
 | AGG
Arg | AAT
Asn | GAC
Азр | ACA
Thr | GGA
Gly
 | CCC
Pro | TAT
Tyr | GAG
Glu | тст
Суз
 | GAA
Glu | 720
216 |
| лта
11• | CAG
Gln | AAC
As n | CCA
Pro | GTG
Val
 | AGT
Ser | GCG
Ala | ААС
Авл | CGC
Arg
 | NGT
Ser | GAC
Азр | CCA
Pro | GTC
Val | ACC
Thr
 | TTG
Leu | AAT
Asn | GTC
Val | ACC
Thr | TAT
Tyr
 | GGC
Gly | CCG
Pro | GAC
And | ACC
Thr
 | ccc
Pro | 792
240 |
| ACC
Thr | ATT
Ile | TCC
Ser | CCT
Pro | TCA
Ser
 | GAC
Авр | ACC
Thr | TAT
Tyr | TAC
Tyr
 | CGT
Arg | CCA
Pro | GGG
Gly | GCA
Ala | ААС
Азп
 | CTC
Leu | AGC | CTC | TCC | TGC
 | TAT | GCA
Ala | GCC | TCT
Ser
 | AAC | 864 |
| CCA
Pro | CCT | GCA
Ala | CAG | TAC
 | TCC | TGG
Trd | CTT | ATC
 | AAT | GGA | ACA | TTC | CAG
 | CAA | AGC | ACA | CAA | GAG
 | CTC | TTT | ATC | CCT
 | AAC | 936 |
| ATC | ACT | GTG | AAT | AAT
 | AGT | GGA | TCC | TAT
 | ACC | TGC | CAC | GCC | AAT
 | AAC | TCA | GIC | ACT | GGC
 | Tec | AAC | ACG | ACC
 | <u>лэл</u>
лсл | 1008 |
| GIC | ANG | ACG | ATC |
ATA
 |
GTC | | A
GAG | CTA
 | AGT | CCA | GTA | GTA | GCA
 | ANG | CCC | CAA | ATC | 01 y
 | Y | AGC | ANG | MCC
 | Thr
MA | 1080 |
| Val | Lys | Thr | 11. | Il•
 | Val | Thr | Glu | Leu
 | Ser | Pro | Val | Val | Ala
 | Lys | Pro | Gln | 11e | Lys
 | Ala | Ser | Lys | Thr
 | Thr | 336 | | |
| GTC | 101 | | |
 | | | |
 | | | | _ |
 | | | _ | |
 | | | |
 | | |
| Val | Thr | GLY | Asp | AAG
Lys
 | GAC
Asp | TCT
Ser | CTC
Val | AAC
Asn
 | CTG
Leu | ACC
Thr | 7GC
Cys | TCC
Ser | ACA
Thr
 | እእፓ
እያስ | GAC
Asp | ACT | GGA
Gly | ATC
11e
 | TCC | ATC
Ile | CGT | TGG
Trp
 | TTC
Phe | 1152
360 |
| Val
TTC
Phe | Thr
AAA
Lys | Gly
AAC | GAT
Asp
CAG
Gln | AAG
Lys
AGT
Ser
 | GAC
Asp
CTC
Leu | TCT
Ser
CCG
Pro | GTG
Val
TCC
Ser | AAC
Asn
TCG
Ser
 | CTG
Leu
GAG
Glu | ACC
Thr
AGG
Arg | TGC
Cys
ATG
Met | TCC
Ser
AAG
Lys | ACA
Thr
CTG
Leu
 | AAT
Asn
TCC
Ser | GAC
Asp
CAG
Gln | ACT
Thr
GGC
Gly | GGA
Gly
AAC
Asn | ATC
Ile
ACC
Thr
 | TCC
Ser
ACC
Thr | ATC
Ile
CTC
Leu | CGT
Arg
Agc
Ser | TGG
Trp
ATA
Ile
 | TTC
Phe
AAC
Asn | 1152
360
1224
384 |
| Val
TTC
Phe
CCT
Pro | Thr
AAA
Lys
GTC
Val | AAC
AAG
Lys | GAT
Asp
CAG
Gln
AGG
Arg | AAG
Lys
AGT
Ser
GAG
Glu
 | GAC
Asp
CTC
Leu
GAT
Asp | TCT
Ser
CCG
Pro
GCT
Ala | CTG
Val
TCC
Ser
GGG
Gly | AAC
Asn
TCG
Ser
ACG
Thr
 | CTG
Leu
GAG
Glu
TAT
Tyr | ACC
Thr
AGG
Arg
TGG | TGC
Cys
ATG
Met | TCC
Ser
AAG
Lys
GAG
GIu | ACA
Thr
CTG
Leu
GTC
Val
 | AAT
Asn
TCC
Ser
TTC
Phe | GAC
Asp
CAG
Gln
AAC
Asp | ACT
Thr
GGC
Gly
CCA
Pro | GGA
Gly
AAC
Asn
ATC | ATC
Ile
ACC
Thr
AGT
 | TCC
Ser
ACC
Thr
AAG | ATC
Ile
CTC
Leu
AAC | CGT
Arg
AGC
Ser
CAA
Gln | TGG
Trp
ATA
Ile
AGC
Ser
 | TTC
Phe
AAC
Asn
GAC
Asn | 1152
360
1224
384
]296
408 |
| Val
TTC
Phe
CCT
Pro | Thr
AAA
Lys
GTC
Val | AAC
AAAC
AAAA
Lys
ATG | GAT
Asp
CAG
Gln
AGG
Arg
CTG | AAG
Lys
AGT
Sef
GAG
Glu
AAC
 | GAC
Asp
CTC
Leu
GAT
Asp
GTA | TCT
Ser
CCG
Pro
QCT
Ala
AAC | GTG
Val
TCC
Ser
GIY
SP | AAC
Asn
TCG
Ser
ACG
Thr

AAT
 | CTG
Leu
GAG
Glu
TAT
Tyr

GCT | ACC
Thr
ACG
Arg
TGG
Trp
 | TGC
Cys
ATG
Met
TGT
Cys
CCA | TCC
Ser
AAG
Lys
GAG
Glu
CAA | ACA
Thr
CTG
Leu
GTC
Val
 | AAT
Asn
TCC
Ser
TTC
Phe
AAT | GAC
Asp
CAG
Gln
AAC
Asn
GGC | ACT
Thr
GGC
Gly
CCA
Pro
CTC | GIY
AAC
Asn
ATC
Ile
TCA | ATC
Ile
ACC
Thr
ACT
Ser
CCT
 | TCC
Ser
ACC
Thr
AAG
Lys
GGG | ATC
Ile
CTC
Leu
AAC
Asn
GCC | CGT
Arg
AGC
Ser
CAA
Gln
ATT | TGG
Trp
ATA
Ile
AGC
Ser
GCT
 | TTC
Phe
AAC
Asn
GAC
Asp
GGC | 1152
360
1224
384
]296
408
1368 |
| Val
TTC
Phe
CCT
Pro
Pro | Thr
AAA
Lys
GTC
Val
ATC
Ile | AAC
AAC
AAA
Lys
ATG
Met | CAG
Gln
ASG
Arg
CTG
Leu | AAG
Lys
AGT
Ser
GAG
Glu
AAC
Asn
 | GAC
Asp
CTC
Leu
GAT
Asp
GTA
Val | TCT
Ser
Pro
Pro
Ala
Aan | GTG
Val
TCC
Ser
GGG
Gly
sp
TAT
Tyr | AAC
Asn
TCG
Ser
ACG
Thr

AAT
Asn
 | CTG
Leu
GAG
Glu
TAT
Tyr

GCT
Ala | ACC
Thr
ACG
Arg
Trp
CTA
Leu | TGC
Cys
ATG
Met
TGT
Cys
Cys
CCA
Pro | TCC
Ser
AAG
Lys
GAG
Glu
CAA
Gln | ACA
Thr
CTG
Leu
GTC
Val
GAA
Glu
 | AAT
Asn
TCC
Ser
TTC
Phe
AAT
Asn | GAC
Asp
Gln
Gln
Asn
GGC
Gly | ACT
Thr
GGC
Gly
CCA
Pro
CTC
Leu | GIY
GIY
AAC
Asn
ATC
Ile
TCA
Ser | ATC
11.
 | ACC
Thr
AACC
Thr
Lys
GGG
Gly | ATC
11e
CTC
Leu
AAC
Asn
GCC
Ala | CGT
Arg
AGC
Ser
CAA
Gln
ATT
Ile | TGG
Trp
ATA
Ile
AGC
Ser
GCT
Ala
 | TTC
Phe
AAC
Asn
GAC
Asp
GGC
GIY | 1152
360
1224
384
296
408
1368
432 |
| Val
TTC
Phe
CCT
Pro
ATT
Ile | Thr
AAAA
Lys
GTC
Val
ATC
Val
Ile
GTG
Val | CGA
Gly
AAC
AAG
Lys
ATG
Met
XTT
Ile | GAT
Asp
CAG
Gln
Arg
CTG
Leu
GQA
Gly | AAG
Lys
Sof
GAG
Glu
AAC
Asn
GTA
Val
 | GAC
Asp
CTC
Leu
GAT
Asp
GTA
Val
GTC
Val | TCT
Ser
CCG
Pro
OCT
Ala
Agn
QCC
Ala | CTC
Val
TCC
Ser
Gly
sp
Tyr
Tyr
CTG
Leu | AAC
Asn
TCG
Ser
ACG
Thr
AAT
Asn
GTT
Val
 | CTG
Leu
GAG
Glu
TAT
Tyr

GCT
Ala | ACC
Thr
ACC
Arg
TCG
Trp
CTA
Lou | ATGC
Cys
ATGT
TGT
Cys
CCA
Pro
ATA | TCC
Ser
AAG
Lys
GAG
Glu
CAA
Gln
GCA
Ala | ACA
Thr
CTG
Leu
GAA
Glu
GAA
Glu
GTA
Val
 | AAT
Asn
TCCC
Ser
TTCC
Phe
AAT
Asn
GCCC
Ala | GAC
Asp
CAG
Gln
AAC
CAG
Gln
CAG
Gln
CTG
CTG
Lou | ACT
Thr
GGC
Qly
CCA
Pro
CTC
Leu
CTC
Leu
Ala | GIY
AAC
Asn
ATC
TCA
Sef
Cys | ATC
11-
ACC
Thr
ACT
Ser
CCT
Pro
TTT
Phe
 | TCC
Ser
ACC
Thr
AAG
CI
Y
CTG
Leu | ATC
Ile
CTC
Leu
AAC
AAC
Ala
CAT
His | CGT
Arg
Agc
Ser
CAA
Gln
ATT
Ile | ATA
ATA
ATA
ATA
ACC
Ser
GCT
Ala
 | TTC
Phe
AAC
Asn
GAC
GAC
GIY
Lys | 1152
360
1224
384
296
408
1368
432
1440
456 |
| Val
TTC
Phe
Pro
Pro
ATT
Ile | Thr
Thr
Lys
Val
ATC
Val
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC | Gly
Gly
AAC
AAG
Lys
ATG
Het
XTT
Ile
AGG
Arg | CAT
Asp
CAG
Gln
AGG
Arg
CTG
Leu
CCTG
CTG
Leu
Ale | AAG
Lys
AGT
GAG
Glu
AAC
Asn
GTA
Val
 | CAC
Asp
CTC
Leu
GAT
Asp
CTC
Val | CCCG
Pro
Ala
AAC
Ala
Gln | CTC
Val
TCC
Ser
Ser
Tyr
Tyr
CTC
Lou
Arg | AAC
Asn
TCG
Ser
Thr
AAT
Asn
GTT
Val
 | CTG
CAG
CAG
Clu
TAT
Tyr
CCT
Ala
CCT
Ala | ACC
Thr
ACC
Arg
Trg
Trp
CTA
Lou
CTG
Lou
CTG
Lou
CTG
Lou
CTG
Trp | ATG
Cys
ATG
Met
Cys
Cys
Cys
Cys
ATA
Ile
Glu | TCC
Ser
AAG
Lys
GAG
Glu
CAA
Gln
GCA
Ala
Bis | ACA
Thr
Thr
CTG
Leu
GTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
 | AAT
Asn
TCC
Ser
TTC
Ser
AAT
Asn
GCC
Ala
Pro | CAC
Asp
CAC
CIn
CAC
CIn
CAC
CIn
CIN
CIN
CIN
CIN
CIN
CIN
CIN
CIN
CIN
CIN | ACT
Thr
Thr
QGC
Qly
Pro
CCA
Pro
CTC
Leu
QCA
Ala | GGA
Gly
AAC
Asn
Atrc
TCA
Sef
TGT
TGT
Cys
Ser | ATC
III-
ACC
Thr
ACT
Thr
ACT
Pro
TTT
Phe
ASC
ASC
 | ACC
Thr
AAC
Lys
CrG
Leu
His | ATC
IIe
CTC
Leu
AAC
AAC
AAC
ACAT
Hie
Thr | CGT
Arg
Arg
Ser
CAA
Gln
ATT
TIC
Phe
erg
Gln | ATA
ATA
ATA
ACC
ALC
Ser
ALC
CCT
ALC
CCT
ALC
CCC
CCT
ALC
CCC
ALC
CCC
ALC
CCC
ALC
CCC
CCC
ALC
CCC
CC
 | TTC
Phe
AAC
Asn
GAC
Cly
CaC
Bis | 1152
360
1224
384
]296
408
1368
432
1440
456
1512
480 |
| Val
TTC
Phe
Dro
Pro
ATT
Thr
CCCC
Pro | Thr
AAA
Lys
GTC
Val
ATC
Ile
GTC
CTC
GI
Val
In*
AAT | AAC
Gly
AAC
ASD
Lys
ATG
ATG
ATG
ATG
AAC
AAC
AAC
AAC
AC
AC
AAC
AC
AC
AC
AC
A | CAC
CAG
CIn
AGG
Arg
CTG
CTG
CLeu
CCA
CCA | AAG
Lys
AGT
Ser
GAG
Glu
Asn
GTA
Val
ASS
Ser
CCT
 | GAC
Asp
CTC
Lau
GAT
Asp
Val
GTG
Val
GTG
Val | CCG
Ser
CCG
Pro
Ala
Asn
Ala
CCC
Ala | CTC
Val
TCC
Ser
Gly
PAT
Tyr
CTG
Leu
Arg | AAC
Asn
TCG
Ser
Thr
AAT
Asn
GTT
Val
 | CTC
Leu
GAG
Glu
TAT
Tyr
CCT
Ala
GCT
Ala
CTC
Leu
GAA | ACC
Thr
ACC
Arg
TGG
Trp
CTA
Leu
CTG
Leu
CTG
Leu
CTG
CTG
Thr | ATC
Cys
ATC
Met
TGT
Cys
CCA
Pro
ATA
Ile
Glu
ACT | TCC
Ser
AAG
Lys
GAG
Glu
CAA
Gln
GCA
Ala
Eis
TAT | ACA
Thr
Thr
CTG
Leu
Val
GAA
Glu
GAA
Glu
Val
Lys
TCT
 | AAT
Asn
TCC
Ser
TTC
Phe
AAT
Asn
GCC
Als
Pro
ACC | CAC
Asp
CAC
Gln
AAC
Asn
CTC
Leu
XCA
Sor
CTC | ACT
Thr
Thr
GGC
Qly
CCA
Pro
CTC
CTC
CTC
Lau
GCA
Ala
Val | GGA
AAAC
Aan
AAAC
TIL
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TCA | ATC
11.0
ACC
Thr
ACC
Thr
ACC
Thr
CCT
PFO
CCT
PFO
ASC
CCT
CCT
CCT
CCT
CCT
CCT
CCT
C
 | ACC
Thr
ACC
Thr
Lys
Crc
Crc
Leu
Str
Scc
Scr
Scc
Scc
Scc | ATC
Ile
CTC
Leu
AAC
AAC
AAS
AS
CAT
His
CAT
Thr
CAG | CGT
Arg
Agc
Ser
CAA
Gln
TTC
Phe
erg
Gln
CAA | ATA
ATA
ATA
ATA
ACC
Ser
GCT
Ala
GCC
Gly
Lyp
GAC
Asp
CCC
 | TIC
Phe
AAC
Asn
GAC
Asn
GAC
Asp
GCC
CLy
Sis
CAC
Bis
ACA | 1152
360
1224
384
296
408
1368
432
1440
456
1512
480 |
| Val
TTC
Phe
Pro
Pro
ATT
Ile
ATT
Thr
•uG
TCC
Ser | AAAA
Lys
GTC
Val
ATC
GTC
GIY
In*
AAT
AAT
Asn | Gly
AAC
AAC
AAG
Lys
ATG
ATG
ATG
ATG
ATG
ATG
ATG
ATG | CAC
Asp
CAG
Gln
AGG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
C | AAG
Lys
AGT
Sar
GAG
Glu
AAC
Clu
AAS
CTA
Val
Ser
CCT
Pro
 | GAC
Asp
CTC
Leu
GAT
Asp
GTA
Val
GTG
Val
Asp
Asp | CCCG
Ser
CCCG
Pro
Ala
Asn
Acc
Ala
CCCT
Ala
Asn
CCCT
Ala | CTC
Val
TCC
Ser
Ser
Tyr
Tyr
Tyr
CTG
Leu
Arg
ATG
Het | AACC
ACC
Ser
ACC
Thr
AAT
ASD
ASD
AAT
ASD
 | CTC
Letu
GAG
Glu
TAT
Tyr

GCT
Ala
GCT
Leu
GAA
Glu | ACC
Thr
ACG
Trrp
Trp
CTA
Leu
CTG
Leu
CTG
CTG
Thr
GTT
Thr | ATG
Cys
ATG
Met
Cys
Cys
Cys
Cys
CCA
Pro
ATA
Ile
Glu
ACT
Thr | TCC
Ser
AAG
Lys
GAC
Glu
CAA
Glu
CAA
Glu
CAA
Ala
His
TAT
Tyr | ACA
Thr
Thr
CTG
Leu
GAA
Clu
GAA
Clu
GAA
Clu
CTA
Lys
Ser
 | AAT
Asn
TCC
Ser
TTC
Phe
AAT
Asn
GCC
ALS
Pro
ACC
Thr | GAC
Asp
CAG
Gln
AAC
CIC
Gln
CTG
CTG
Leu
CTG
Leu
CTG
Leu | ACT
Thr
CCA
Cly
CCA
Pro
CTC
Lev
Val | AAC
Gly
AAC
Asn
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TCA | ATC
III
ACC
Thr
ACC
Thr
CCT
Pro
CCT
Pro
CCT
Pro
CCT
Pro
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CC
 | ACC
Thr
AAC
Lys
CTG
CTG
Leu
CTG
Elis
GCC
Ala | ATC
11e
CTC
Leu
AAC
Asn
CAT
His
CAT
Thr
CAG
Gln | CGT
Arg
Arg
CAA
Ser
CAA
Gln
ATT
TIC
CAG
Gln
CAA
Gln
CAA
Gln | The Trp
ATA AACC
ACC
ACC
ACC
ACC
ACC
ACC
AC
 | TIC
Phe
AAC
Asn
GAC
GLy
CAC
CLys
roL
Lys
roL
Bis
ACA
Thr | 1152
360
1224
384
296
409
1368
432
1440
456
1512
480
1564
504 |
| Val
TTC
Phe
DCT
Pro
Pro
ATT
Thr
CCC
CCC
Pro
ATT
Chr
Gln | AAAA
Thr
Thr
Lys
GTC
Val
ATC
Cly
In*
AAT
Asn
CCA
Fro | AAC
AAC
AAAG
AAAG
AAAG
AAAG
AAAG
AAAG
A | CAG
GIN
ASD
CAG
GIN
ACG
CIN
CCTG
CLeu
CCTG
CLeu
CCTG
CLeu
CCTA
Ser
TCA
Ser | AAG
Lys
AGT
Ser
GAG
Glu
AAC
Clu
AAC
Ser
CCT
Pro
GCC
Ala
 | GAC
Asp
CTC
Leu
GAT
Leu
GAT
Asp
GTA
Val
GTG
CTG
Val
Asp
Asp
AAC
Ser | TCT
Ser
Ser
Pro
QCT
Ala
Asn
QCT
Ala
Asn
QCT
Ala
CAC
Gln
AAC
CAC
Fro
CCCG
Pro | CTC
Val
TCC
Ser
Gly
sp
Tyr
CTC
CTC
Leu
ATG
Het
TCC
Ser | AAC
Asn
TCG
Ser
Thr
AAT
Asn
GTT
Yal
SAT
Asn
AAT
Asn
CTA
Leu
 | CTG
Leu
GAG
Glu
TAT
Tyr
Tyr
CGCT
Ala
GCT
CAL
au
GAA
Glu
CAA
Glu
ACA
Thr | ACC
Thr
ACC
Thr
Arg
Trp
Trp
CTA
Leu
CTG
Leu
CTG
CTA
Leu
CTG
CTA
Leu
CTG
CTA
Leu
CTG
Ala | ATG
Cys
ATG
Met
TGT
Cys
Cys
CCA
Pro
ATA
Cys
CCA
Cys
CCA
Cys
CCA
Cys
Cys
CCA
Cys
Cys
Cys
Cys
Cys
Cys
Met | TCC
Ser
AAG
GAG
Glu
CAA
Glu
CAA
Glu
CAA
CAC
CAA
CAC
CAA
CAA
CAA
CAA
CAA
CA | ACA
Thr
Thr
CTG
Leu
GAA
Glu
GAA
Glu
GAA
CAA
Lys
TCT
Ser
Ser
ATA
 | AAT
Asn
TCC
Ser
TTC
Phe
AAT
Asn
GCC
GCC
ACC
Thr
ATT
Ile | CAC
Asp
CAC
Gln
Asn
CAC
Gln
CTC
CTC
CTC
Leu
TAT
Tyr | ACT
Thr
Thr
GGC
Gly
CCA
Pro
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CT | GGA
Gly
AAC
AAn
ATC
TCA
TCA
TCA
TCA
TCA
Ser
TTT
Phe
GAA | ATC
1110
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
COT
TTT
Thr
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CC
 | TCC
Ser
ACC
Thr
Lys
GGG
GGL
CIG
U
CIG
U
CIG
U
CIG
U
SAC
CIG
U
SAC
CIG
U
SAC
CIG
U
SAC
CIG
U
SAC
CIG
U
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
CIG
SAC
SAC
SAC
SAC
SAC
SAC
SAC
SAC
SAC
SAC | ATC
Ile
CTC
Leu
AAC
ASD
CAT
SCT
Thr
CAG
Gln
AAG
Lys | CGT
Arg
AGC
Ser
CAA
Gln
ATT
TTC
CAG
Gln
CAA
Gln
CAA
Gln
CAA
Gln
CAA | TIGG
Trp
ATA
ATA
AGC
Ser
GCT
GCT
GCT
GCC
GCC
CCC
Pro
TAA
 | TIC
Phe
AAC
Asn
GAC
Asp
GCC
CL
SC
Els
AAG
Bis
ACA
Thr
TGA | 1152
360
1224
384
]296
408
1368
432
1440
456
1512
480
1554
504 | | | | | | | | | | | | | | | | | | | | | | | |
| Val
TTC
Phe
CCT
Phe
CCT
Pro
Pro
Pro
ATT
Thr
•uG
TCC
Thr
•uG
CCC
Thr
•uG
CCC
Thr
•ACC
Thr | AAAA
Thr
AAAA
Lys
GTC
Val
ATC
GC
Gly
In*
AAT
AAAA
AAAA
AAAA
AAAA
AAAA
CCA
Pro
CTG | Gly
AMC
ASA
ASA
ASA
ASA
ATG
ATG
ATG
ACT
Thr
TCCC
GGT | CAG
CAG
CAG
CAG
CIN
ACC
CAG
CIN
ACC
CAG
CIN
CCAG
CIN
CCAG
CIN
CCAG
CIN
CCAG
CIN
CCAG
CIN
CAG
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CAG
CIN
CIN
CAG
CIN
CIN
CAG
CIN
CIN
CAG
CIN
CIN
CIN
CIN
CIN
CIN
CIN
CIN
CIN
CIN | AAG
Lys
AGT
Ser
GAG
Glu
AAC
Clu
AAA
Ser
CCT
Pro
GCC
Ala
AAAA
CGG
GGC | GAC
Asp
CTC
Leu
GAT
Asp
GTA
Val
GTG
Val
GTG
Sor
Sor
CTG
GTG
GTG
GTG
GTG
GTG
GTG
GTG
GTG
GTG | CCCG
Pro
Pro
Ala
Asn
GCC
Ala
CCC
Ala
CCC
Ala
CCC
Ala
CCC
CCC
CCCG
Pro
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CC | CTC
Val
TCC
Ser
Ser
Tyr
Tyr
CTG
Leu
Arg
Arg
Arg
Arg
Arg
Arg
Arg | AACC
Asn
TCG
Ser
Thr
AAT
Asn
GTT
Val
Asn
CTA
Asn
CTA
Leu
TGA | CTG
Ley
GAG
Glu
TAT
Tyr
CCT
Ala
GCT
Ala
GCT
GLU
GAA
Glu
GAA
Glu
CAA
Glu
CAA
GLU
CTA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA | ACC
Thr
ACC
Thr
ACC
Arg
Trp
CTA
Lou
CTG
Lou
CTG
Lou
CTG
Lou
CTG
Lou
CTG
Ala
ATT
ACT | ATC
CYs
ATC
Met
TCTT
CYs
CCA
Pro
ATA
Fro
Glu
ATA
Thr
Thr
TCA | TCC
Ser
Lys
GAG
Glu
CAA
Glu
CAA
Glu
CAA
Ala
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA | ACA
Thr
Thr
CTG
Leu
Val
GAA
Glu
GAA
Glu
Val
Lys
TCT
Ser
ATA
Ile
CTC
TAC | AAT
Asn
TCC
Ser
TTCC
Ser
TTCC
AAT
Asn
GCC
ALL
ASN
ACC
Thr
Sign
ATT
TCA
ATT
TCA | CAC
Gln
CAC
Gln
Asn
CAC
CAC
Gln
CAC
CLOU
CTC
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CC | ACT
Thr
GGC
Gly
CCA
Pro
CTC
Leu
CCC
Ala
Ala
Asn
TCA
Ser
TCA | AAC
Gly
AAC
Asn
ATC
TCA
TCA
TCA
TCA
TCA
TCA
TCA
Glu
TTT
TTT
Phe
GAA
CTA | ATC
11-
ACC
Thr
ACT
Thr
ACT
TTT
Pho
CCT
PFO
CCT
PFO
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CC | ACC
Thr
AAG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CT | ATC
III
CTC
Leu
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | CGT
Arg
AGC
Ser
CAA
Gln
ATT
TIC
CAG
Gln
CAA
Gln
CAA
Gln
CAA
Gln
TCC
TCC | The Transformer of transformer of the transformer o | TIC
Phe
AAC
Asn
GAC
Asp
GGC
Clys
rol
Lys
rol
CAC
Bis
ACA
Thr
TGA | 1152
360
1224
384
]296
408
1368
432
1440
456
1512
480
1564
504 |
| Val
TTC
Phe
CCT
Phe
Pro
Pro
ATT
Thr
CCC
Pro
ATT
Thr
CCC
CTC
Thr
CCC
CTC
Thr
CCCC
CTC
Thr
CCCC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
C | Thr
Thr
AAAA
Lys
GTC
Val
ATC
Val
Ile
GTC
Gly
In*
AAT
Asn
CCA
TAG
GGT
TAG
GGT
CAT | Gly
AAC
ASD
ASD
ASD
ASD
ASD
ASD
ASD
ASD | CAG
GIN
AGG
GIN
AGG
GIN
CTG
CTG
Leu
GGA
AIg
CTG
GIN
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG |
AAG
Lys
AGT
Ser
GAG
Glu
ABA
GTA
ASA
GTA
ASA
CCT
Fro
GCC
Ala
ATCA
GGG
CCT
TIG | GAC
Asp
CTC
Leu
GAT
Leu
GAT
Asp
GTA
Val
GTA
Val
GTA
Ser
CTG
GTG
GTG
GTG
GCA | CCCG
CCCG
Pro
CCCG
Pro
CCC
Ala
Asn
CAC
CCA
CAC
CCA
CCA
CCA
CCA
CCA
CCA
CC | CTC
Val
TCC
Ser
Gly
sD
TTY
TY
TTY
CTG
Leu
ATG
Het
TCC
Ser
TCC
Ser
TCC
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG |
AAC
ACG
Ser
Thr
AAT
AAT
AAT
AAT
AAT
CTA
ASD
CTA
CTA
CTA | CTG
Leu
GAG
Glu
TAT
Tyr

GCT
Ala
GCT
Ala
GCT
Leu
GAA
Glu
ACA
AAT
TTG
TACA | ACC
Thr
ACC
Arg
Trp
Trp
CTA
Leu
CTG
Leu
CTG
Leu
CTG
ACA
Thr
GTT
Trp
CTA
Leu
CTG
ACA
ACA
ACA
ACA
ACA
ACA
ACA
ACA
CTA
ACA
CTA
ACA
CTA
CT | ATC
CYS
ATC
Met
TOT
CYS
CYS
CCA
Pro
CYS
CCA
Pro
Glu
ATA
ACA
Thr
TCA
Thr
TCA
TACA | TCC
Ser
Lys
GAG
Glu
CAA
Glu
CAA
Glu
CAA
Ala
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA | ACA
Thr
Thr
CTG
Leu
Val
GAA
Glu
GAA
Glu
GAA
Lys
TCT
Ser
TCT
Ser
TAC
TAC
TAC
TAC
GAC
 | AAT
Asn
TCC
Ser
TTC
Phe
AAT
Asn
GCC
Ala
ACC
Thr
ATT
Ile
ATT
TCA | CAC
Gln
CAC
Gln
Asn
CAC
Cln
CTG
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC | ACT
Thr
Thr
CCA
Qly
Pro
CTC
Leu
CCA
Ala
Ala
Ala
Ala
Ala
CCA
Ala
CTC
Ala
CTC
CCA
TCA
Asn
TCA | AAC
Gly
AAC
Asn
ATC
TCA
TCA
TCA
TCA
Sef
TTT
TCA
Sef
TTTT
Phe
GAA
CTG
CCC | ACC
Thr
ACC
Thr
ACC
Thr
Ser
CCT
Pro
CCT
Pro
CCT
Pro
CCT
Pro
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CC
 | ACC
Sor
ACC
Thr
Lys
CTG
CTG
Leu
CTG
ALS
CTG
ALS
CTG
ALS
CTG
ALS
CTG
CCC
ALS
CTG
CCC
CTG
CCC
CCC
CCC
CCC
CCC | ATC
III
CTC
Leu
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | CGT
Arg
Acc
Sor
Gln
ATT
CAA
Gln
CAA
Gln
CAA
Gln
CAA
Gln
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA | TAGE
Trp
ATA
ATA
ACC
Ser
Ala
CCT
Ala
CCC
Pro
TAA
ASP
CCC
TTAA
 | TIC
Phe
AAC
Asn
GAC
Asp
GCC
CCC
CCC
CCC
CCC | 1152
360
1224
384
2296
409
1368
432
1440
455
1512
480
1564
504 |
| Val
Trc
Phe
DCT
Pro
Pro
ATT
Thr
Ser
CAN
Gln
AACC
ToT
TAA | AAAA
Thr
Thr
AAAA
Lys
GTC
Vall
GTC
Cly
In*
AAT
CCA
CCA
TTCC
GGT
CAT
TTCC | Gly
Gly
AAC
ASD
ASD
ASD
ASD
ASD
ASD
ASD
ASD | CAC
Gln
Arg
CLau
CAC
Gln
Arg
CLau
CCLau
CCLau
CCLau
CCLau
CCLau
CCLau
CCLau
CCLau
CCLau
CCLau
CCLau
CCLau
CCLau
CCLau
CLau |
AAG
Lys
AGT
Sor
GAG
Glu
AGC
Glu
AAC
ASA
Sor
Fro
GCC
CCT
Pro
GCC
CCT
Pro
GCC
CTTG
TTTG | GAC
Asp
CTC
Leu
GAT
Leu
GAT
Asp
GTA
Val
GTA
Sor
CTG
GTG
GTG
GTG
GTG
GTG
GTG
GTA
AAA | CCG
GCT
Ala
Asn
GCC
Ala
CCG
CCG
Pro
GCC
Ala
Asn
CAC
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG | CTC
Val
TCC
Ser
Gly
JAT
Tyr
CTG
Lou
Arg
ATG
AAT
TCC
Ser
TCC
Ser
TCC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CT |
AACC
Ser
Thr
AACG
Ser
Thr
AAT
ASD
AAT
ASD
AAT
ASD
AAT
TGA
ASD
AAT
TGA
ASD
AAT | CTG
Leu
GAG
Glu
TAT
Tyr

GCT
Ala
GCT
Ala
GCT
Ala
GCT
TAT
TTG
ACA
Thr
TGA
AAT
TTG | ACC
ACC
Thr
ACC
ACC
ACC
ACC
ACC
ACC
ACC
AC | ATG
CYs
ATG
Met
CYs
CYs
CYs
CYs
CCA
Pro
ATA
CCA
CYs
CYs
CCA
ATA
CCA
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs | TCC
Ser
AAG
Lys
GAG
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA | ACA
Thr
Thr
CTG
CTG
Val
GAA
Glu
GAA
Clu
GAA
Clu
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
 | AAT
TCC
Ser
TTCC
Ser
TTCC
Cor
AAT
AAT
AAT
ACC
Thr
ACC
Thr
ACC
Thr
ACC
TTCC
ACC
TTCC
ACC
TTCC
ACC
TTCC
ACC
A | CAG
Gln
Asp
Gln
AAC
Gln
Asn
CCG
Gly
CTG
Lou
Lou
TAT
Tyr
CCC
CCC
GGG
AAC
CCCC
CCC
CCCC
CCCC
CCC | ACT
Thr
Thr
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
Cly
CCCA
CCCA | GIA
GIY
AAC
Asn
ATC
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TTT
TPhe
GAA
CTA
CTA
CTA
CTA | ACC
Thr
ACC
Thr
ACT
Thr
Ser
CCT
TTT
Phe
CCT
TTT
Phe
CCT
ACC
CCT
TTT
TPhe
CCT
ACC
CCT
TTT
TCC
ACC
CCT
ACC
CCT
CCT
 | TCC
Ser
ACC
Thr
Lys
CTG
CCC
CTG
CCC
CCC
CCC
CCC
CCC | ATC
III
CTC
Lou
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | CGT
Arg
Arg
Gln
ATT
CAA
Gln
ATT
CAA
Gln
CAA
Gln
CCAG
Gln
TCC
CCAG
CCAG
CCAG
CCAG
CCAG
CCAG
CCAG | TGG
Trp
ATA
ACCC
Ser
GCT
CCC
TCC
TAA
ACCC
CCC
TTAA
ACCC
CCC
TTAA
 | TIC
Phe
AAC
Asn
GAC
GC
GC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC | 1152
360
1224
384
]296
408
1368
432
1440
456
1512
480
1554
526 | | | | | | | | | | | | | | | | | | | | | | | |
| Val
The
Phe
ECT
Pro
Pro
ATT
Thr
CCC
Thr
CAA
CCC
Thr
CAA
CCC
Thr
CCA
CCC
Thr
CCA
CCC
Thr
CCC
CCC
Thr
CCCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC | AAAA
Thr
Thr
AAAA
Lys
GTC
Val
ATC
Val
Ile
GTG
Gly
In*
AAT
Asn
CCA
GTC
CAT
TGG
GTTC
CAT
TGG
GAA | Gly
Gly
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | CAC CIG
CAC CIG
CAC CIG
Arg
CCAC CIG
Arg
CCAC CIG
CIG
CAC CIG
CIG
CAC CIG
CIG
CAC CIG
CIG
CAC CIG
CIG
CAC CIG
CIG
CAC CIG
CIG
CAC CIG
CAC CIG | AAG
Lys
AGT
Ser
GAG
Glu
AAC
Clu
AAC
Clu
AAC
Clu
AAC
Clu
AAC
Clu
CCT
Ser
CCT
Fro
GCC
ALa
AAAA
TCA
GCC
TTT
TCA
GCC
CCC
CCC
CCC
CCC
CCCC
C | GAC
Asp
CTC
CTC
Levu
GAT
Asp
CTA
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC | CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCCC | CTC
Val
TCC
Ser
Gly
DAT
Tyr
CTG
Leu
Arg
ATG
Ket
TCC
Ser
TCC
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG | AAC
TCG
Ser
Thr
ACG
Ser
Thr
AAT
AAT
AAT
AAT
CTA
Leu
AGA
GTC
TCTA
CTA | CTG
Leu
TAT
Tyr
Tyr
CGCT
Ala
GCT
Ala
GCT
Leu
GAA
GIU
ACA
Thr
TGA
AAAT
TTGA
TTA | ACC
ACC
Thr
ACC
Arg
Trp
CTA
Leu
CTA
Leu
CTA
Leu
CTA
ACA
ATT
ACA
ATT
ACA
ATT
ACA
ATT
TAT | ATG
CYs
ATG
Met
CYs
ATG
TOTT
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs | TCC
Ser
AAG
GL
GL
GL
GL
GL
GL
GL
GL
GL
GL
GL
GL
GL | ACA
Thr
Thr
CTG
CTG
CTG
CTG
CTG
CTG
CTC
CTC
CTAC
CCA
AAA | AAT
TCC
Ser
TTCC
Ser
TTCC
Ser
AAT
Asn
GCC
ALA
ASN
ACC
TTR
ATT
IIIe
ATT
TCA
TTR
CCT
TTR
CCT
TTR
CCT
TTCC | GAC
Asp
CAG
Gln
Asn
Gly
CTG
Gly
CTG
Gly
CTG
Leu
Ser
CTG
Leu
TAT
TYr
CCCC
CCG
AAC
CCCC
AAC | ACT
Thr
Thr
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cly
CCA
Cla
Cly
CCA
Cly
CCA
Cla
Cly
Cly
CCA
Cla
Cly
Cly
CCA
Cla
Cly
Cly
Cly
Cly
Cly
Cly
Cly
Cly
Cly
Cly | GIA
GIY
AAC
Ann
ATC
TCA
ANT
TCA
TCA
TCA
Sor
TTTT
TPho
GAA
CTNG
GIU
TCA
AC
CTNG
GAA | ACC
110
ACC
Thr
ACC
Thr
ACC
Thr
CCT
Pho
CCT
Pho
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CC | ACC
Ser
ACC
The
AAG
CTG
Lys
CTG
CTG
Lys
CTG
AAG
CTY
SH15
CCC
AAA
CCC
CTG
CCC
ACC
CTG
CCC
CTG
CCC
CCC
CCC
CCC
C | ATC
IIe
CTC
Leu
AAC
CAT
His
CAT
The
CAG
GIN
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
CAT
AAG
AAG
AAG
AAG
AAG
AAG
AAG
A | CGTT
Arg
Arg
CAAA
Gln
ATT
CAAA
Gln
CAG
Gln
CAG
Gln
CAG
Gln
CCAG
Gln
CCAG
CAG
CAG
CAG
CAG
CAG
CAG
CAG | TGG
Trp
ATA
ACC
Trp
ACC
Ser
GCT
ACC
Cly
Pro
TAA
CCC
TTA
ACC
TTC
ACC
CCA
ACC | TIC
Phe
AAC
Asp
GGAC
Cly
Asp
GGC
Cly
AAG
Cly
ToL
CAC
Bis
ACA
Thr
TCA
ATAT | 1152
360
1224
384
]296
408
1368
432
1440
456
1512
480
1564
504 |
| Val
Trc
Phe
ECT
Pro
CCC
Pro
ATT
Lle
CCC
Ser
CAN
Gln
AAC
CCC
CCC
CCC
CCC
CCC
CCC
CC | AAAA
ATC
GTC
Val
ATC
CVal
ATC
CVal
CTG
GCV
AAT
AAT
AAT
CCA
Fro
CTG
GGT
TTC
GAAA
AAAA | Gly
Gly
AACC
AAC
AAC
AAC
ATG
ATG
ATG
ATG
ATG
ATG
ATG
ATG | GAT
Asp
CAG
GIN
Arg
CTG
Leu
GIN
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG |
AGC
Lys
AGT
Sei
GAG
Glu
AAC
Glu
AAC
Glu
GC
AAC
Ser
Fro
GCC
CCT
Fro
GCC
CCT
TTG
GCC
CCC
GCC
GCC
GCC
GCC
GCC | GAC
Asp
CTC
Leu
GAT
Asp
GTA
Val
GTA
Asp
AAC
Asp
AAC
Asp
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CT | CCCG
CCCG
Pro
CCCG
Pro
CCCG
CCC
Ala
Asn
CCCG
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC
C | CTC
Val
TCC
Ser
Gly
DAT
Tyr
CTG
Leu
Arg
Arg
Arg
Arg
Arg
Arg
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG |
ACC
Ser
TCG
Ser
ACC
Ser
Thr
AAT
ASA
ASA
ASA
CTA
ASA
CTA
ASA
CTA
ASA
CTA
ASA
CTA
ASA
CTA
ASA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CT | CTG
Leu
GAG
Glu
TAT
Tyr
Ala
GCT
Ala
GCT
Ala
GCT
Ala
GCT
TTAT
TTA
TTA
TTA
TTA
TTA
TTA
TTA
TTA | ACC
Thr
ACC
Thr
ACC
Arg
Trp
Trp
CTA
Leu
CTA
Leu
CTA
Leu
CTA
ACA
Thr
ACA
CTA
Thr
ACA
CTA
ACG
CTA
ACG
CTA
ACG
CTA
ACG
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA | ATGC
CYs
ATGC
CYs
ATGC
CYs
ATG
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs | TCC
Ser
AACC
GAC
GIU
CAA
GIU
CAA
GIU
CAA
GIU
CAA
GIU
CAA
GIU
CAA
GIU
CAA
GIU
CAA
GIU
CAA
CAA
GIU
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA |
ACA
Thr
Thr
CTG
Leu
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
Cau
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
GTC
CAU
CAU
CAU
CAU
CAU
CAU
CAU
CAU
CAU
CA | AAT
TCC
Ser
TTCC
Ser
TTCC
Ser
Phe
AAT
Asn
GCC
Als
ACC
Thr
TCA
ATT
TCA
ATT
TCA
TCA
CTT
TCA
CTT
CTT | CAC
Gla
Gla
CAC
Gla
Asp
CAC
Gla
Asp
CAC
Gla
CTG
Leu
CTG
Leu
TAT
Tyr
CCCC
GGG
AAC
CAC
CCCC
CCCC
CCCCCCCCCCC | ACT
Thr
Thr
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
CCA
Club
Club
Club
Club
Club
Club
Club
Club | GIA
GIY
AAC
Ann
ATC
TGT
TGT
TGT
TGT
TGT
TGT
TGT
TGT
TGT
T
 | ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
TTT
Pho
CCT
Pro
CCT
Pro
CCT
Pro
CCT
CCT
TTT
Pho
CCT
ACC
CCT
CCT
CCT
CCT
CCT
CCT | ACC
Ser
ACC
Thr
AACC
Thr
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG | ATC
IIe
CTC
Leu
AAC
CAT
LS
CAT
HIE
CAT
CAT
CAT
CAT
CAT
CAT
CAT
CAT | CGAT
Arg
Agg
Ser
CAA
ATT
TTC
CAA
ATT
TTC
CAA
Gln
CAA
Gln
CAA
Gln
CAA
Gln
TTC
CAA
CAA
CAA
AAC
 | TAG
Trp
ATA
ACC
Trp
ACC
ACC
ACC
ACC
ACC
ACC
ACC
ACC
ACC
AC | TIC
Phe
AAC
Asn
GAC
Asn
GAC
Lys
roll
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
C | 1152
360
1224
384
1284
408
1368
432
1440
456
1512
480
1554
504 |
| Val
TTC
Phe
CCT
Pro
CCC
Pro
ATT
Ile
CCC
Thr
CCC
Ser
CCC
CCC
CCC
CCC
CCC
CCC
CCC
C | AAAA
Thr
AAAA
Lys
GTC
Val
ATC
CVal
ATC
GCC
Gly
In*
AAT
Asn
CCA
Pro
CTG
GGT
TAG
GGT
TAG
GAA
AAAA
AAT
CCA
ATC
CTG
CTG
CTG
CTG
CTG
CTG
CTG
C | Gly
Gly
AAC
AAG
AAG
AAG
AAG
AAG
AAG
AAG | GAT Asp
CAG GIR
AGO
GIR
AGO
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG |
AAG
Lys
AGT
Ser
GAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
CCT
CCT
Ser
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CC | GAC
Asp
CTC
Leu
GAT
Asp
GTA
Asp
CTC
Val
GTG
CTC
Asp
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CT | CCCC
CCCC
Pro
Pro
CCCC
Ala
Ala
Ala
CCCC
CCCC
Pro
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCCC | CTCC
Val
TCCC
Ser
Gly
BD
TTY
TY
TTY
CTGU
ATG
TTY
ATG
TCCC
GCG
Gly
SD
TTY
TTY
TTY
TTY
TTG
CTGU
ATG
TTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC |
AAC
AG
TCG
Ser
TCG
Ser
ACG
CTA
ASIN
AAT
AAT
AAT
AAT
AAT
AAT
AAT
CTA
AG
CTA
AG
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA | CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTT
TTA
CCT
Ala
CCT
Ala
CCT
Ala
CCT
Ala
CCT
Ala
CCT
CTA
CCT
CCT
CCT
CCT
CCT
CCT | ACC
ACC
Thr
ACC
ACC
ACC
TTP
TTP
CTA
ACC
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
CTA
ACC
ACC | ATA
ATG
CYs
ATG
CYs
ATG
CYs
ATG
CYs
CYs
CYs
CYs
CYs
CYs
CYs
CYs | TCC
Ser
AAG
Glu
GAG
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA | ACA
Thr
Thr
CTG
CTG
CTG
CVal
GAA
Glu
GTC
CVal
GAA
Glu
GTC
CVal
CTA
CCTA
CCTA
CCTA
CCTA
CCTA
CCTA
CCTA
 | AAT
AAT
TCCC
Ser
TTCC
Phe
AAT
ASD
ACC
Thr
ATT
CAC
ACC
TTA
CAC
CTT
TTA
CAC
CTT
TTA
CAC
CCT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTTA
CCTT
TTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTTA
CCTT | GAC
Asp
Gla
Gla
AACC
Gla
AACC
Gla
CTG
Gla
CTG
Gla
CTG
Gla
CTG
Gla
CTG
Gla
CTG
CAC
Gla
CTG
Gla
CTG
CAC
CAC
Gla
CAC
CAC
CAC
CAC
CAC
CAC
CAC
CAC
CAC
CA | ACT
Thr
Thr
GGC
Clau
CCCA
Pro
CCCC
Pro
CCC
Leu
CCCA
Pro
CCCA
Pro
CCCA
A
Sor
TCA
Sor
TCA
Sor
TCA
AGA
TTCA
AGGC
GGT
TTCC
ACA
AGGC
GGT
CCCA
CCCA | GIA
GIY
AAC
Aan
TCA
Aan
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TCA
TCA | ATC
III
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
ACC
ACC
ACC
ACC
ACC
ACC
AC
 | ACC
Thr
AAG
CC
Thr
Lys
GGG CC
Ly
CC
GGC CC
Ly
CC
GGC CC
Ly
CC
GGC CC
Lys
CC
GGC CC
Lys
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC | ATC
IIe
CTC
Leu
AAC
AAS
AS
AS
AS
AS
AS
AS
AS
AS
A | CGT
Arg
AGC
Ser
CAA
Gln
ATT
TTC
CAG
Gln
CAA
Gln
CAA
Gln
CCAG
Gln
TCC
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA | TGG
Trp
ATAA
ATAA
AAGC
Ser
Ala
GGG
CCT
Ala
CCCC
Pro
TAA
CCCT
TTCA
AGG
CCT
TTCA
AGA
CCTAT
TCAA
CCAA
TAAA
 | TIC
Phe
AAC
Asp
GAC
Asp
GGC
Lys
CCAC
Bis
AAG
AAG
Thr
TCA
TTA
AAG
CCC
CCC
GAG
AAAA
CCAC
CCAC | 1152
360
1224
384
1286
408
1368
432
1440
456
1512
480
1554
526 |
| Val
TTCC
Phe
Phe
ECCT
Pro
Pro
ATT
Thr
ACC
Thr
ACC
CTC
Thr
ACC
CTCT
TAT | AAA
Thr
Thr
Lys
GTC
Val
ATC
GCV
In*
AAT
ASD
CTG
GGLY
In*
AAT
CTG
GGAT
TAG
GGAT
TGA
AAAT
AAA
CTG
TGA
AAAT
TGA | Gly
Gly
AAC
AAG
AAG
AAG
AAG
AAG
AAG
AAG | GAT Asp
CAG GIN
AGO
AGO
AGO
AGO
AGO
AGO
AGO
AGO |
AAG
Lys
AGT
Ser
GAG
Glu
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
GIU
GIU
GIU
GIU
GIU
GIU
GIU
GIU
GIU | GAC
Asp
CTC
Leu
GAT
Asp
CTA
Asp
CTA
Asp
CTA
Asp
CTA
Asp
CTA
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG | ACT
Ser
CCCG
Pro
CCCG
Pro
CCCG
Ala
Assn
CCCG
Assn
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG | GTG
Val
TCC
Ser
Gly
BD
TTY
TTT
CTG
Gly
BD
TTT
TTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG |
AAC
ASD
TCG
Ser
TCG
Ser
ACG
Ser
Thr
AAT
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
ASD
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA | CTG
GAG
Glu
TAT
Tyr

GCT
Ala
GCT
Ala
GCT
Ala
GCT
TAT
TTA
TTA
TTA
TTA
TTA
TTA
TTA
CAA
GCT
TTA
CAA
GCT
TTA
TTA
TTA
TTA
TTA
TTA
TTA
TTA
TTA
T | ACC
Thr
ACC
Arg
Trp
Trp
CTA
Arg
CTA
CTA
Arg
CTA
CTA
Arg
CTA
CTA
Arg
CTA
CTA
Arg
CTA
CTA
Arg
CTA
CTA
Arg
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA | ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CYA
ATCC
CCCA
ATCC
CYA
ATCC
CCCA
ATCC
CCCA
ATCC
CCCA
ATCC
CCCC
CTCC
CYA
ATCC
CCCC
CTCC
CT | TCC
Ser
AAG
Glu
GAG
Glu
CAA
Gla
Glu
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
Gla
CAA
CAA
Gla
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA | ACA
Thr
Thr
CTG
Leu
Val
GAA
Glu
GTC
Val
GAA
Clu
GTC
Val
GAA
Clu
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
 | AAT
ASD
TCC
Ser
TTCC
Phe
AAT
ASD
ACC
TTC
ATT
TCA
ATT
TTA
ATT
TCA
ATT
TTA
CTT
TTA
CCT
CCC
CCT
TTCC
CCC
C | GAC
Asp
CAC
Gln
Asp
CAC
Gln
CTG
CI
CAC
CI
CI
CI
CI
CI
CI
CI
CI
CI
CI
CI
CI
CI | ACT
Thr
Thr
GGC
Qly
CCCA
Pro
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA | CGA
Gly
AACC
Ile
AACC
ICA
ACC
Ser
TCT
Ser
TCTA
CCC
CCC
TCTA
CCC
CCC
TCTA
ACC
CCC
C | ATC
III
ACCC
Thr
ACCC
Thr
Ser
TTT
Phe
CCT
ACC
CCT
ACCC
CCT
CCT
CCT
CC
 | ACC
Thr
AACC
Cleu
Cleu
Cleu
Cleu
Cleu
Cleu
Cleu
Cl | ATC
III
CTC
Leu
AAC
AAA
CAT
LEU
AAAC
CAT
AAA
CAT
THE
CAG
GIN
AAC
CAT
CAG
GIN
AAC
CAT
CAT
CAT
CAT
CAT
CAT
CAT | CGT
Arg
AGC
Ser
CAA
Gln
CAA
Gln
CAA
Gln
CAA
Gln
CCAC
Gln
TCC
CCAC
CCA
Gln
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CC | TOG
Trp
ATTA
ATTA
ATTA
ATTA
ATTA
AAGC
CTT
TIA
AAGC
CCCC
Pro
TAA
AGC
CCTC
TTA
ACC
TTA
ACC
TTA
AACC
TTA
AACC
TTA
TTT
TTT
 | TTC
Phe
AAC
Asn
GAC
Asn
GAC
Asn
GAC
Lys
rol
Lys
rol
CAC
Bis
ACA
Thr
TCA
ATG
AAAG
CTCA
CTCG
CTCG
CTCG
CTCG
C | 1152
360
1224
384
1296
408
1368
432
1440
456
1512
480
1554
504 |
| Val
TTCC
Phe
Pro
ATT
Pro
Pro
ATT
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
TTT
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT | AAA
Thr
Thr
Lys
GTC
Val
ATC
GIV
Ile
GTC
GIV
In*
AAT
AAA
CCCA
TTGC
GAA
AAAT
AAAC
CCA
TTGC
CAA
TTAG
CCA
TTGC
CAA
TTAG
CCA
TTGC
CAA
TTAG
CCA
TTGC
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
CAA
TTAG
TTAG | Gly
Gly
AAC
Lys
ATG
ATG
ATG
ATG
ATG
ATG
ATG
ATG | GAT Asp
GAT Asp
CAG GLA
CAG GLA
CTG GLA
CTG GLA
CTG GLA
CTG GLA
CTG GLA
CTG CTG CTG CTG
CTG CTG CTG CTG
CTG CTG CTG CTG
CTG CTG CTG CTG CTG
CTG CTG CTG CTG CTG
CTG CTG CTG CTG CTG
CTG CTG CTG CTG CTG CTG
CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG |
AAG
Lys
AGT
Saf
GAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
Glu
AAG
AAG
AAG
Glu
AAG
Glu
AAG
Glu
AAG
AAG
AAG
AAG
AAG
AAG
AAG
AAG
AAG
AA | GAC
Asp
CTC
Leu
GAT
Asp
GTA
Asp
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CT | ALL
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG
CCCG | CTG
Val
TCCC
Ser
Cly
Sp
TAT
Tyr
CTG
Leu
CTG
Leu
CTG
Leu
CTG
Leu
CTG
Cly
Scr
Tyr
Tyr
Tyr
TGC
Cly
Scr
Tyr
Tyr
CTG
Leu
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
Tyr
CTG
Cly
Scr
Tyr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
Tyr
CTG
Cly
Scr
T
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C |
AACC
ACC
Ser
ACC
Ser
ACC
CTA
AAT
AAT
AAT
AAT
AAT
AAT | CTG
GAG
Glu
TAT
Tyr

GCT
Tyr

GCT
Tyr

GCT
Ala
Glu
GAA
Glu
GAA
Glu
ACA
Thr
TGT
TCA
CAC
TTG
TTA
TTG
CAC
GCT
CAC
GAC
GCT
CAC
GCT
CAC
CAC
CAC
CAC
CAC
CAC
CAC
CAC
CAC
C | ACC
The
ACC
The
ACC
Trp
Trp
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA | ATCC
CYA
ATCC
CYA
Met
TGTTC
CYA
ATA
CCA
Pro
ATA
ATA
CCA
CYA
ATA
CCA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
ATA
CYA
CYA
ATA
CYA
CYA
ATA
CYA
CYA
CYA
CYA
CYA
CYA
CYA
CYA
CYA
CY | TCC
Ser
AAAG
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
CAA
CGLU
CAA
CAA
CGLU
CAA
CAA
CAC
CAA
CAC
CAA
CAC
CAA
CAC
CAA
CAC
CAA
CAC
CAA
CAC
CAA
CAC
CAA
CAC
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAAA
CAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAA
CAAAAA
CAAA
CAAAAAA | ACA
Thr
Thr
CTG
CTG
CTC
Val
GAA
Glu
GTA
Val
GAA
Clu
Val
CTC
Ser
Ser
TCT
Ser
TCT
TAC
TAC
CTAC
CAA
AAA
ATA
ACA
TTC
CTG
GAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA
 | AAT
Asn
TCCC
Ser
TTCC
Phe
AAT
Asn
GCCC
ALSS
ACC
Thr
ASSN
ACC
Thr
ASSN
ACC
TTTA
ATT
CTT
TTA
CTT
TTTA
CTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTTA
CCTT
TTA
CCTT
TTTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
TTA
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTT
CCTTT
CCTTT
CCTTT
CCTTT
CCTTT
CCTTT
CCTTTTC
CCTTTT
CCTTT
CCTTTT
CCTTTTC
CCTTTTTCTT
CCTTTTTC | GAC
Asp
CAG
Gln
AAC
CAG
Gln
AAC
CAG
CIN
CTG
Loui
TAT
Tyr
CCC
GGAC
AGT
CTG
GAC
CAG
GIN
CTG
Loui
TAT
Tyr
CCC
GGAC
CAG
GIN
CAG
CON
CON
CON
CON
CON
CON
CON
CON
CON
CON | ACT
Thr
Thr
CCA
Cly
CCA
Pro
CCA
Pro
CCA
Pro
CCA
Pro
CCA
Pro
CCA
Pro
CCA
CCA
Pro
CCA
CCA
Pro
CCA
CCA
Pro
CCA
CCA
CCA
CCA
CCA
CCA
CCA
CC | CIA AAC
CIA
AAC
AAC
AAC
AAC
AAC
AAC | ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
TTT
Pho
CCT
TTT
Pho
CCT
TTT
TTT
Pho
CCT
TTT
TTT
TTT
TTT
TTT
TTT
TT
 | ACC
Ser
ACC
Thr
AAC
CTG
Lys
GGG Cly
CTG
Lys
GGC Cly
GGC Cly
GGC Cly
GGC Cly
GGC Cly
CTG
Lys
GGC Cly
CTG
CTG
CCC
CTG
CCC
CCC
CCC
CCC | ATC
III
CTC
Leuu
AAC
AAC
AAC
AAC
AAC
AAC
AAC
A | CAT
Arc
Acc
Ser
CAA
Cln
ATT
CAG
Gln
ATT
CAG
Gln
CAA
CAG
Gln
CAA
CAG
Gln
CCAG
CAG
CAA
CAG
CTT
CCAA
CTT
CCAA
CTT
CCAA
CTTTC
CAA
CTTTC
CAA
CTTTCC
CAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAA
CTTTCCCAAA
CTTTCCCAAA
CTTTCCCAAA
CTTTCCCAAA
CTTTCCCAAAA
CTTTCCCAAAA
CTTTCCCAAAAAAAA | TAG
Trp
ATA
ATA
ACC
ACC
ACC
ACC
ACC
ACC
 | TTC
Phe
AAC
Asn
GAC
Asn
GGC
Lys
rol
CAC
Bis
ACA
Thr
TGA
ATG
CAC
GAC
CTA
GAC
AAT
TCA
ATG
CAC
AAT
TAA
TAA
TAA
CAC
CTA
ASN
ASN
CAC
CAC
CAC
CAC
CAC
CAC
CAC
CAC
CAC
CA | 1152
360
1224
384
1296
408
1368
432
1440
456
1512
480
1554
504
1656
526 | | | | | | | | | | | | | | | | | | | | | | | |
| Val
TTCC
Phe
Pro
ATT
Pro
Pro
ATT
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
TH
ACC
TH
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
AC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC
THR
ACC | AAAA
Thr
Thr
Thr
Crc
GTC
Val
ATC
GTC
GTC
GTC
GTC
GTC
GTC
GTC
G | Gly
Gly
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | GAT TAAAA SP
CAG GLA
GLAU
CTG GLAU
CTG GLAU
CTG GLAU
CTG GLAU
CTG CAG ACCA
CTG AAAA TAGT
TAAAA TGT TAAAA TGT TAAAAA TGT TAAAAA TGT TAAAA TGT TAAAAA TGT TAAAA TGT TAAAA TGT TAAAA TGT TAAAA TGT TAAAAA TAAT TAAAAA TAAAA TAAAA TAAAAA TAAAAA TAAAA TAAAAAA | AAG
Lys
AGT
Jar
GAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
AAG
AAG
AAG
AAG
AAG
AAG
AAG
AAG
AA | GAC
Asp
CLeu
GAT
Asp
GAT
Asp
CTA
Asp
CTA
Asp
CTA
Asp
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA
CTA | CCG
Pro
CCG
Pro
CCG
Pro
CCG
Pro
CCG
Pro
CCG
CCG
Pro
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CC | CTG
Val
TCCC
Ser
Gly
sp
Gag
Gly
sp
Tyr
Tyr
CTG
Leu
Cag
Arg
Atg
Het
TCCC
Ser
TTG
Cag
Cag
Cag
Cag
Cag
Cag
Cag
Cag
Cag
Cag | AACC
ACC
Ser
ACC
Ser
ACC
CTA
ASIN
GTT Thr
AASIN
GTT Thr
CVal
CTA
ASIN
CTA
ASIN
CTA
ASIN
CTA
ACC
CCA
ACC
CCA
CCA
CCA
CCA | CTG
CAG
CAAG
CIU
TAT
Tyr
Tyr
CCT
Ala
CCT
Ala
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CC | ACC
The
ACC
The
Arcq
Trp
Trp
CTA
Levu
CTG
Levu
CTG
Levu
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG | ATC
CY
ATC
CY
ATC
CY
CY
CY
CY
CY
CY
CY
CY
CY
CY
CY
CY
CY | TCC
Ser
AAAG
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
Glu
CAA
CAA
Glu
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA | ACA
Thr
Thr
CTO
Levu
GAA
Glu
GTC
Val
GAA
Colu
CTA
CTA
CTAC
CTAC
CTAC
CTAC
CTAC
CTAC | AAT
Asn
TCC
Ser
TTC
Phe
AAT
Asn
GCC
ALS
ACC
TTA
CTT
ATT
CCT
TTA
CCT
TTA
CCT
TTA
CCT
TTA
CCT
TTA
CCT
TTA
CCC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
CCT
ACC
ACC | GAC
Asp
CAG
Gln
AAC
CAG
Gln
AAC
CAG
Ser
CTG
Leu
TAT
Tyr
CCC
GGAG
CAG
CAG
CAG
CAG
CAG
CAG
CAG
CA | ACT
Thr
Thr
CCCA
Cly
CCCA
Pro
CCCA
Pro
CCCA
Pro
CCCA
Pro
CCCA
Pro
CCCA
CCCA
Pro
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCC | CIA
AACC
AAAC
AAAC
AAAC
AAAC
AAAC
AAAC
AAAC
AAAC
AAAC
AAAC
AAAC
TCA
TCA | ACC
Thr
ACC
Thr
ACC
Thr
Thr
CCT
Prove
CCT
Prove
CCT
Prove
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CC | ACC
Ser
ACC
Thr
Lys
CIG
Lys
CIG
Lys
CIG
Lys
CIG
CIG
Lys
CIG
CIG
CIG
CIG
CIG
CIG
CIG
CIG | ATC
III
CTIC
Leu
AAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
III
CAC
CAT
IIII
CAC
CAT
IIII
CAC
CAT
IIII
CAC
CAT
IIII
CAC
CAT
IIIII
CAC
CAT
IIIII
CAC
CAT
IIIIII
CAC
CAT
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII | CAT
Arg
Acc
Ser
CAA
Cln
ATT
CAG
CAG
CAG
CAG
CAG
CAG
CAA
CCAC
CAA
CCAC
CAA
CCAC
CAA
CCAC
CAA
CCAC
CAA
CCAC
CAA
CCAC
CAA
CCAC
CAA
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCAC
CCC | TAG
Trp
ATA
ATA
ACC
ACC
ACC
ACC
ACC
ACC | TTC
Phe
AAC
Aan
GAC
Aan
GAC
Aan
GAC
Aan
GAC
Aan
GAC
Aan
Cac
Cac
Bis
ACA
AAC
Cac
Cac
Cac
Cac
Cac
Cac
Cac
Ca | 1152
360
1224
384
]296
408
1368
432
1440
456
1512
480
1554
504
1656
526 |
| Val
TTCC
Phe
CCT
Pro
ATT
Fro
ACC
Thr
ACC
Thr
ACC
CTC
CCC
CCC
CCC
CCC
CCC
CCC | AAAA
Thr
Thr
Thr
Thr
Thr
Chr
Qual
AATC
CVal
AATC
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVal
CVA
CVA
CVA
CVA
CVA
CVA
CVA
CVA | Gly
Gly
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | CAC GIN
CAC GIN
CAC GIN
CTC | AAG
Lys
AGT
Sgr
GAG
Glu
AAC
Glu
AAC
Glu
AAC
Glu
AAC
Glu
AAC
Glu
AAC
Glu
AAC
Glu
AAC
Glu
AAC
Glu
AAC
Glu
AAC
Glu
AAC
GLUSS
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC | GAC
Asp
CLou
GAT
CLou
GAT
Asp
GTA
Asp
CTG
GTA
Asp
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG | CCG
Pro
CCG
Pro
CCG
Pro
CCG
Pro
CCG
Pro
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CC | GTG
Val
TOC
Ser
Gly D
TTY
TYT
CTG
U
CTG
U
ATG
ATG
ATG
CTG
CCC
TTCC
ATG
CTG
CCC
TTY
TTYT
TTYT
TTYT
TTYT
TTYT | AAC
ASS
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
Ser
TCG
Ser
Ser
TCG
Ser
Ser
Ser
Ser
Ser
Ser
Ser
Ser | CTG
CAG
CAG
CAG
CAG
CA
CA
CA
CA
CA
CA
CA
CA
CA
CA | AGG
AFG
AFG
TTP
TTP
TTP
TTP
TTP
TTP
TTP
TTP
TTP
TT | ATG
CY
ATG
CY
ATG
CY
CY
ATG
CY
CY
ATG
CY
CY
ATG
CY
CY
CY
ATG
CY
CY
CY
CY
CY
CY
CY
CY
CY
CY
CY
CY
CY | TCC
Ser
AAG
Glu
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GAL
GLU
GLU
GLU
GLU
GLU
GLU
GLU
GL | ACA
Thr
Thr
CTG
Leu
GTC
Val
GTC
Val
GTC
Val
GTC
Val
GTC
Val
CTC
GAA
GTU
CTC
GAA
GTU
CTC
GAA
CTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
GTC
CTC
C | AAT
Asn
TCCC
Ser
TTCC
Phe
AAT
Asn
CCCa
ATT
CCCC
CCC
AAT
Asn
CCCC
AAT
Asn
CCCC
AAT
Asn
CCCC
AAT
Asn
CCCC
AAT
Asn
CCCC
AAT
Asn
CCCC
AAT
Asn
CCCC
AAT
Asn
CCCC
AAT
Asn
Asn
Asn
Asn
Asn
Asn
Asn
Asn | GAC
Asp
CGAG
GIn
AACC
Asn
GIV
CTG
GIN
CTG
CCG
CAG
CGIN
CTG
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG | ACT
The
GGC OLY
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAAA
CCAA
CCAA
CCAAA
CCAAA
CCAAA
CCAAAA
CCAAA
CCAAA
CCAAA
CCAAAAAA
CCAA | GIA
GIY
AACC
III
TCA
Sor
TTT
TCA
Sor
TTT
TTA
Sor
TTT
TTA
Sor
TTT
TTA
Sor
TTT
TTA
Sor
TTT
TTA
Sor
TTTT
TCA
Sor
TTTT
TCA
Sor
TTTT
TCA
Sor
TTTT
TCA
Sor
TTTT
TCA
Sor
TTTT
TCA
Sor
TTTT
TCA
Sor
TTTT
TCA
Sor
TTTT
TCA
Sor
TTTTT
TCA
Sor
TTTTT
TCA
Sor
TTTTTTT
TCA
Sor
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT | ACC
IIIe
ACC
Thr
ACC
Thr
Thr
ACC
Thr
CTA
ACC
Thr
CTA
ACC
Thr
CTA
ACC
CTA
TTTT
CCT
TTTT
CCT
TTTT
CCT
TTTT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CCT
C | ACC Ser
ACC The
Lys CC U
SHI CC ACC ACC ACC ACC ACC ACC ACC ACC ACC | ATC
III
CTC
Leu
AAC
Leu
CAT
CAT
CAT
CAT
CAT
CAT
CAT
CAT | CALL AFG
AFG
AFG
CALL AFG
GIN
ATT
TIC
CALL
GIN
CALL
GIN
CALL
GIN
CALL
GIN
CALL
GIN
CALL
GIN
CALL
GIN
CALL
GIN
CALL
GIN
CALL
CALL
CALL
CALL
CALL
CALL
CALL
CAL | TOG
Trp
ATA
III
ACCC
CT
ACC
CCT
CCA
ACCC
CCA
CCA | TTC
Phe
AAC
Aan
GAC
Lys
GGC
CIA
GGC
CAC
Bis
ACA
TTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CAC
CTC
CT | 1152
360
1224
384
]296
408
1368
432
1440
456
1512
480
1554
526 |
| Val
TTCC
Pho
CCT
Pro
ATT
Fro
ACC
Thr
ACC
Thr
ACC
CTC
CCT
ACC
CTC
CCT
CCT
CCT | Thr
Thr
Thr
Thr
AAAA
Lys
GTC
Val
AAC
Val
CVal
CVal
CVal
CVal
CVal
CVal
CVal | Gly
Gly
AAC
AAG
Lys
AAC
Lys
AAC
AAG
AAC
AAG
AAC
AAC
AAG
AAC
AAC
AAC | GAT A Sp
CAG GIN
ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC | AAG
Lys
AGT
Sgr
GAG
Glu
AAC
Glu
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
GIU
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | GAC
Asp
GAT
Levu
GAT
CTC
Levu
GAT
CASP
GTA
Asp
CTC
CTC
CTC
CTC
CVal
CVal
CVal
CVal
CVal
CTC
CTC
CTC
CTC
CVal
CVal
CVal
CVAL
CTC
CTC
CVAL
CVAL
CVAL
CVAL
CVAL
CVAL
CVAL
CVA | CCCC
Ser
Ser
Ser
Pro
CCCC
Pro
CCCC
Pro
CCCC
Pro
CCCC
Pro
CCCC
Pro
CCCC
CCCC
Pro
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCC
CCCC
CCCCC
CCCCC
CCCC
CCCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCC
CCCCC
CCCCC
CCCCCC | GTG
Val
TCC
Ser
Gly
DAT
TYT
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
C
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
CTG
U
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C
C | AAC
ASD
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
TCG
Ser
Ser
Ser
TCG
Ser
Ser
Ser
Ser
Ser
Ser
Ser
Ser | CTG
CAG
GAG
GAG
GU
TATT
Tyr
Tyr
Tyr
Ala
GCT
Ala
GCT
Ala
GCT
Ala
GCT
Ala
CTT
CTT
CTT
Ala
CTT
CTT
Ala
CTT
CTT
Ala
CTT
Ala
CTT
Ala
CTT
Ala
CTT
Ala
CTT
Ala
CTT
Ala
CTT
Ala
CTT
Ala
CTT
CTT
CTT
Ala
CTT
CTT
CTT
Ala
CTT
CTT
CTT
Ala
CTT
CTT
CTT
CTT
Ala
CTT
CTT
CTT
CTT
CTT
CTT
Ala
CTT
CTT
CTT
CTT
CTT
CTT
CTT
CT | AGG
AFG
AFG
TTP
TTP
TTP
TTP
TTP
TTP
TTP
TTP
TTP
TT | ATGC CYM ATGC ATGC ATGC ATGC ATGC ATGC ATGC ATGC | TCC
Ser
AAG
GLU
GAL
GLU
GAL
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
GLU
CAA
CAA
GLU
CAA
CAA
GLU
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CA | ACA
Thr
Thr
CTG
U
U
U
U
U
U
U
U
U
U
U
U
U
U
U
U
U
U
U | AAT
Asn
TCCC
Ser
TTCC
Phe
AAT
Asn
CCC
ACC
ATT
CAC
CCCT
ACA
CCTT
ACC
CCTT
AGC
CCTT
AGC
CCTT
AGC
CCCT
AGC
CCCT
AGC
CCCT
AGC
CCCT
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
CCCC
AGC
CCCC
CCCC
AGC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCCC | GAC
Asp
CGC
GIn
AACC
CAG
GIN
CTG
CCG
CAG
CGIN
CTG
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG
CCG | ACT
The
GGC GGC
GGC GGC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
AGC
CCCC
CCCC
AGC
CCCC
CCCC
AGC
CCCC
CCCC
AGC
CCCC
CCCC
AGC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCC
CCCC
CCCC
CCCCCC | GIA
GIY
AACC
III
TCA
Soft
TCA
Soft
TTT
TPho
GAN
CCCCCTGAACTAACCGATTGGA
CAACCGATTGGA
CCCCCTGAAATACCGATTGGA
CCCCCTGAAATACCGATTGGA | ACC
IIIe
ACC
Thr
ACC
Thr
ACC
Thr
Thr
ACC
Thr
ACC
Thr
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
Thr
ACC
ACC
ACC
ACC
ACC
ACC
ACC
AC | ACC TALL CLUB COLOR CLUB CLUB CLUB CLUB CLUB CLUB CLUB CLUB | ATC
III
CTC
Leu
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | COT Arcg
Arcg
MGC Ser
Gln ATT TIC
CAA ACA
Gln CAA
Gln CAA
Gln CCAA
Gln CCAA
Gln TCCC
CAA ACA
GGAC
AACA GGAC
AACA GGAC
ACA ACA GGAC
ACA ACA ACA GGAC
ACA ACA ACA ACA ACA ACA ACA ACA ACA A | THE ATAA AACC CAA AACA AAACA AACA AACA AACA AACA AACA AACA AACA AAACA AACA AACA AACA AACA | TIC
Phe
AAC
Asp
GCC
Cly
Asp
GCC
Cly
CAC
Bis
ACA
Thr
TCA
ASA
CCC
CCC
CCC
CCAC
CCAC
CCAC
CCAC | 1152
360
1224
384
1296
408
1368
432
1440
456
1512
480
1554
526 |
| Val
TTCC
Pho
CCT
Pro
ATT
Fro
ACC
Thr
ACC
Thr
ACC
CTT
ACC
CCT
CCT
ACC
CCT
CCT | GTC
Val
GTC
Val
ATC
GTC
GTC
GTC
GTC
GTC
GTC
GTC
CAT
Fro
CAT
TAG
GTC
CAT
TAG
GTC
CAT
TAG
GTC
CAT
TAG
GTC
CAT
TAG
GTC
CCA
TAG
TAG
TAG
TAG
TAG
TAG
TAG
TAG
TAG
TA | Gly
Gly
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | GAT
Asp
CAG
CIL
CLOU
CCC
CLOU
CCC
CLOU
CCC
CLOU
CCC
CLOU
CCC
CLOU
CCC
CLOU
CCC
CCC
CLOU
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC
C | AAG
Lyse
AGT
Ser
GAG
GIU
AAG
GIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAA
AAAAAAG
CIU
AAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAA | GAC
Asp
GAT
Asp
GTA
Ser
Val
GTA
Ser
CTC
GTG
GTA
Asp
Val
GTA
Asp
Val
CVal
Asp
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CTC
CT | CCCC
Ser
Ser
Ser
Pro
CCCC
Pro
CCCC
Pro
CCCC
Pro
CCCC
Pro
CCCC
Pro
CCCC
Pro
CCCC
CCCC
Pro
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCC
CCCCCC | GTG
Val
TCC
Ser
Ser
Tyr
Tyr
Tyr
CTG
GGG
Ser
Tyr
Tyr
Tyr
Tyr
Tyr
Tyr
Tyr
Tyr
Tyr
Ty | ACC
Ser
Thr
AAT
AAT
AAT
AAT
AAT
AAT
AAT
AAT
AAT
AA | CTG
CAG
GAG
GAG
GIU
TATT
Tyr
Tyr
C
SCT
Ala
GCT
Ala
GCT
Ala
CCT
CCT
Ala
CCT
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
Ala
CCT
CCT
CCT
Ala
CCT
CCT
CCT
CCT
CCT
CCT
CCT
CC | ACC
ACC
ACC
ACC
ACC
ACC
ACC
ACC | ATGC CYM ATGC ATGC ATGC ATGC ATGC ATGC ATGC ATGC | AAAG
GLU
GLU
GLU
GLU
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
GLU
CAAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
CAA
C | ACA
Thr
Thr
CTG
Lau
GAA
Glu
GAA
Calu
GAA
Calu
GAA
Calu
Calu
Calu
Calu
Calu
Calu
Calu
Calu | AAT
AST
TCCC
Ser
TTCC
Ser
TTCC
Ser
AAT
AST
TCCC
Ser
AAT
AST
TCCC
Ser
AAT
AST
TCCC
Ser
CAA
AST
AST
AST
AST
AST
AST
AST
AST
AST
A | GAC
Asp
Gla
Gla
Gla
Gla
Gla
CCC
Gla
Ser
CTG
Gla
Ser
CTG
Leu
CTG
Leu
TAT
Tyr
CCC
CGG
AACC
Ast
TTYr
CCC
CGG
AACC
CAG
Gla
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CCC
CCGG
CAG
CC
CL
Ser
CTG
CL
Ser
CTG
CL
Ser
CTG
CCC
CCGG
CAG
CCC
CCC
CCG
CL
Ser
CTG
CCC
CCGG
CAG
CCC
CCC
CCG
CCC
CCC
CCGG
CAG
CCC
CCC | ACT
Thr
Thr
CCCA
Gly
CCCA
Gly
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA | GIA
GIY
AGC
ASA
ASA
ASA
ASA
ASA
ASA
ASA
ASA
ASA
AS | ACC The ACC ACC ACC ACC ACC ACC ACC ACC ACC AC | ACC The Ser Concernment of the Ser Concernmen | ATC
III
CTC
Leu
AAC
AAC
AAC
AAC
AAC
AAC
AAC
AA | CGT
AFG
GIN
ACC
GIN
ATT
FIC
GIN
ATT
FIC
GIN
CAA
GIN
CCAA
GIN
CCAA
GIN
CCAA
GIN
CCAA
GIN
CCAA
GIN
CCAA
GIN
CCAA
CCTTTTTC
CCAA
CCAA
GIN
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CCAA
CC | ATA A
ATA A
IIIe
ATA A
ATA A
A
A
A
A
A
A
A
C
C
C
C
C
C
C
C
C
C
C
C
C | TIC
Phe
AAC
Asp
GCC
Cly
CAC
Bis
Asp
GCC
Cly
CAC
Bis
ACA
Thr
TCA
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC
CCC | 1152
360
1224
384
1296
408
1368
432
1440
456
1512
480
1554
526 |
| Val
TTCC
Pho
CCT
Pro
ATT
Fro
CCC
Pro
ATT
ACC
Thr
GCC
CCC
CCT
TTT
ACC
CCC
CCC
CCC
C | Thr
Thr
Thr
Thr
Thr
Thr
Thr
Chr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Corr
Co | Gly
Gly
AACC
ABL
Lys
ATG
ACC
ACC
ACC
ACC
ACC
ACC
ACC
ACC
ACC
AC | GAT
Asp
CAG
CIN
CAG
CIN
CTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CCTG
CIN
CIN
CIN
CIN
C
CIN
CIN
CIN
CIN
CIN | AAG
Lys
AGT
Jer
GAG
GIU
AAG
GIU
AAG
GIU
AAG
GIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAAG
CIU
AAAG
CIU
AAAG
CIU
AAAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAAG
CIU
AAG
CIU
AAG
CIU
AAAG
CIU
AAAG
CIU
AAAG
CIU
AAG
CIU
AAAG
CIU
AAAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
AAG
CIU
CIU
AAG
CIU
CIU
AAG
CIU
AAG
CIU
CIU
CIU
CIU
CIU
CIU
CIU
CIU
CIU
CIU | GAC Asp
GAC Asp
GAT Asp
GTA Sor
GTA Sor
GTA Asp
Asp
Asp
Asp
Asp
Asp
Asp
Asp | CCCC
Ser
Ser
Ser
Pro
SCC
Pro
SCC
Ala
AAG
AAG
CAG
CAG
CAG
CAG
CAG
CAG | GTG
Val
TCC
Ser
Ser
Tyr
Tyr
Tyr
Tyr
CTG
Leu
Arg
Arg
Arg
Mat
TCC
Ser
TGC
Arg
Arg
Arg
Arg
CCC
TGC
Arg
Arg
Arg
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
CCC
TCC
TCC
CCC
TCC
TCC
CCC
TCC
TCC
CCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TCC
TC | ACC
Ser
Thr
AAT
AAT
AAT
AAT
AAT
AAT
AAT
AAT
AAT
AA | CTC
CAG
GIU
TATT
TYT
CCT
ALA
CTAT
CCT
ALA
CTAT
CAC
ALA
CTAT
CAC
CAA
CAA
CTAT
CAC
CAA
CAA | ACC
ACC
ACC
ACC
ACC
ACC
ACC
ACC | ATC CY ATA ATC CY CY ATA ATC CY CY ATA ATC CY ATA ATA ATC CY ATA ATC CY ATA ATA ATC CY ATA ATC CY ATA ATC CY ATA ATA ATA ATC CY ATA ATA ATA ATA ATA ATA ATA ATA ATA AT | AAGG
Glu
CAAAGG
Glu
CAAAGG
Glu
CAAAGG
Glu
CAAAGG
Glu
CAAAGG
Glu
CAAAGG
Glu
CAAAGG
Glu
CAAAGG
Glu
CAAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAGG
CAAG | ACA
Thr
Thr
CTG
Law
GCC
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG
CTG | AAT
AAT
TCCC
Ser
TTCC
Ser
TTCC
Ser
TTCC
TCCC
Ser
AAT
AAT
AAT
AAT
AAT
AAT
AAT
AA | GAC
Asp
Gla
Gla
Gla
Ser
CTG
Gla
Ser
CTG
Gla
Ser
CTG
CCC
CGG
CAC
CCC
CCC
CCC
CCC
CCC
CCC
CC | ACT
Thr
Thr
CCCA
Gly
CCCA
Gly
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA
CCCA | GIA
GIY
ANCC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ANC
ASS
ASS
ASS
ASS
ASS
ASS
ASS
ASS
ASS
AS | ACC TALL ACC TALL AND | ACC The Ser Control of | ARC Leve CAT CALL IN CALL AND | CONT
AFG
GIN
ACC
GIN
ATT
FIC
GIN
CAA
GIN
CCAA
GIN
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CCAA
CONT
CONT
CONT
CONT
CONT
CONT
CONT
CONT | ATA A
ATA A
ATA A
ATA A
A
A
A
A
A
A
A
A
A
A
A
A
A | TIC
Phe
AAC
Asp
GCC
Cly
Asp
GCC
Cly
Asp
GCC
Cly
CAC
Bis
AC
ASP
CCC
CCC
CCC
CCC
CCCC
CCCC
CCCC
C | 1152
360
1224
384
1296
408
1368
432
1440
456
1512
480
1554
526 |

Figure 2. DNA and deduced amino acid sequences of TM-CEA cDNAs. The DNA sequences were compiled from double-strand dideoxy termination reactions on both strands of all clones. The full sequence from nucleotide 1 to 3,464 corresponds to TMI-CEA. The initiating methionine residue of the apoprotein is at position -34 relative to the proposed amino terminus of the mature protein at +1. Symbols \downarrow demarcate the likely beginning and end of the first loop-domain based on similarity with CEA and NCA; solid bars note consensus sequences for N-linked glycosylation sites; are cysteine residues in the extracellular portion of the molecule; Cys) is the only unpaired cysteine residue in the extracellular region; the thick barberpole line indicates a hydrophobic transmembrane region. The open box around nucleic acid sequence corresponds to the "half" loop-domain (LD IIa) region deleted in TM2-CEA; the single amino acid that is altered as a result of "half" loop-domain deletion, i.e., aspartic acid (Asp), is shown as a split residue above the corresponding split codon in the TM1-CEA sequence; dots on either side of the Asp residue indicate that all adjacent sequences are as shown for TMI-CEA. The 53 nucleotides that are deleted in the cytoplasmic region of TM3-CEA are indicated by the shaded portion of the broken box. The new amino acid, serine (Ser), that is generated at the splice junction, and subsequent new amino acids that are translated by the induced frameshift, are shown above their corresponding new codons; dots amino-terminal to Ser indicate that all sequence to that point is identical to TM1-CEA. Termination of translation is indicated by ***. Solid and broken triangles show potential polyadenylation signals for TM1/TM2-CEA and for TM3-CEA, respectively; the up arrow after nucleotide 1,662 indicates the presence of a short oligo A stretch detected in the TM3-CEA cDNA.



also generates TM1-CEA. The predicted apoprotein size of TM2-CEA is 47 kD.

DNA Sequence of TM3-CEA. Northern analysis of poly A+ RNAs indicated a prominent 2.2-kb RNA species when hybridized with LV7 probe DNA (Fig. 1). cDNA from HT-29 poly A+ RNA yielded two partial clones that could be distinguished from TM1- and TM2-CEA cDNAs by restriction site analysis. Efforts to isolate full-length versions of TM3-CEA cDNA from the 23.4 11+ library were not suc-

Table 1	I. Amin	o Acid	Sequence	Identities
among	CEA F	amily.	Members	

	Regions compared (in percent)								
compared	Leader	NH ₂ -terminal	Loop-domain	HR					
NCA/CEA	73	89	85 (CEA I)	76					
			77 (CEA II)						
			76 (CEA III)						
CEA/TM1-CEA	76	87	83 (CEA I)	67					
			79 (CEA II)						
			73 (CEA III)						
NCA/TM1-CEA	76	89	86	62					
TM1-CEA:									
domain I vs. IIa	-	-	44	_					

For each comparison, the different regions correspond to signal sequences (amino acids -34--1), NH₂-terminal segments (amino acids 1-108), loop-domains (amino acids 109-286, 287-464, and 465-642 for CEA, and 109-286 for NCA [Barnett et al., 1988], 109-285 and 286-389 for loop-domains I and IIa, respectively, of TM-CEA), and hydrophobic regions (*HR*) (amino acids 643-668 for CEA, 287-310 for NCA, and 390-421 for TM-CEA). Sequence comparisons are strictly on the basis of amino acid identity; conservative substitutions have not been permitted. Where one region is slightly different in length from its counterpart, percent identity is calculated on the shorter sequence.

Figure 3. Schematic representation of proposed structural domains of CEA-isoantigens TM1-CEA, TM2-CEA, TM3-CEA, TM4-CEA, and their comparison with CEA (180 kD) and NCA. Thin lines correspond to the 5' and 3' untranslated regions. The coding regions of the apoproteins (rectangle) initiate with ATG and are divided into leader sequences (\blacksquare), NH₂-terminal sequences (\blacksquare), loop-domains (\square), hydrophobic regions (\blacksquare), and a cytoplasmic region for TM-CEA (\blacksquare). TAG and TAA are translation stop codons. Lines connecting structural domains indicate where the TM-CEA apoproteins differ from TM1-CEA. The symbol | indicates the segment resulting from the translational frame shift in TM3- and TM4-CEA (see text for details). A_n is poly A.

cessful. The DNA sequence of these partial cDNAs corresponds exactly to nucleotides 345-1,448 and 1,502-1,662 of TMI-CEA cDNA. The deletion of 53 bp between nucleotides 1448 and 1502 occurs in the DNA coding for the cytoplasmic region (shaded box of Fig. 2). As 53 bp is not a multiple of a codon triplet, nucleotides after 1,448 yield a new translation reading frame and generate a new cytoplasmic component for TM3-CEA (schematically drawn in Fig. 3). The sequence indicates a translation of six new amino acids, followed by a TGA stop codon (nucleotide 1518). This results in a 9-amino acid cytoplasmic region for TM3-CEA compared to a 71-amino acid region for TM1- and TM2-CEA producing an apoprotein of 51 kD. A potential polyadenylation consensus sequence AATAAT (Wickens and Stephenson, 1984) in the 3' untranslated region (UTR) of TM3-CEA mRNA is found at a position that precedes a short run of oligo A nucleotides where the sequence diverges from that of TM1- and TM2-CEA. The likely polyadenylation site occurs in sequence that codes for the TMI- and TM2-CEA cytoplasmic domain.

DNA Sequence of TM4-CEA. Although we have yet to obtain a representative cloned cDNA sequence, we speculate that there exists an additional TM-CEA molecule, TM4-CEA (calculated M_r of 40 kD), in which both LD IIa and the cytoplasmic region are alternatively spliced as in TM2-CEA and TM3-CEA, respectively (illustrated in Fig. 3). Experimental support for a fourth TM-CEA molecule comes from our observation of a 1.8 kb mRNA that is observed in Northern blots of 23.4 11+ cell mRNA (Fig. 1, lane *a*). The estimated size of TM4-CEA mRNA without a poly A⁺ tail is 1.4 kb. As our size estimates for each of the TM-CEA RNAs by Northern blot analysis is consistently 300-400 bases longer than that of the cDNA, it is likely that the 1.8-kb RNA codes for TM4-CEA.



Figure 4. Tumor cell distribution of CEA-related poly A+ RNA. Two micrograms of poly A+ RNAs from various tumor cell lines were subjected to electrophoresis on a 2.2 M formaldehyde-0.8% agarose gel, transferred to nylon membrane, and hybridized with ³²P-labeled CYTO probe DNA ([A] lane a, SW403; lane b, LoVo; lane c, DLD-1; lane d, HT-29; lane e, KG-1; lane f, BeWo; lane g, SCaBER; and lane h, MIA PaCa-2) or LV7 DNA ([B]; for lanes c, d, f, g, and h; CYTO and LV7 patterns are identical). Autoradiographic exposures have been adjusted to accommodate the differences in TM-CEA mRNA concentration displayed by the cell lines SCaBER and MIA PaCa-2.

Expression of TM-CEA cDNAs

To investigate the specific expression of TM-CEA mRNAs in tumor cells of different origins, we probed cellular RNAs with a cDNA probe derived from the cytoplasmic region of the TMI-CEA cDNA (CYTO probe; nucleotides 1,443-1,676). Fig. 4 A shows hybridization of the CYTO probe to poly A + RNAs from a variety of tumor cell lines (lanes a-h). All of the tumor cell lines shown, representing colorectal adenocarcinomas (SW403, lane a; LoVo, lane b; DLD-1, lane c; HT-29, lane d), acute myelogenous leukemia (KG-1, lane e), choriocarcinoma (BeWo, lane f), squamous bladder adenocarcinoma (SCaBER, lane g), and pancreatic adenocarcinoma (MIA PaCa-2, lane h) express the TM-CEA mRNAs. While some of these tumors express other CEA-isoantigen mRNAs based on hybridization to LV7 cDNA (Fig. 4 B, lanes a, b, and e), BeWo, SCaBER, and MIA PaCa-2 express only the transmembrane CEA isoantigen mRNAs (data not shown).



We have reported the transfection of the TM-CEA transcription unit into primary and secondary L-cell transfectants (Kamarck et al., 1987). To clone genomic segments corresponding to this transcription unit, a λJam recombinant DNA library of 2×10^6 independently derived recombinants were obtained from an Mbo I partial digest of transfectant cell 23.4 11+ DNA. 96 phage that hybridized with radiolabeled human DNA were plaque purified and a number of these phage isolates ($\lambda 11$, $\lambda 22$, and $\lambda 39$) also hybridized with the LV7 probe. DNA inserts of these three phage ranged in size from 12 to 20 kb and contained the 5.3-kb Bam HI restriction segment shared by all of the primary and secondary L-cell transfectants (Kamarck et al., 1987).

Defined restriction fragments of TMI-CEA cDNA were subcloned for analysis of the genomic DNA. A 5' untranslated region probe (cDNA nucleotides 0-72) hybridized to genomic phage $\lambda 11$, $\lambda 22$, and $\lambda 39$. A 3' UTR probe (3'UTR3;



Figure 5. Map of TM-CEA genomic DNA. (A) Extent of human genomic DNA contained in each λ phage selected from the transfectant cell library. Dots indicate that the phage insert extends beyond that shown. (B) Composite restriction map contained in the five λ phage shown above. Symbols are: R, Eco RI; B, Bam HI; Bg, Bgl II; H, Hind III; and X, Xho I. (C) Restriction segments of genomic DNA homologous to subcloned portions of cDNA representing structural domains in the apoprotein. These segments are described in the text.

cDNA nucleotides 2,917-3,461) did not hybridize to these phage, but did identify two additional human recombinant phage (λ 47 and λ 55) derived from the transfectant genomic library. Restriction site mapping demonstrated that all of the identified phage overlapped (Fig. 5, *A* and *B*). Localization of the 5' and 3' UTR to genomic restriction segments indicated that the cDNA coding sequence is contained within ~23 kb of genomic DNA (Fig. 5, *B* and *C*).

cDNA restriction segments defining structural domains within TM-CEA apoprotein were used to determine their location within genomic DNA (Fig. 5 C). For example, cDNA sequences coding for the LD IIa structural domain (cDNA nucleotides 1017-1343; Fig. 2) were localized to a 3.8-kb Hind III genomic segment (Fig. 5 C). In a similar manner, coding sequences for the leader and NH₂-terminal domains (L-NH₂), the first disulfide-loop domain (LD I), the transmembrane domain (TM), and cytoplasmic domains (CYT1 and CYT2) were localized to genomic restriction segments (Fig. 5 C).

To look for other genomic sequences related to the TM-CEA cytoplasmic domain, the CYTO probe (nucleotides 1,443–1,676) was hybridized to a Southern blot of human genomic DNA cleaved with several restriction endonucleases. In contrast to the complex pattern obtained at high stringency with the domain probe LV7 (Kamarck et al., 1987), hybridization performed even at reduced stringency (60°C; $6 \times$ SSPE) produced one band per lane indicating that there is a single cytoplasmic region in the genome (data not shown).

Delineation of Exon-Intron Boundaries

To support the evidence for alternative RNA splicing, we sequenced selected exon boundaries of the TM-CEA genomic DNA. For example, the 3' end of LD I was sequenced from a subclone of the 2.5-kb Bg1 II genomic segment (Fig. 5 B) that hybridizes with the loop-domain probe, LV7. Doublestranded DNA sequencing using the dideoxy chain termination method was performed after extending an oligonucleotide primer from the 3' end of LD I, in this example, into adjacent intron sequence. A summary of selected exon-intron junction sequences is presented in Table II. Our data demonstrate that the structural domains defined as LD IIa, transmembrane, and cytoplasmic regions are coded for by discrete exons flanked by splice donor-acceptor sites. In addition, the nucleotide sequences of the exon-intron boundaries confirm the proposed location of junctions for alternative splicing based on the cDNA sequences of TM1- through TM4-CEA.

Biochemical Analysis of TM-CEA Isoantigens

Western immunoblot analyses, using a broadly reactive anti-CEA polyclonal antibody on lysates of 23.4 11+ cells, showed a complex array of antigens with molecular masses ranging from 50 to 150 kD (Fig. 6, lane A). This result is consistent with the existence of multiple TM-CEA isoantigens displaying extensive N-linked glycosylation. To assess the number of different TM-CEA apoprotein species actually present, we subjected 23.4 11+ polypeptides to enzymatic deglycosylation with peptide/N-glycanase isolated from F. meningosepticum (Tarentino et al., 1985). As shown in Fig. 6, lane B, western analysis of deglycosylated transfectant cell proteins demonstrates the presence of four to five CEA isoantigen apoproteins. In contrast a lysate of control L cells shows no reactive species (Fig. 6, lane C). The mobilities of the major bands in lane B are consistent with the molecular sizes of apoproteins calculated from translation of cDNA sequences: TM1, 58 kD; TM2, 47 kD; TM3, 51 kD; TM4, 40 kD. The presence of an apparent doublet for the smallest species is presently not explained, but may be due to inaccessibility of some N-linked oligosaccharides to enzyme.

Discussion

In the past 20 yr, an array of 8-10 CEA-related glycoproteins has been identified by biochemical methods in a variety of tumor and nontumor cells (for review see Shively and Beatty, 1985). Recently, the genes for two members of this isoantigen family, CEA and NCA, have been cloned (Beauchemin et al., 1987; Oikawa et al., 1987; Barnett et al., 1988; Neumaier et al., 1988; Tawaragi et al., 1988). In this study, we identify mRNAs, cDNAs, and a genomic transcription unit that encode novel CEA isoantigen members with transmembrane and cytoplasmic domains (TM-CEAs). Immunological data confirm that TM-CEAs are expressed at the cell surface. Although we cannot assign TM-CEA glycoproteins to previously identified CEA isoantigens (Shively and Beatty, 1985), on the basis of molecular size we suggest that TMI-CEA may be related to the so-called "128K antigen" (Neumaier et al., 1985).

We have shown that an L-cell secondary transfectant, 23.4 11+, and a number of tumor cell lines express a set of four RNA transcripts (3.9, 3.7, 2.2, and 1.8 kb) that are unrelated to CEA (4.0 and 3.6 kb) or NCA (3.0 kb) poly A+ RNAs. As the secondary transfectants likely contain a single genomic transcription unit, we proposed that these mRNAs represented the products of alternative RNA splicing (Kamarck

Table II. DNA Sequence at Exon-Intron Junctions of TM-CEA Genomic DNA

Junction	Donor	Acceptor
LD I-LD IIa	GTC ACT G gtaag	gacag AG CTA AGT
	1030	1031
LD IIa-TM	GTA AAC T gtaag	gacag AT AAT GCT
	1318	1319
TM-CYT1	ACC GGC AG gtatg	tttag G GCA AGC
	1448	1449
CYT1-CYT2	AAC CAC A gtaag	ag CT CAG GAC
	1501	1502

Appropriate genomic restriction segments were subjected to DNA sequencing near the regions that are deleted or altered by splicing in TM2- and TM3-CEA cDNAs. Capitalized letters represent sequences contained in cDNA and genomic DNA, and adjacent small type represents intron sequence. Numbers refer to their positions in TM1-CEA cDNA. Abbreviations describing the exons are as defined in the text.



Figure 6. TM-CEA protein isoantigens expressed by 23.4 11+ transfectants. After transfer from gels, CEA isoantigens were detected using DAKO anti-CEA Ig followed by goat anti-rabbit alkaline phosphatase. (Lane A) Lysate from 23.4 11+ cells; (lane B) enzymatically deglycosylated lysate from 23.4 11+ cells; and (lane C) lysate from mouse L cells. Numbers on the left indicate the apparent molecular masses in kilodaltons of prestained molecular mass markers phosphorylase B (106), BSA (68), ovalbumin (45), and α -chymotrypsin (23).

et al., 1987). The cDNA sequences of the TM-CEAs reported here, coupled with genomic DNA sequences at exonintron junctions, demonstrate that the mRNAs are indeed derived by alternative splicing from a single transcript. While TM1-CEA represents the largest transmembrane molecule, TM2-CEA lacks LD IIa, and TM3-CEA is defined by an alternate splice within the cytoplasmic domain, resulting in a unique intracellular peptide sequence. The existence of an mRNA that contains both alternate splices, TM4-CEA, is supported by expression in the transfectants of a 1.8-kb poly A + RNA (Fig. 1, lane a) and a 40-kD apoprotein (Fig. 6).

We also have cloned, as overlapping recombinant phage, the complete TM-CEA transcription unit which is encoded in ~ 23 kb of genomic DNA. Hybridization with cDNA probes indicates that structural domains defined by comparison with other CEA-related genes correspond to discrete exons (Fig. 5). This was confirmed by the demonstration that regions adjacent to predicted exons are not found in cDNA and have consensus splice donor-acceptor sequences (Table II). These exon-intron junctions correspond exactly to the points of difference detected in the sequences of the alternatively spliced TM-CEAs.

Cell lines derived from colorectal tumors primarily produce CEA poly A+ RNA (Barnett et al., 1988), consistent with their secretion of 180-kD CEA. Using specific cDNA probes we have demonstrated that TM-CEAs are also expressed by a number of these colorectal lines, in addition to other tumor types. In some cell lines, only the set of TM-CEA mRNAs are expressed (e.g., BeWo; Fig. 4, lane f), while in others, numerous CEA-related RNAs are additionally produced (e.g., HT-29; Fig. 1, lane b). The expression of TM-CEAs is evidently not restricted to neoplastic tissue as TM1-CEA cDNAs have been cloned from an apparently normal fetal liver cDNA library. We cannot as yet correlate the expression of one or more CEA isoantigens with biological function, cellular morphology, substrate adhesion, or metastatic potential.

Members of the CEA isoantigen family display considerable nucleotide and amino acid sequence conservation (Table I). With the exception of the "half" disulfide loop-domain of TM1-CEA, all of the proposed extracellular regions are \geq 70% similar at the amino acid level and \geq 80% similar at the nucleotide level. This explains the ability of the TM-CEAs to be detected by several anti-CEA monoclonal antibodies and also illustrates the general cross-reactivity of anti-CEA immunoreagents.

Extracellular loop domains in CEA isoantigens demonstrate conserved amino acids surrounding the cysteine residues that are reminiscent of immunoglobulins and immunoglobulin superfamily proteins (Hunkapiller and Hood, 1986). We have noted previously (Barnett et al., 1988), and also observe here, the overall structural and amino acid similarity of all CEA glycoproteins to members of the immunoglobulin gene superfamily (Williams and Barclay, 1988). This family extends to cell surface recognition proteins (Thy-1 [Williams and Gagnon, 1982]), cell adhesion proteins (neural cell adhesion molecule [N-CAM; Cunningham et al., 1987], ICAM [Simmons et al., 1988], T cell erythrocyte receptor CD2 [Peterson and Seed, 1987], and S- and L-MAG [Salzer et al., 1987; Arquint et al., 1987]), receptor proteins (T4 receptor [Maddon et al., 1985] and PDGF receptor [Yarden et al., 1986]), and oncogenes (CSF-1 receptor/v-fms [Hampe et al., 1984]). Although their ligands have not yet been identified, we propose by analogy that members of the CEA isoantigen family act as cell surface recognition proteins.

This analogy can be extended to the alternative structural forms of neural cell adhesion molecule (N-CAM) and myelinassociated glycoprotein (MAG). Three different N-CAM isoantigens are generated from a single genomic segment (Cunningham et al., 1987): the ssd form, with a short hydrophobic COOH-terminus (like CEA and NCA), the sd form, with a short cytoplasmic domain (like TM3/4-CEA), and the ld form, with a longer cytoplasmic domain (TM1/2-CEA). Two cytoplasmic forms of MAG (S-MAG and L-MAG) have also been described (Salzer et al., 1987). For MAG, N-CAM, and TM-CEA, the alternative cytoplasmic regions are derived by differential splicing of those exons that code for intracellular peptide segments. A novel N-CAM has been described recently in human skeletal muscle where the ssd form contains a previously unidentified 37-amino acid extracellular domain (Dickson et al., 1987). This may compare with the TMI- and TM3-specific "half" disulfide-loop in these forms of CEA.

Alternative forms of the cytoplasmic region of N-CAM and L-MAG are derived by appropriate splice selection of inframe exons (Cunningham et al., 1987; Salzer et al., 1987). In contrast, TM3-CEA and the proposed TM4-CEA contain a short but novel cytoplasmic peptide generated by a splice that results in out-of-frame translation. With this splice, the coding sequence for TM1/2-CEA cytoplasmic region becomes the 3' UTR and also contains the consensus polyadenylation sequence (Figs. 2 and 3). To our knowledge, this result is unique among alternatively spliced RNAs. Unusual alternate COOH-terminal splices accompany the synthesis of the glial and endothelial cell forms of platelet-derived growth factor A chain (Tong et al., 1987; Bonthron et al., 1987) and of the hydrophobic and hydrophilic termini of decay accelerating factor (Caras et al., 1987), but in each case, the 3' untranslated sequence and polyadenylation site remains constant.

The cytoplasmic domain of TM-CEA is unique, as it is represented as a single band in Southern blot genomic analysis and bears no significant resemblance to other CEA- or non-CEA-related sequences based on computer-aided searches EPPELDLSYSHSDKGKRPTKDSYTLTEELAEYAEIRVK *!! ***i! !!! *!! *!!** DPPMNEVTYSEAQQPTQPTSASPSLTATEIIYSEVKKQ TM1/2-CEA NK TLNF

Figure 7. Amino acid similarity in the cytoplasmic region of immunoglobulin gene superfamily members, L-MAG and TM-CEA. The carboxyl-terminal residues of L-MAG and TM1-/TM2-CEA are shown; the absolute COOH-terminus for each is on the right. Regions of identity or similarity are indicated, respectively, by a solid line (1) or an asterisk (*) extending between compared residues. For best physical alignment, deletion of some amino acids in TM-CEA relative to L-MAG is required; this is indicated by \hat{A} .

of the GenBank or EMBL data bases, or by direct comparison of 20 different immunoglobulin superfamily members. However, a segment of limited similarity has been detected between the COOH-terminal 44 amino acids of the TMI/ TM2-CEA cytoplasmic domain and the COOH-terminal 38 amino acids of the L-MAG cytoplasmic domain (19/38 amino acids; Fig. 7). This cytoplasmic segment is relatively high in amino acids that are potential substrates for phosphorylation i.e., serine, threonine, and tyrosine (13/48 residues), and it is largely these amino acids that are shared between the molecules. Consensus signals for associated enzymatic activities like tyrosine kinase (Bairoch and Claviere, 1988) are not found.

In Fig. 8, we illustrate our interpretation of the extensive structural similarity among CEA isoantigens using immunoglobulin molecules as a model for their organization (Hunkapiller and Hood, 1986). CEA isoantigens are depicted



Figure 8. Hypothetical single chain and dimer forms of CEA isoantigens as immunoglobulin-like molecules. Each CEA family member is depicted with extracellular disulfide loops and is embedded in the lipid bilayer via phosphoinositol linkage (for CEA and NCA) or is an integral cell surface molecule (for TM-CEAs). Some TM-CEA members are depicted as homo- or heterodimers. as membrane bound: for CEA and NCA, the membrane attachment may occur via a phosphoinositol glycan moiety (Takami et al., 1988), while the TM-CEAs have structural domains that make them integral membrane glycoproteins. While we depict all CEA isoantigens as single polypeptide chains, sequence analysis of the TM-CEAs indicates that these proteins have the potential to exist in vivo as cell surface homo- or heterodimers. LD I of TM-CEAs contains a single cysteine residue (amino acid position 308; Fig. 2) that may not be involved in intrachain disulfide linkage and therefore by analogy, TM-CEA homodimers and heterodimers are depicted as immunoglobulin-like in surface character.

We wish to thank Mary Ann Nothdurft, Judy Dziuba, and Karen Wallberg for excellent technical assistance; Drs. Frank Ruddle and Vincent Marchesi for support; and gratefully acknowledge the many astute and helpful suggestions of Peter M. M. Rae. Quality graphics were by Suzy Pafka.

Received for publication 11 August 1988, and in revised form 10 October 1988.

Note Added in Proof: Hinoda et al. (1988. Proc. Natl. Acad. Sci. USA. 84:5350-5354) recently described the sequence of biliary glycoprotein I (BGP-I) which appears virtually identical to TM1-CEA. There are several significant differences between BGP-I mRNA and TM1-CEA mRNA described here. These include a shortened BGP-I cytoplasmic coding region which is attributable to a single nucleotide deletion at their position 1,401, a completely unrelated 3' UTR when compared to TM1-CEA, and drastically different Northern blot patterns when using unique cDNA segments.

References

- Arquint, M., J. Roder, L.-S. Chia, J. Down, D. Wilkinson, H. Bayley, P. Braun, and R. Dunn. 1987. Molecular cloning and primary structure of myelin-associated glycoprotein. Proc. Natl. Acad. Sci. USA. 84:600-604.
- Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid cellulose. Proc. Natl. Acad. Sci. USA. 69:1408-1412.
- Bairoch, A., and J.-M Claviere. 1988. Sequence patterns in protein kinases. Nature (Lond.). 331:22.
- Barnett, T., S. Goebel, M. A. Nothdurft, and J. Elting. 1988. Carcinoembryonic antigen family: characterization of cDNAs coding for NCA and CEA, and suggestion of non-random sequence variation in their conserved loop domains. *Genomics*. 3:59-66.
- Beauchemin, N., S. Benchimol, D. Cournoyer, A. Fuks, and C. Stanners. 1987. Isolation and characterization of full-length functional cDNA clones for human carcinoembryonic antigen. *Mol. Cell. Biol.* 7:3221–3230.
- Benton, W., and R.W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science (Wash. DC). 196:180-182.
- Bonthron, D., C. Morton, S. Orkin, and T. Collins. 1988. Platelet-derived growth factor A chain: gene structure, chromosomal location, and basis for alternative mRNA splicing. Proc. Natl. Acad. Sci. USA. 85:1492-1496.
- Caras, I., M. Davitz, L. Rhee, G. Weddell, D. Martin, and V. Nussenzweig. 1987. Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. *Nature (Lond.)*. 325:545-549.
- Cunningham, B., J. Hemperley, B. Murray, E. Prediger, R. Brackenbury, and G. Edelman. 1987. Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. *Science (Wash. DC)*. 236:799-806.
- Darcy, D., C. Turberville, and R. James. 1973. Immunological study of carcinoembryonic antigen (CEA) and a related glycoprotein. Br. J. Cancer. 28:147-154.
- Dickson, G., H. Gower, C. Barton, H. Prentice, V. Elsom, S. Moore, R. Cox, C. Quinn, W. Putt, and F. Walsh. 1987. Human muscle neural cell adhesion molecule (N-CAM): identification of a muscle-specific sequence in the extracellular domain. *Cell*. 50:1119-1130.
- DiLella, A., and S. Woo. 1987. Cloning large segments of genomic DNA using cosmid vectors. *Methods Enzymol.* 152:199-212.
- Engvall, E., J. Shively, and M. Wrann. 1978. Isolation and characterization of the normal crossreacting antigen: homology of its NH₂-terminal amino acid sequence with that of carcinoembryonic antigen. *Proc. Natl. Acad. Sci.* USA. 75:1670-1674.
- Feinberg, A., and B. Vogelstein. 1984. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 137:266-267.
- Gold, P., and S. Freedman. 1965. Specific carcinoembryonic antigens of the human digestive system. J. Exp. Med. 122:467-481.
- Gubler, U., and B. Hoffman. 1983. A simple and efficient method for generat-

ing cDNA libraries. Gene (Amst.). 25:263-269.

- Hampe, A., M. Gobet, C. Sherr, and F. Galibert. 1984. Nucleotide sequence of the feline retroviral oncogene v-fms shows unexpected homology with oncogenes encoding tyrosine-specific protein kinases. Proc. Natl. Acad. Sci. USA. 81:85-89.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene (Amst.). 28:351-359.
- Hunkapiller, T., and L. Hood. 1986. The growing immunoglobulin gene superfamily. Nature (Lond.). 323:15-16.
- Kamarck, M., J. Elting, J. Hart, S. Goebel, P. M. M. Rae, M. A. Nothdurft, J. Nedwin, and T. Barnett. 1987. Carcinoembryonic antigen family: expression in a mouse L-cell transfectant and characterization of a partial cDNA in bacteriophage \gt11. Proc. Natl. Acad. Sci. USA. 84:5350-5354.
- Laemmli, U. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (Lond.). 227:680-685.
- Lehrach, H., D. Diamond, J. Wozney, and H. Boedtker. 1977. RNA molecular weight determinations by electrophoresis under denaturing conditions, a crit-ical reexamination. *Biochemistry*. 16:4743–4751.
- Maddon, P., D. Littman, M. Godfrey, D. Maddon, L. Chess, and R. Axel. 1985. The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family. Cell 42:93-104
- Maniatis, T., E. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 368-369
- Mullins, J., D. Brody, R. Binari, and S. Cotter. 1984. Viral transduction of c-myc gene in naturally occurring feline leukaemias. Nature (Lond.). 308: 856-859
- Neumaier, M., U. Fenger, and C. Wagener. 1985. Monoclonal antibodies for carcinoembryonic antigen (CEA) as a model system: identification of two novel CEA-related antigens in meconium and colorectal carcinoma tissue by Western blots and differential immunoaffinity chromatography. J. Immunol. 135:3604-3609.
- Neumaier, M., W. Zimmermann, L. Shively, Y. Hinoda, A. Riggs, and J. Shively. 1988. Characterization of a cDNA clone for the nonspecific crossreacting antigen (NCA) and a comparison of NCA and carcinoembryonic antigen. J. Biol. Chem. 263:3202-3207.
- Oikawa, S., H. Nakazoto, and G. Kosaki. 1987. Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequence. *Biochem.* Biophys. Res. Commun. 142:511-513. Paxton, R., G. Moser, H. Pande, T. Lee, and J. Shively. 1987. Sequence analy-
- sis of carcinoembryonic antigen: identification of glycosylation sites and homology with immunoglobulin supergene family. Proc. Natl. Acad. Sci. USA. 84:920-924.
- Peterson, A., and B. Seed. 1987. Monoclonal antibody and ligand binding sites of the T cell erythrocyte receptor (CD2). Nature (Lond.). 329:842-846.
- Salzer, J., W. P. Holmes, and D. R. Colman. 1987. The amino acid sequences

of the myelin-associated glycoproteins: homology to the immunoglobulin gene superfamily. J. Cell Biol. 104:957-965.

- Sanger, F., S. Nicklen, and A. Coulson. 1977. DNA sequencing with chainterminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467. Shively, J., and J. Beatty. 1985. CEA-related antigens: molecular biology and
- clinical significance. CRC Crit. Rev. Oncol. Hematol. 2:355-399.
- Simmons, D., M. Makgoba, and B. Seed. 1988. ICAM, an adhesion ligand of LFA-1, is homologous to the neural cell adhesion molecule NCAM. Nature (Lond.). 331:624-627.
- Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.
- Takami, N., Y. Misumi, M. Kuroki, Y. Matsuoka, and Y. Ikehara. 1988. Evidence for carboyl-terminal processing and glycolipid-anchoring of human carcinoembryonic antigen. J. Biol. Chem. 263:12716-12720.
- Tarentino, A., C. Gomez, and T. Plummer. 1985. Deglycosylation of asparagine-linked glycans by peptide:N-glycanase F. Biochemistry. 24: 4665-4671
- Tawargi, Y., S. Oikawa, Y. Matsuoka, G. Kosaki, and H. Nakazoto. 1988. Primary structure of nonspecific crossreacting antigen (NCA), a member of the carcinoembryonic antigen (CEA) gene family, deduced from cDNA sequence. Biochem. Biophys. Res. Commun. 150:89-96.
- Thompson, J., H. Pande, R. Paxton, L. Shively, A. Padma, R. Simmer, C. Todd, A. Riggs, and J. Shively. 1987. Molecular cloning of a gene belonging to the carcinoembryonic antigen gene family and discussion of a domain model. Proc. Natl. Acad. Sci. USA. 84:2965-2969. Tong, B., D. Auer, M. Jaye, J. Kaplow, G. Ricca, E. McConathy, W. Drohan,
- and T. Deuel. 1987. cDNA clones reveal differences between human glial and endothelial cell platelet-derived growth factor A-chains. Nature (Lond.). 328:619-621
- Towbin, H., T. Stahelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA. 76:4350-4354
- von Kleist, S., G. Chavanel, and P. Burtin. 1972. Identification of an antigen from normal human tissue that crossreacts with the carcinoembryonic antigen. Proc. Natl. Acad. Sci. USA. 69:2492-2496.
- Wickens, M., and P. Stephenson. 1984. Role of conserved AAUAAA sequence: four AAUAAA point mutations prevent messenger RNA 3' end formation. Science (Wash. DC). 226:1045-1051
- Williams, A., and J. Gagnon. 1982. Neuronal cell Thy-1 glycoprotein: homology with immunoglobulin. Science (Wash. DC) 216:696-703
- Williams, A., and A. Barclay. 1988. The immunoglobulin superfamily-domains for cell surface recognition. Annu. Rev. Immunol. 6:381-405.
- Yarden, Y., J. Escobedo, W.-J. Kuang, T. Yang-Feng, T. Daniel, P. Tremble, E. Chen, M. Ando, R. Harkins, U. Francke, V. Fried, A. Ullrich, and L. Williams. 1986. Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. Nature (Lond.) 323:226-232.