



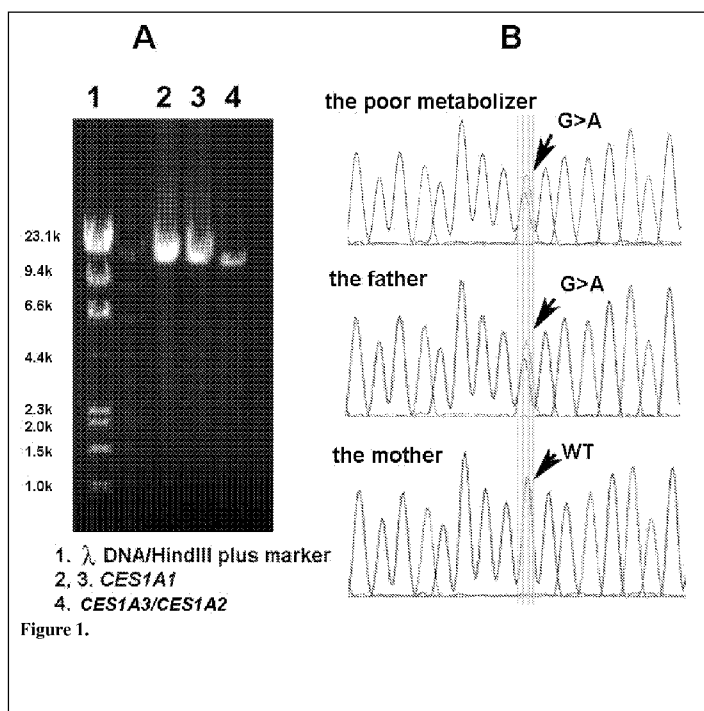
- (51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 9/16* (2006.01)
C12N 15/55 (2006.01)
- (21) International Application Number:
PCT/US2012/034255
- (22) International Filing Date:
19 April 2012 (19.04.2012)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/477,475 20 April 2011 (20.04.2011) US
61/491,668 31 May 2011 (31.05.2011) US
- (71) Applicant (for all designated States except US): **UNIVERSITY OF FLORIDA RESEARCH FOUNDATION INC.** [US/US]; P.O. Box 115575, Gainesville, FL 32611-5575 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **MARKOWITZ, John, Seth** [US/US]; 1600 SW Archer Road, Room PG-23, Gainesville, FL 32610 (US). **ZHU, Haojie** [US/US];

1600 SW Archer Road, Room PG-06, Gainesville, FL 32610 (US).

- (74) Agents: **CORLESS, Peter, F.** et al.; Edwards Wildman Palmer LLP, P.O. BOX 55874, Boston, MA 02205 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on next page]

- (54) Title: COMPOSITIONS AND METHODS FOR GENOTYPING CES1 GENETIC VARIANTS AND USE THEREOF



(57) Abstract: The invention features compositions and methods that are useful for genotyping CES1 isoforms (*CES1A1*, *CES1A2*, *CES1A3*). The ability to specifically genotype one or more CES1 isoforms (e.g. *CES1A1*) is useful for assessing drug metabolism in a subject and guiding treatment selection.



SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG). **Published:**

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

COMPOSITIONS AND METHODS FOR GENOTYPING CES1 GENETIC VARIANTS AND USE THEREOF

CROSS-REFERENCE TO RELATED APPLICATION

5 This application claims the benefit of the following U.S. Provisional Application No.: 61/477,475, filed April 20, 2011, and 61/491,668, filed May 31, 2012, the entire contents of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

10 Mammalian carboxylesterases (CESs) have prominent roles in the hydrolysis of numerous and diverse compounds including carboxylic acid esters, carbamates, thioesters, and amide containing agents. These substrates are represented in almost every drug class, as well as in many prodrugs and agents purposely formulated as esters for the purpose of improving oral bioavailability of the active moiety. Human carboxylesterase 1 (hCES1) is
15 the major carboxylesterase expressed in the liver, and contributes approximately 80% of total hepatic hydrolytic activity and its functionality is important for the inactivation of numerous xenobiotics including therapeutic agents, such as methylphenidate (MPH; Ritalin[®]) which serve as substrates. Further, functional hCES1 is important for the bioactivation of a number of specific prodrugs, such as oseltamivir, andtrandolapril. Finally, hCES1 is recognized as
20 the hepatic enzyme catalyzing transesterification reactions between xenobiotics and some endogenous compounds with orally ingested ethanol.

 hCES1 is encoded in humans by the *CES1* gene. The expression of *CES1* and catalytic activity of hCES1 exhibit substantial interindividual variability. Three isoforms of the *CES1* gene have been identified, i.e. *CES1A1*, *CES1A2*, and *CES1A3*. Although the DNA
25 sequences of the three genes have high similarity, *CES1A1* and *CES1A2* are functional whereas *CES1A3* is not. Genetic variation is one of the major contributing factors of varied hCES1 function in humans.

 In the liver, the majority of hCES1 is the product of the *CES1A1* gene because transcription of the *CES1A2* gene is substantially lower than that of the *CES1A1*. Thus,
30 *CES1A1* variants influence CES1 enzyme activity more than *CES1A2* variants. However, current Taqman[®]-based methods are incapable of distinguishing variants in *CES1* genes due to the high similarity of *CES1* DNA sequences. For example, homozygous variants of either *CES1A1* or *CES1A2* are often mistakenly read as heterozygous mutations in a Taqman[®] *CES1* assay.

The ability to genotype *CES1* variants has the potential to improve a variety of pharmacotherapeutic treatments directed at curing or ameliorating symptoms of selected diseases and disorders. Despite rapid progress in methods for detecting *CES1* genetic variants, current methods are unable to distinguish between *CES1* variants encoded by *CES1A1*, *CES1A2*, and *CES1A3* and/or to distinguish *CES1* homozygotes from heterozygotes, (e.g., between the presence of two or one *CES1* variant alleles). Thus, there is an urgent need to develop a discriminative *CES1A1* genotyping assay in order to better guide pharmacogenetic-based therapy.

SUMMARY OF THE INVENTION

As described below, the present invention features compositions and methods for genotyping *CES1* genetic variants, which is useful for predicting whether a subject will respond to a drug treatment, identifying a subject as responsive to drug treatment, identifying the cause of exaggerated pharmacodynamics responses or overt toxicity related to specific drugs, or selecting an appropriate therapy for a subject. The present invention further provides compositions and methods to personalize a therapy and/or avoid adverse consequences of altered metabolism of a therapeutic agent or compound (e.g., enalapril, methylphenidate), associated with the presence of a *CES1A1* variant allele linked to reduced drug metabolism in a subject.

In one aspect, the invention provides a method of genotyping a subject for *CES1*, the method involving amplifying a *CES1* nucleic acid in a biological sample from the subject by long range polymerase chain reaction (PCR); and detecting a variant allele of *CES1*, thereby genotyping the subject for *CES1*.

In another aspect, the invention provides a method of genotyping a subject for *CES1*, the method comprising amplifying a *CES1* nucleic acid in a biological sample from the subject by long range polymerase chain reaction (PCR); and detecting a variant allele of *CES1*, thereby genotyping the subject for *CES1*, wherein the alteration is in a residue that stabilizes substrate-enzyme intermediates or that alters enzyme activity; and wherein analyzing comprises nucleic acid sequencing, allele-specific hybridization, allele-specific PCR, oligonucleotide microarray analysis, or mass spectrometry.

In another aspect, the invention provides a method of selecting an appropriate therapy for a subject, the method involving genotyping *CES1* in a subject, where a genotype having a variant allele having reduced carboxylesterase activity or expression relative to a reference

indicates that drug therapy is not appropriate for the subject, and where a genotype that is homozygous wild-type indicates that drug therapy is appropriate for the subject.

In yet another aspect, the invention provides a method of identifying a subject as responsive to drug therapy, the method involving genotyping *CES1* in a subject, where a
5 genotype having a variant allele having reduced carboxylesterase activity or expression relative to a reference identifies the subject as not responsive to drug therapy, and where a genotype that is homozygous wild-type identifies the subject as responsive to drug therapy.

In another aspect, the invention provides a method of monitoring drug therapy in a subject, the method involving genotyping *CES1* in a subject, identifying a subject as having a
10 variant allele of *CES1*, and measuring plasma levels of the drug in the subject to determine an appropriate dosage.

In still another aspect, the invention provides a method of characterizing subject sensitivity to a drug, the method comprising genotyping *CES1* in a subject, where a genotype having a variant allele having reduced carboxylesterase activity or expression relative to a
15 reference identifies the subject as sensitive to the drug, and where a genotype that is homozygous wild-type identifies the subject as not sensitive to the drug.

In an additional aspect, the invention provides a method of genotyping a subject, the method involving amplifying a *CES1* nucleic acid in a biological sample from the subject by long range polymerase chain reaction (PCR); and detecting a variant allele of *CES1* relative
20 to a wild-type reference sequence, where the variant allele encodes a *CES1* polypeptide having reduced carboxylesterase activity or expression relative to a reference selected from the group consisting of Gly143Glu and Asp260Glu frameshift.

In a further aspect, the invention provides a kit for genotyping a *CES1* isoform in a subject containing a set of nucleic acid probes, where the nucleic acid probes can selectively
25 bind to and amplify a nucleic acid at a *CES1* isoform locus.

In various embodiments of any of the aspects delineated herein, the *CES1* or *CES1* isoform is one or more of a *CES1A1*, *CES1A2*, or *CES1A3* isoform. In particular embodiments, the *CES1* or *CES1* isoform is a *CES1A1* isoform. In various embodiments of any of the aspects delineated herein, the method distinguishes a heterozygous genotype from
30 a homozygous genotype. In various embodiments of any of the aspects delineated herein, the subject is heterozygous or homozygous for a variant allele having reduced carboxylesterase activity or expression relative to a reference. In various embodiments of any of the aspects delineated herein, analyzing comprises nucleic acid sequencing, allele-specific hybridization, allele-specific PCR, oligonucleotide microarray analysis, or mass spectrometry.

In various embodiments of any of the aspects delineated herein, the variant allele encodes a CES1 polypeptide having reduced or increased carboxylesterase activity or expression relative to a reference. In various embodiments of any of the aspects delineated herein, the method identifies at least one alteration in an evolutionarily conserved residue of CES1. In various embodiments, the conserved residue is in a triad residue, in the active site, or in the oxyanion hole. In particular embodiments, the conserved residue serine 221 (S), glutamic acid 354 (E), histidine 468, or Gly141-143.

In various embodiments, the alteration is in a residue that stabilizes substrate-enzyme intermediates or that alters enzyme activity. In particular embodiments, the variant allele is selected from the group consisting of Gly143Glu and Asp260Glu frameshift. In other embodiments, the variant allele is one or more of an allele set forth in Table 1.

In various embodiments of any of the aspects delineated herein, the variant allele is indicated by the presence of a single nucleotide polymorphism (SNP). In particular embodiments, the SNP is selected from the group consisting of 428G>A and T891del. In other embodiments, the SNP is one or more of a SNP set forth in Table 1.

In various embodiments of any of the aspects delineated herein, the method further comprises assaying the carboxylesterase activity of the protein encoded by the variant allele of CES1. In various embodiments of any of the aspects delineated herein, the method further comprises assaying substrate hydrolysis for the variant allele of CES1. In various embodiments of any of the aspects delineated herein, substrate hydrolysis is determined by contacting a drug with the protein and detecting the half-life ($t_{1/2}$), pharmacokinetics of the drug, or altered pharmacodynamic (PD) effect of the drug.

In various embodiments of any of the aspects delineated herein, the method involves selecting an appropriate drug therapy for the subject. A genotype comprising the presence of a variant allele having reduced carboxylesterase activity or expression relative to a reference identifies the subject as sensitive to the drug. A genotype that is homozygous wild-type identifies the subject as not sensitive to the drug. In various embodiments of any of the aspects delineated herein, the subject is administered an effective amount of the drug. In other embodiments of any of the aspects delineated herein, drug administration to the subject is discontinued.

In various embodiments of any of the aspects delineated herein, the drug is a compound selected from the group consisting of a prodrug, an illicit drug, an opioid, a dopaminergic or noradrenergic drug, an ACE Inhibitor, an HMG-CoA reductase inhibitor, a statin, an anesthetic, a toxin, a chemical warfare agent, an insecticide (e.g., an

organophosphate insecticide), an antiviral drug, and an anti-cancer drug. In various embodiments of any of the aspects delineated herein, the drug is a compound selected from the group consisting of lidocaine, cilazapril, delapril, imidapril, cocaine, enalapril, quinapril, temocapril, methylphenidate, benazepril, trandolapril, lovastatin, oseltamivir, meperidine, prasugrel, simvastatin, valacyclovir, capecitabine, heroin, clopidogrel, sarin, soman, tabun, cholesterol, irinotecan (CPT-11), and mycophenolate.

In various embodiments of any of the aspects delineated herein, the drug is a prodrug that requires *CES1* activity for bioactivation. In specific embodiments, the prodrug is a compound selected from the list consisting of oseltamivir, cilazapril, enalapril, capecitabine, delapril, quinapril, imidapril, temocapril, and lovastatin. In various embodiments of any of the aspects delineated herein, the method further comprises selecting an alternative therapy.

In various embodiments of any of the aspects delineated herein, the subject is identified as having an adverse reaction to a drug that requires *CES1* activity for bioinactivation. In some embodiments, the drug is a compound selected from the list consisting of cocaine, methylphenidate, meperidine, capecitabine, and clopidogrel. In other embodiments, the drug is a compound selected from the list consisting of an insecticide, an organophosphate insecticide, or paraoxon.

In various embodiments of any of the aspects delineated herein, the subject is heterozygous or homozygous for a variant allele having reduced carboxylesterase activity or expression relative to a reference. In various embodiments of any of the aspects delineated herein, genotyping *CES1* in a subject involves amplifying a *CES1* nucleic acid by long range polymerase chain reaction (PCR) and analyzing the sequence of the amplified nucleic acid for a variant allele of *CES1* relative to a wild-type reference sequence. In various embodiments of any of the aspects delineated herein, the method further involves decreasing a dosage of a drug administered to the subject or discontinuing treatment in the subject.

In various embodiments of any of the aspects delineated herein, the kit contains a nucleic acid probe set forth in Table 3. In various embodiments of any of the aspects delineated herein, the kit contains instructions for selecting an appropriate therapy for a subject, identifying a subject as responsive to drug therapy, monitoring drug therapy in a subject, or characterizing subject sensitivity to a drug based on the genotype of a *CES1* isoform of the subject.

In various embodiments of any of the aspects delineated herein, the kit contains a microarray for detecting an SNP in a gene encoding an enzyme involved in drug metabolism

or a drug transporter. In specific embodiments, the enzyme is one or more of cytochrome P450 2D6 and cytochrome P450 2C19.

Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

5

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them below, unless specified otherwise.

10

15

By "alteration" is meant an increase or decrease. An alteration may be by as little as 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, or by 40%, 50%, 60%, or even by as much as 75%, 80%, 90%, or 100%.

By "allele" is meant one of a series of two or more different gene sequences that occupy the same position or locus on a chromosome.

20

By "amplify" is meant to increase the number of copies of a molecule. In one example, the polymerase chain reaction (PCR) is used to amplify nucleic acids.

By "binding" is meant having a physicochemical affinity for a molecule. Binding is measured by any of the methods of the invention, e.g., hybridization of a detectable nucleic acid probe, such as a TaqMan[®] based probe, Pleiades based probe.

25

By "biological sample" is meant any tissue, cell, fluid, or other material derived from an organism (e.g., human subject).

By "bioactivation" is meant the chemical alteration of a compound within a subject to generate a compound or form of a compound having a pharmacological effect. For example, a prodrug undergoes bioactivation to remove protective groups, thereby producing a drug in active form (e.g., increasing the pharmacodynamic (PD) effect of the drug). In the context of the invention, the hCES1 enzyme bioactivates specific prodrugs and/or selected non-prodrugs by carboxylesterase catalytic activity (hydrolysis of a carboxylic acid ester, carbamate, thioester, or amide).

30

By “bioinactivation” is meant the chemical alteration of a compound within a subject to decrease or eliminate the pharmacological effect of the compound or otherwise serve in a detoxifying capacity. For example, bioinactivation results in a decrease in the half-life ($t_{1/2}$) of the drug or a decrease in the pharmacodynamic (PD) effect of the drug. In the context of the invention, the *CES1* enzyme bioinactivates compounds by carboxylesterase activity (hydrolysis of a carboxylic acid ester, carbamate, thioester, or amide).

By “biomarker” is meant a polypeptide or polynucleotide that is differentially present in a sample taken from a subject having a disease or disorder relative to a reference. Exemplary biomarkers include nucleic acid molecules encoding variant alleles.

By “carboxylesterase 1 polypeptide” or “*CES1* polypeptide” is meant a polypeptide or fragment thereof having at least 85% amino acid identity to NCBI Accession No. AAI10339 and having carboxylesterase activity.

By “carboxylesterase 1 nucleic acid molecule” or “*CES1* nucleic acid molecule” is meant a polynucleotide encoding a *CES1* polypeptide. Exemplary *CES1* nucleic acid molecules are provided at NCBI Accession Nos. AB119997 (*CES1A1*); AB119998 (*CES1A2*); and NT_010498 (*CES1A3*).

By “carboxylesterase activity” is meant the hydrolysis of a carboxylic acid ester, carbamate, thioester, or amide containing agent.

By “detect” refers to identifying the presence, absence, level, or concentration of an agent.

By “detectable” is meant a moiety that when linked to a molecule of interest renders the latter detectable. Such detection may be via spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes, magnetic beads, metallic beads, colloidal particles, fluorescent dyes, electron-dense reagents, enzymes (for example, as commonly used in an ELISA), biotin, digoxigenin, or haptens.

By “drug” is meant a chemical compound, composition, agent (e.g., a pharmaceutical agent) capable of inducing a pharmacological effect in a subject. A drug when properly administered to a patient as a pharmaceutical agent has a desired therapeutic effect.

Additionally, the term “drug” also encompasses drugs regarded as illicit or having the potential for abuse, such as cocaine, meperidine (Demerol[®]) or heroin.

By “genotype” is meant the genetic composition of a cell, organism, or individual. With reference to the invention, the genotype of an individual is determined as heterozygous or homozygous for one or more variant alleles of interest.

By "genotyping" is meant the characterization of the two alleles in one or more genes of interest (i.e., to determine a genotype).

By "heterozygous" is meant that a chromosomal locus has two different alleles. In one embodiment of the invention, heterozygous refers to a genotype in which one allele has a wild-type *CES1* sequence (e.g., encoding a *CES1* that has carboxylesterase activity) and the other allele has a sequence encoding a *CES1* variant that does not have carboxylesterase activity (e.g., an alteration in serine 221 (S), glutamic acid 354 (E), histidine 468 (H), Gly141-143 (G), Asp260Glu frameshift).

By "homozygous" is meant that a chromosomal locus has two identical alleles. In the invention, homozygous wild-type is meant to refer to a genotype in which both alleles have a wild-type *CES1* sequence (e.g., encoding a *CES1* that has carboxylesterase activity). In some embodiments, homozygous can refer to a genotype in which both alleles have a sequence encoding a *CES1* variant that does not have carboxylesterase activity (e.g., an alteration in serine 221 (S), glutamic acid 354 (E), histidine 468 (H), Gly141-143 (G), Asp260Glu frameshift). In particular embodiments, the inactive *CES1* variant alleles are identical at one or more SNPs.

By "increases" is meant a positive alteration of at least 10%, 25%, 50%, 75%, 100%, 200%, 300%, 400%, 500%, 1000%, or more.

By "long range polymerase chain reaction" or "long range PCR" is meant the amplification of a nucleic acid having a length of 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 167, 18, 19, 20, 21, 22, 23, 24, 25 kb, or more. Long range PCR involves one or more thermostable DNA polymerases, any one of which may have 3'→5' exonuclease activity.

By "marker" is meant any protein or polynucleotide having an alteration in activity, expression level, or sequence that is associated with a disease, disorder, or condition.

By "native" is meant endogenous, or originating in a sample.

As used herein a "nucleic acid or oligonucleotide probe" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (i.e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labeled with

isotopes, for example, chromophores, lumiphores, chromogens, or indirectly labeled with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of a target gene of interest.

By “prodrug” is meant any compound that must undergo bioactivation before exhibiting its intended pharmacological effects. Prodrugs can be viewed as compounds which have incorporated specialized non-toxic protective groups which are intended to exist only transiently to alter or eliminate undesirable characteristics of the active compound. Such undesirable qualities or impediments to adequate delivery of the therapeutic moiety to the intended site of action often relate to poor aqueous solubility, absorption and permeability as well as high first-pass hepatic extraction—all factors which contribute to overall poor oral bioavailability.

By “reduces” is meant a negative alteration of at least 10%, 25%, 50%, 75%, or 100%.

By “reference” is meant a standard or control condition. In various embodiments of the invention, the reference is the wild-type sequence of a gene or gene isoform.

By “subject sensitivity” is meant the pharmacological response of a subject to a compound or drug. From a drug metabolism perspective, an increase in subject sensitivity may result from a decrease in bioactivation and/or a decrease in bioinactivation of a specific compound. For example, a decrease in bioactivation of a prodrug or a decrease in bioinactivation of a drug may lead to an accumulation of the inactive prodrug, as well as non-prodrug agents which serve as hCES1 substrates. Such accumulation may lead to unanticipated concentrations of these agents following typical clinical dosing which may lead to adverse side effects, as well as the potential for toxicity in a given patient. Furthermore, in the case of prodrugs, the lack of requisite bioactivation would likely lead to therapeutic failure since the active moiety may be only minimally liberated from the prodrug, or not liberated at all. In the context of the invention, *CES1* enzyme activity is indicative of subject sensitivity. Subjects having one or more variant alleles at a *CES1* locus (e.g., *CES1A1*) have reduced *CES1* enzyme activity. In a particular embodiment, a decrease in bioinactivation of an insecticide due to reduced *CES1* enzyme activity causes subject sensitivity when the insecticide accumulates to toxic levels in the subject.

By “responsive” or “responsiveness” is meant that a subject to which a compound is administered will obtain the pharmacological effect of the compound. A decrease in responsiveness may result from a decrease or complete loss in bioactivation or an increase in bioinactivation of a compound. For example, a decrease in bioactivation of an administered

prodrug results in a decrease in systemic concentrations of the intended active compound such that the subject does not obtain or benefit from the intended pharmacological effect(s). In the context of the invention, *CES1* enzyme activity is indicative of subject responsiveness. Subjects having two wild-type alleles at a *CES1* locus (e.g., *CES1A1*) have normal *CES1* enzyme activity and are responsive. Subjects having one or more variant alleles at a *CES1* locus (e.g., *CES1A1*) have reduced *CES1* enzyme activity and are not responsive. In one embodiment, a subject having normal *CES1* enzyme activity indicates that drug therapy is appropriate. In another embodiment, a subject having reduced *CES1* enzyme activity indicates that drug therapy is inappropriate.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (for example, total cellular or library DNA or RNA).

By "single nucleotide polymorphism" or "SNP" is meant a DNA sequence variation occurring when a single nucleotide in the genome differs between members of a biological species or paired chromosomes in an individual. SNPs are used as genetic markers for variant alleles.

By "target nucleic acid molecule" is meant a nucleic acid or biomarker of the sample that is to be detected.

By "variant" as is meant a polynucleotide or polypeptide sequence that differs from a wild-type or reference sequence by one or more nucleotides or one or more amino acids. Exemplary *CES1* variants include *CES1A1*, *CES1A2*, and *CES1A3*.

Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having "substantial identity" to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. By "hybridize" is meant pair to form a double-stranded molecule between complementary polynucleotide sequences (e.g., a gene described herein), or portions thereof, under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507).

For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM

trisodium citrate, and more preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and more preferably at least about 50% formamide. Stringent

5 temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these

10 various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate,

15 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt concentration and by

20 temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include a temperature of at least about 25° C, more preferably of at least about 42° C, and even more preferably of at least about 68° C. In a

25 preferred embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS.

Additional variations on these conditions will be readily apparent to those skilled in the art.

30 Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel et al. (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning

Techniques, 1987, Academic Press, New York); and Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

By "substantially identical" is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence (for example, any one of the amino acid sequences described herein) or nucleic acid sequence (for example, any one of the nucleic acid sequences described herein). Preferably, such a sequence is at least 60%, more preferably 80% or 85%, and more preferably 90%, 95% or even 99% identical at the amino acid level or nucleic acid to the sequence used for comparison.

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B depict genotyping of individuals heterozygous for *CES1A1* 428G>A (Gly143Glu) allele, which is a marker for poor drug metabolism. Figure 1A is a gel showing PCR amplification of ~14kb fragments from *CES1A1* and *CES1A3/CES1A2* genes. Figure 1B depicts sequencing chromatograms of *CES1A1* exon 4 of a poor metabolizer (PM) and the biological parents, indicating that the PM and the father are heterozygous for *CES1A1* 428G>A (Gly143Glu) and that the mother is wild-type (WT).

Figure 2 is a DNA sequencing chromatogram depicting the identification of an individual homozygous for the *CES1A1*/Gly143Glu (428G>A) allele.

Figure 3 is a graph depicting the detection of variant Gly143Glu (428G>A) of *CES1A1* using a previously developed Taqman[®] assay. The Taqman[®] assay was not able to distinguish the homozygote from the heterozygotes.

Figure 4 provides exemplary sequences of human *CES1A1*, *CES1A2*, and *CES1A3* polypeptides and nucleic acid molecules.

DETAILED DESCRIPTION OF THE INVENTION

As described below, the present invention features compositions and methods that provide for genotyping *CES1* isoforms (*CES1A1*, *CES1A2*, and *CES1A3*).

Advantageously, the present invention provides genotype information for one or more of *CES1A1*, *CES1A2*, and *CES1A3* genes. The invention is based, at least in part, on the discovery that genotyping *CES1A1* alleles characterizes carboxylesterase activity in a subject. A reduced level of carboxylesterase activity in a subject is indicative of abnormal drug metabolism. Current Taqman[®]-based methods, which employ hydrolysis probes relying on the 5'-3' exonuclease activity of Taq polymerase to cleave a dual-labeled probe during hybridization to the complementary target sequence (with fluorophore-based detection) are incapable of distinguishing variants between *CES1* gene isoforms due to the high similarity of their DNA sequences. The current Taqman[®]-based methods for detecting *CES1* alleles are incapable of distinguishing between homozygous and heterozygous genotypes for one or more of the *CES1* gene isoforms (*CES1A1*, *CES1A2*, and *CES1A3*). The results presented herein describe the successful application of a novel specific *CES1* genotyping assay to human DNA samples. Genotyping performed in accordance with the methods of the invention are useful for the diagnosis, monitoring, or characterization of virtually any disease, disorder, or condition characterized by an alteration in nucleic acid sequence in a gene having high sequence similarity to one or more genes, for example, *CES1A1*, *CES1A2*, and *CES1A3*.

Mammalian Carboxylesterases (CESs)

Mammalian carboxylesterases (CESs) are members of the α/β -hydrolase family of proteins encoded by multiple genes (Imai et al., 2006) and ~1% of the human proteome is comprised of serine hydrolases (Ross and Crow, 2007). CESs assume prominent roles in the hydrolysis of numerous and diverse compounds including carboxylic acid esters, carbamates, thioesters, and amide containing agents. These substrates are represented in every therapeutic drug class inclusive of many prodrugs, agents purposely formulated as esters for the purpose of improving oral bioavailability of the active moiety. Further, endogenous substrates such as triacylglycerols, fatty acyl-CoA esters, and some environmental toxins have been identified as substrates (Satoh et al., 2002). In general, CESs catalyze the conversion of lipophilic ester substrates into more water-soluble carboxylic acids facilitating their elimination (Brzezinski et al., 1997). CES-mediated hydrolysis is a catabolic process that

transforms an active substrate to an inactive metabolite (e.g., metabolism of methylphenidate or clopidogrel (Plavix®), or yields more active products (e.g., conversion of heroin to monoacetylmorphine and morphine) (Brzezinski et al., 1997; Bencharit et al., 2003; Bencharit et al., 2006; Imai, 2006). The expression of CESs is generally highest in the
5 epithelia of most organs within the endoplasmic reticulum. Localization suggests that CES enzymes provides a protective function against toxins. The highest CES hydrolytic activity occurs in the liver.

At least four main groups of CESs (CES1-CES4) as well as several subgroups are recognized in humans according to the homology of the amino acid sequence (Imai, 2006).

10 However, two human isoenzymes hCES1 and hCES2, which belong to classes CES1 and CES2, respectively, are relevant to human drug metabolism. hCES1 and hCES2 are known to exhibit broad substrate specificities. In some cases, both hCES1 and hCES2 contribute sequentially to the metabolism of the same molecule (e.g. irinotecan [CPT-11]). However, significant differences exist between hCES1 and hCES2 relative to localization, substrate
15 specificity, immunological properties and gene regulation (Satoh and Hosokawa, 1998; Imai et al., 2006).

Human Carboxylesterase 1 (hCES1)

hCES1 is the most abundant carboxylesterase expressed in the human liver (~50-fold
20 greater than hCES2), and contributes approximately 80% of total hepatic hydrolytic activity (Imai, 2006). hCES1 is essential for the activation of a number of specific prodrugs, such as oseltamivir, trandolapril (Shi et al., 2006; Zhu et al., 2009b). Additionally, hCES1, (but not hCES2) catalyzes transesterification reactions with ester drugs and endogenous substrates when ethanol is co-ingested (i.e. forming ethyl-esters). Transesterification products include a
25 variety of agents which can be active, inactive, or potentially toxic (Bencharit et al., 2003; Bencharit et al., 2006). hCES1 is also known to exhibit stereoselectivity for a number of substrates, such as methylphenidate (MPH) and cocaine (Brzezinski et al., 1997; Sun et al., 2004). Drug metabolism and dispositional data generated in animal models is problematic due to interspecies differences in CESs (Hosokawa, 2010). For example, humans are one of
30 the few mammals that have no plasma carboxylesterase activity, and even non-human primates differ significantly from humans relative to tissue expression of CES1 and CES2.

hCES1 is encoded in humans by the *CES1* gene. Three isoforms of the *CES1* gene have been identified, i.e. *CES1A1*, *CES1A2*, and *CES1A3* (Hosokawa et al., 2007; Fukami et al., 2008). *CES1A1* and *CES1A3* are inversely located on chromosome 16 q13-q22.1 while

CES1A2 is a variant of the *CES1A3* gene (Fukami et al., 2008). Both *CES1A1* and *CES1A2* are functional whereas *CES1A3* is a pseudogene due to a premature stop codon located in exon 3. *CES1A1* is identical to *CES1A2* except its promoter region and exon 1 (Fukami et al., 2008). *CES1A1* and *CES1A2* isoforms produce an identical mature protein. In the liver, the majority of hCES1 is the product of the *CES1A1* gene because the transcription efficiency of the *CES1A2* gene is substantially lower than that of the *CES1A1* due to the different promoter structures associated with the two genes (Fukami et al., 2008; Hosokawa et al., 2008). *CES1A1* variants have a more significant influence on enzyme activity than *CES1A2* and *CES1A3* variants, which contribute less to hCES1 activity. Thus, genetic variants present in *CES1A1* have a greater impact on hCES1 function than when the same variants are present in *CES1A2* or *CES1A3*.

The expression and activity of hCES1 exhibit substantial interindividual variability (Yang et al., 2009; Zhu et al., 2009a). Genetic variation is a contributing factor to varied hCES1 function in humans. Significant adverse effects can result as a consequence of variable metabolism or pharmacokinetics of a drug between individuals. Adverse effects (e.g., toxicity, etc.) may occur when a typically therapeutic amount or dose of a medicinal or therapeutic agent, e.g., methylphenidate (Ritalin[®]) is administered to an individual and results in elevated systemic concentrations (e.g., blood concentration), and attendant CNS concentrations, due to altered metabolism of the therapeutic agent. For example, therapeutic doses of *dl*-methylphenidate can occasionally cause significant increases in a number of cardiovascular parameters due to both central dopaminergic effects and increased plasma epinephrine concentrations. In rare instances, stroke or sudden death have been reported in patients with underlying risk factors and has led to the United States Food and Drug Administration to mandate that drug manufacturers of psychostimulants provide educational literature on the risks to patients. In the instance of the prescribing and administering of a prodrug serving as a hCES1 substrate, such as the antiviral oseltamivir (Tamiflu[®]), a patient with deficient hCES1 activity is subject to risks of both therapeutic failure (in lieu of requisite bioactivation to active drug), and potential toxicity secondary to the unintended accumulation of the non-hydrolyzed prodrug.

Previous studies identified two novel SNPs Gly143Glu and Asp260fs within the *CES1A1* and *CES1A2* genes, respectively, in a normal volunteer who participated in a pharmacokinetic study of *dl*-MPH (Patrick et al., 2007; Zhu et al., 2008). The subject was found to have enormous systemic blood concentrations of both *d*- and *l*-MPH and an unprecedented enrichment of *l*-MPH relative to *d*-MPH following a single modest dose (0.3

mg/kg) (Patrick et al., 2007). Additionally, a prolonged half-life ($t_{1/2}$) and significantly altered pharmacodynamic (PD) effects were observed (Zhu et al., 2008). Subsequent investigations have revealed the minor allele frequency (MAF) of Gly143Glu to be 3.7%, 4.3%, 2.0, and 0% in Caucasian, Black, Hispanic, and Asian populations, respectively. The
5 Asp260fs variant was extremely rare as none of the 925 screened subjects carried this mutation.

The functional consequences of both mutations have been investigated utilizing cell lines stably expressing each mutant (Zhu et al., 2008; Zhu et al., 2009b; Zhu and Markowitz, 2009). Briefly, cell lines stably expressing wild-type or a mutant *CES1* are established using
10 standard co-transfection methods, and a cleared supernatant from lysates of each of the cell lines is prepared. Hydrolysis reactions of a substrate (e.g., *p*-nitrophenol acetate (PNP), methylphenidate (MPH)) are performed using the *CES1* wild-type and *CES1*mutant supernatants, using a range of substrate concentrations. Collected data, accounting for spontaneous hydrolysis, were fit to the Michaelis-Menten equation, and kinetic parameters
15 were calculated with nonlinear regression analysis with Graphpad Prism software (Graphpad Software, San Diego, CA, USA).

In vitro incubation studies demonstrated that the catalytic function of both Gly143Glu and Asp260fs was profoundly impaired in terms of hydrolyzing the hCES1 selective substrates MPH, trandolapril, and oseltamivir (Zhu et al., 2008; Zhu et al., 2009b; Zhu and
20 Markowitz, 2009). This finding represented the first direct correlation between a *CES1* mutation and aberrant drug concentration of a medication in a human subject.

A clinical study investigating the association of the Gly143Glu mutation with response to MPH therapy in ADHD patients (Nemoda et al., 2009) expands and supports Applicants' discoveries regarding the function of Gly143Glu in CES1 activity. Preliminary
25 data from the study demonstrated that patients carrying the Gly143Glu variant required significantly lower doses of MPH for symptom reduction relative to the subjects carrying WT *CES1* (0.410 ± 0.127 vs. 0.572 ± 0.153 mg/kg, $t(1,88) = 2.33$, $p = 0.022$). However, this study did not examine MPH blood concentrations (Nemoda et al., 2009).

In addition to the novel variants, the *CES1* SNPs Cyc88Phe, and Arg200His have
30 been determined to exert significantly *less* catalytic activity, while Ser76Asn exhibits slightly *higher* enzymatic activity relative to WT enzyme *in vitro* (Shi et al., 2006). To date, a total of 26 nonsynonymous SNPs of *CES1* have been recorded in the NCBI dbSNP database (Table 1). Except for the two variants (rs71647871 and rs71647872) identified by the Applicants,

the genomic locations (in *CESIA1*, *CESIA2*, or *CESIA3* genes) of the *CESI* variants listed in Table 1 have not been determined.

Table 1. Nonsynonymous CES1 variants

Region	dbSNP rs#	Mutations	Minor Allele Frequency (MAF)
Exon 1	rs111604615	G119C, Arg4Pro	NA
	rs114788146	A127G, Ile7Val	NA
	rs116258771	C139A, Leu11Ile	NA
	rs28563878	G142T, Ser12Ala	NA
Exon 2	rs3826190	G164T, Gly19Val	NA
	rs3826191	C208A, Leu34Met; C208G, Leu34Val	NA
	rs3826192	G223A, Val39Ile	NA
	rs72808055	A256G, Ile50Val	NA
	rs3177828	C278A, Ala57Gly	NA
	rs75463934	C284T, Pro59Leu	NA
	rs2307240	G335A, Ser76Asn	NIHPDR: 0.051; CEU: 0.034, HCB: 0.068, JPT: 0.023; YRI: 0.009
			NA
Exon 3	rs62028647	C356T, Ser83Leu	NA
	rs5023782	G371T, Cys88Phe	NA
	rs5023781	C400A, Leu98Ile	NA
	rs5023780	C421T, Arg105X	NA
Intron 3	rs28760313	C449T, Ser114Phe	NA
	rs469103814	T9298C	W: 0.028, B: 0.000, O:0.000
	rs469103815	A9324G	W: 0.023, B: 0.000, O:0.000
	rs469103816	T9386C	W: 0.000, B: 0.071, O:0.000
Exon 4	rs469103817	C9387A	W: 0.011, B: 0.000, O:0.000
	rs71647871	G539A, Gly144Glu	W: 0.037, B: 0.043, H:0.020, A: 0.000
Exon 5	rs4784575	G629A, Gly174Asp	NA
	rs60054861	G668C, Arg187Pro	NA
Intron 3	rs469103818	C9604T	W: 0.011, B: 0.321, O:0.000
	rs469103819	G9635A	W: 0.006, B: 0.000, O:0.000
	rs2307241	G668del, Arg187fs	NIHPDR: 0.013
	rs2307243	G707A, Arg200His	NIHPDR: 0.012
	rs2307227	C720A, Asp204Glu	NIHPDR: 0.037; CEU: 0.018; HCB: 0.045; JPN: 0; YRI: 0.043
Exon 6	rs71647872	T891del, Asp261fs	NA
Exon 7	rs115629050	G913T, Ala269Ser	NA
	rs114119971	C963G, His285Gln	NA
Exon 11	rs114277361	T1406C, Ile433Thr	NA

CEU: HapMap Utah residents with Northern and Western European ancestry from the CEPH collection

HCB: HapMap Han Chinese in Beijing

JPN: HapMap Japanese in Tokyo

YUI: HapMap Yoruba in Ibadan

NIHPDR: the NIH polymorphism discovery resource

W: White; B: Black; H: Hispanic; A: Asian; O: Other than White, Black

A recently published study reported a number of novel SNPs in both *CESIA1* and *CESIA2* genes (Yamada et al., 2010). However, based upon the primers used in this report, the

investigators were not able to distinguish between *CESIA1* and *CESIA3/CESIA2* genes. For

instance, alignment analysis of *CES1A1* (AB11995) and *CES1A2* (AB11996) cDNA suggested that nonsynonymous SNPs (11G>C, 15C>T, 17T>C, and 19A>G) within *CES1A1* exon 1 were not *CES1A1* variants but rather the wild-type *CES1A2* gene. Thus, genotyping of the nonsynonymous and promoter SNPs reported in the study need to be performed to determine the location of the *CES1* SNPs in *CES1A1*, *CES1A2*, *CES1A3* isoforms.

In addition to nonsynonymous SNPs, a total of 11 SNPs within the promoter region of the *CES1A2* gene were identified and reported in a recent study (Yoshimura et al., 2008). Among them, -816A>C was determined to be significantly associated with the efficacy of the ACE-inhibitor imidapril, a prodrug activated by hCES1 in humans (Geshi et al., 2005). In addition to these promoter and nonsynonymous mutations, a significant number of other *CES1* variants, such as synonymous mutants and SNPs within introns, have been documented in several SNPs databases. However, no functional consequences or clinical significance have been attributed to these mutations to date.

***CES1* Genetic Variants and hCES1 Structure and Function**

The hCES1 enzyme belongs to a larger family of serine hydrolases, which include human acetylcholinesterase (AcChE) and butyrylcholine esterase (BuChE). Crystal structures of human CES1, AcChE, and BuChE indicate that each has an analogous active site groove containing a catalytic triad consisting of a serine, a glutamic acid, and a histidine residue (Fleming et al., 2005).

Glycine at position 143 of hCES1 is important for hCES1 protein function. For hCES1, the corresponding active site triad residues are serine 221 (S), glutamic acid 354 (E), and histidine 468. A series of three consecutive glycine residues are also located in the active site of hCES1 (Gly141-143) and create what is referred to as an "oxyanion hole." The oxyanion hole is thought to stabilize substrate-enzyme intermediates via hydrogen bonds formed with the oxyanion form of the carbonyl oxygen and, thus, would be fundamental to proper hCES1 function (Fleming et al., 2005). The catalytic triad and oxyanion hole are evolutionarily conserved both across species (fish to humans) and within related serine hydrolases (Fleming et al., 2005). When the glycine in hBuChE analogous to Gly143 in hCES1 was mutated both substrate affinity and catalysis were markedly reduced or abolished (Masson et al., 2007). Without being bound to a particular theory, the mutation of Gly143 to glutamic acid (Gly143Glu) in hCES1 results in reduced carboxylesterase activity due to disruption of the oxyanion hole. hCES1 activity can be assessed by in vitro incubation assays previously described (Zhu et al., 2008; Zhu et al., 2009b; Zhu and Markowitz, 2009).

Similarly, the 12785T>del mutation (Asp260fs) causes a significant change in hCES1 structure. The deletion of nucleotide 780 in *CES1* causes an early truncation and alteration of residues 260-299. Thus, this frameshift mutation eliminates two of the three conserved catalytic triad residues as well as other residues involved in protein function.

Both of the polymorphisms Gly143Glu and Asp260fs, either alone or in combination, result in a significant decrease of loss of hCES1 activity. A slow metabolizer was identified as heterozygous for both mutations and each mutation occurred on a different allele. The identification of the slow metabolizer was instrumental in the discovery the two SNPs.

Diagnostic Methods

The present invention provides a number of diagnostic assays that are useful for characterizing the genotype of a subject. The present invention can be employed to genotype a gene of interest in a subject, where the gene of interest has a similar isoform(s). Desirably, the methods of the invention discriminate between the genotype of a gene of interest and the genotype of the similar isoform(s). Preferably, both or all alleles corresponding to a gene of interest are identified. Accordingly, the invention provides for genotyping useful in virtually any clinical setting where conventional methods of analysis are used.

In various aspects, the methods of the invention determine or detect the *CES1* genetic variants comprising the genotype of *CES1A1* and distinguish the *CES1A1* genotype from those of *CES1A2* and *CES1A3*. In contrast to previous methods for detecting *CES1* allelic variants, the genotyping methods described herein are able to more accurately assess CES1 activity by examining the contributions of the *CES1* isoforms, in particular *CES1A1*. The present methods provide a genetic means for the analysis of biomarkers in *CES1* associated with drug metabolism. Results obtained from *CES1* genotyping assays may be used to select an appropriate therapy for a subject, monitor drug therapy in a subject, identify a subject as responsive to drug therapy, or identify a subject as sensitive to a drug. This level of genotyping will better enable individualized pharmacogenetic-based therapy. In particular embodiments, the invention provides for the detection of *CES1* allelic variants and SNPs listed in Table 1 (above). Advantageously, the methods of the invention distinguish between homozygous and heterozygous alleles of *CES1*.

Types of Biological Samples

The genotyping methods of the invention involve detecting or determining a genetic variant or biomarker of interest in a biological sample. In one embodiment, the biologic sample contains a cell having diploid DNA content. Human cells containing 46 chromosomes (e.g., human somatic cells) are diploid. In one embodiment, the biologic sample is a tissue sample that includes diploid cells of a tissue (epithelial cells) or organ (e.g., skin cells). Such tissue is obtained, for example, from a cheek swab or biopsy of a tissue or organ. In another embodiment, the biologic sample is a biologic fluid sample. Biological fluid samples containing diploid cells include saliva, blood, blood serum, plasma, urine, hair follicle, or any other biological fluid useful in the methods of the invention.

Genotyping of *CES1* polymorphisms

A *CES1* isoform is amplified by long range PCR to determine the genotype of the isoform, e.g., *CES1A1*. The amplified nucleic acid corresponding *CES1* isoform may be analyzed using a variety of methods for detecting variant alleles to determine the genotype.

The presence or absence of one or both of the 12754T>del or Gly143Glu (428G>A) polymorphisms in the *CES1* gene may be evaluated using various techniques. For example, the carboxylesterase-1 gene is amplified by long range PCR and sequenced to determine the presence or absence of a single nucleotide polymorphism (SNP). In certain embodiments, real-time PCR may be used to detect a single nucleotide polymorphism of the amplified products. In other embodiments, a polymorphism in the amplified products may be detected using a technique including hybridization with a probe specific for a single nucleotide polymorphism, restriction endonuclease digestion, primer extension, microarray or gene chip analysis, mass spectrometry, or a DNase protection assay.

Long range Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is widely known in the art. For example, U.S. Pat. Nos. 4,683,195, 4,683,202, and 4,800,159; K. Mullis, Cold Spring Harbor Symp. Quant. Biol., 51:263-273 (1986); and C. R. Newton & A. Graham, Introduction to Biotechniques: PCR, 2nd Ed., Springer-Verlag (New York: 1997), the disclosures of which are incorporated herein by reference, describe processes to amplify a nucleic acid sample target using PCR amplification extension primers which hybridize with the sample target. As the PCR amplification primers are extended, using a DNA polymerase (preferably thermostable), more sample target is made so that more primers can be used to repeat the process, thus amplifying the sample target sequence. Typically, the reaction conditions are cycled between

those conducive to hybridization and nucleic acid polymerization, and those that result in the denaturation of duplex molecules.

In the first step of the reaction, the nucleic acid molecules of the sample are transiently heated, and then cooled, in order to denature double stranded molecules. Forward and reverse primers are present in the amplification reaction mixture at an excess concentration relative to the sample target. When the sample is incubated under conditions conducive to hybridization and polymerization, the primers hybridize to the complementary strand of the nucleic acid molecule at a position 3' to the sequence of the region desired to be amplified that is the complement of the sequence whose amplification is desired. Upon hybridization, the 3' ends of the primers are extended by the polymerase. The extension of the primer results in the synthesis of a DNA molecule having the exact sequence of the complement of the desired nucleic acid sample target. The PCR reaction is capable of exponentially amplifying the desired nucleic acid sequences, with a near doubling of the number of molecules having the desired sequence in each cycle. Thus, by permitting cycles of hybridization, polymerization, and denaturation, an exponential increase in the concentration of the desired nucleic acid molecule can be achieved.

The methods of the present invention involve amplifying large regions of a polynucleotide with high fidelity using a thermostable DNA polymerase having 3'→5' exonuclease activity. As defined herein, "3'→5' exonuclease activity" refers to the activity of a template-specific nucleic acid polymerase having a 3'→5' exonuclease activity associated with some DNA polymerases, in which one or more nucleotides are removed from the 3' end of an oligonucleotide in a sequential manner. Polymerase enzymes having high fidelity 3'→5' exonuclease activity are useful, for example, when primer extension must be performed over long distances (i.e., when the desired PCR amplification product is greater than about 5 kb). Polymerase enzymes having 3'→5' exonuclease proofreading activity are known to those in the art. Examples of suitable proofreading enzymes include TaKaRa LA Taq (Takara Shuzo Co., Ltd.) and Pfu (Stratagene), Vent, Deep Vent (New England Biolabs). Exemplary methods for performing long range PCR are disclosed, for example, in U.S. Pat. No. 5,436,149; Barnes, Proc. Natl. Acad. Sci. USA 91:2216-2220 (1994); Tellier et al., Methods in Molecular Biology, Vol. 226, PCR Protocols, 2nd Edition, pp. 173-177; and, Cheng et al., Proc. Natl. Acad. Sci. 91:5695-5699 (1994); the contents of which are incorporated herein by reference. In various embodiments, long range PCR involves one DNA polymerase. In some embodiments, long range PCR may involve more than one DNA polymerase. When using a combination of polymerases in long range PCR, it is preferable to

include one polymerase having 3'→5' exonuclease activity, which assures high fidelity generation of the PCR product from the DNA template. Typically, a non-proofreading polymerase, which is the main polymerase is also used in conjunction with the proofreading polymerase in long range PCR reactions. Long range PCR can also be performed using commercially available kits, such as LA PCR kit available from Takara Bio Inc.

Sequencing

DNA sequencing may be used to evaluate a polymorphism of the present invention. One DNA sequencing method is the Sanger method, which is also referred to as dideoxy sequencing or chain termination. The Sanger method is based on the use of dideoxynucleotides (ddNTP's) in addition to the normal nucleotides (NTP's) found in DNA. Dideoxynucleotides are essentially the same as nucleotides except they contain a hydrogen group on the 3' carbon instead of a hydroxyl group (OH). These modified nucleotides, when integrated into a sequence, prevent the addition of further nucleotides. This occurs because a phosphodiester bond cannot form between the dideoxynucleotide and the next incoming nucleotide, and thus the DNA chain is terminated. Using this method, optionally coupled with amplification of the nucleic acid target, one can now rapidly sequence large numbers of target molecules, usually employing automated sequencing apparatus. Such techniques are well known to those of skill in the art.

Pyrosequencing is another method of DNA sequencing that may be used to evaluate a polymorphism of the present invention, for example as described in U.S. Pat. Publ. No. 2006008824; herein incorporated by reference). Pyrosequencing, which is also referred to as sequencing by synthesis, involves taking a single strand of the DNA to be sequenced, synthesizing its complementary strand enzymatically one base pair at a time, and detecting by chemiluminescence the base that is added. In one embodiment, the template DNA is immobile, and solutions of A, C, G, and T nucleotides are sequentially added and removed from the reaction. Light is produced only when the nucleotide solution complements the first unpaired base of the template. The sequence of solutions which produce chemiluminescent signals allows the determination of the sequence of the template. The templates for pyrosequencing can be made both by solid phase template preparation (streptavidin-coated magnetic beads) and enzymatic template preparation (apyrase+exonuclease).

In a specific embodiment, ssDNA template is hybridized to a sequencing primer and incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase, and with the substrates adenosine 5' phosphosulfate (APS) and luciferin. The addition of one of

the four deoxynucleotide triphosphates (dNTPs)(in place of dATP, dATP α S is added, which is not a substrate for a luciferase) initiates the second step. DNA polymerase incorporates the correct, complementary dNTPs onto the template, and the incorporation of the nucleotide releases pyrophosphate (PPi) stoichiometrically. ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5' phosphosulfate. The ATP generated acts to catalyze the luciferase-mediated conversion of luciferin to oxyluciferin and generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a camera and analyzed in a program. Unincorporated nucleotides and ATP are degraded by the apyrase, and the reaction can restart with another nucleotide.

Pyrosequencing, optionally coupled with amplification of the nucleic acid target, can sequence large numbers of target molecules, usually employing automated sequencing apparatus, including long sequences (e.g., 400 million bp/10 hr in a single run).

15 Real-Time PCR (rtPCR)

The presence or absence of polymorphisms in CES1 isoforms may be detected using real-time PCR. Real-time PCR typically utilizes fluorescent probes for the selective detection of the polymorphisms. Various real-time PCR testing platforms that may be used with the present invention include: 5' nuclease (TaqMan® probes), molecular beacons, and FRET hybridization probes. These detection methods rely on the transfer of light energy between two adjacent dye molecules, a process referred to as fluorescence resonance energy transfer (see, e.g., Espy et al (2006) Clin Microbiol Rev. 2006 January; 19(1): 165-256 for a review of various rtPCR approaches that may be used with the present invention).

25 5' Nuclease Probes

In certain embodiments, a 5' nuclease probe may be used to detect a polymorphism of the present invention. 5' nuclease probes are often referred to by the proprietary name, TaqMan® probes. A TaqMan® probe is a short oligonucleotide (DNA) that contains a 5' fluorescent dye and 3' quenching dye. To generate a light signal (i.e., remove the effects of the quenching dye on the fluorescent dye), two events must occur. First, the probe must bind to a complementary strand of DNA, e.g., at about 60° C. Second, at this temperature, Taq polymerase, which is commonly used for PCR, must cleave the 5' end of the TaqMan® probe (5' nuclease activity), separating the fluorescent dye from the quenching dye.

In order to differentiate a single nucleotide polymorphism from a wild-type sequence in the DNA from a subject, a second probe with complementary nucleotide(s) to the polymorphism and a fluorescent dye with a different emission spectrum are typically utilized. Thus, these probes can be used to detect a specific, predefined polymorphism under the probe in the PCR amplification product. Two reaction vessels are typically used, one with a complementary probe to detect wild-type target DNA and another for detection of a specific nucleic acid sequence of a mutant strain. Because TaqMan® probes typically require temperatures of about 60° C. for efficient 5' nuclease activity, the PCR may be cycled between about 90-95° C. and about 60° C. for amplification. In addition, the cleaved (free) fluorescent dye can accumulate after each PCR temperature cycle; thus, the dye can be measured at any time during the PCR cycling, including the hybridization step. In contrast, molecular beacons and FRET hybridization probes typically involve the measurement of fluorescence during the hybridization step.

Genotyping for the 12754T>del ("Asp260fs") or Gly143Glu (428G>A, "Gly143Glu") in the carboxylesterase-1 gene may be evaluated using the following (5' endonuclease probe) real-time PCR technique. Genotyping assays can be performed in duplicate and analyzed on a Bio-Rad iCycler Iq® Multicolor Real-time detection system (Bio-Rad Laboratories, Hercules, Calif.). Real-time polymerase chain reaction (PCR) allelic discrimination assays to detect the presence or absence of specific single nucleotide polymorphisms in the CES1 gene, Gly143Glu (genomic: nt 9486; Cdna: nt 428) and Asp260fs (genomic: nt 12754; Cdna: nt 780), may utilize fluorogenic TaqMan® Probes.

Real-time PCR amplifications may be carried out in a 10 µl reaction mix containing 5 ng genomic DNA, 900 Nm of each primer, 200 Nm of each probe and 5 µl of 2xTaqMan® Universal PCR Master Mix (contains PCR buffer, passive reference dye ROX, deoxynucleotides, uridine, uracil-N-glycosylase and AmpliTaq Gold DNA polymerase; Perkin-Elmer, Applied Biosystems, Foster City, Calif.). Cycle parameters may be: 95° C. for 10 min, followed by 50 cycles of 92° C. for 15 sec and 60C ° for 1 min. Real-time fluorescence detection can be performed during the 60° C. annealing/extension step of each cycle. The IQ software may be used to plot and automatically call genotypes based on a two parameter plot using fluorescence intensities of FAM and VIC at 49 cycles.

Molecular Beacons

Molecular beacons are another real-time PCR approach which may be used to identify the presence or absence of a polymorphism of the present invention. Molecular beacons are

oligonucleotide probes that are labeled with a fluorescent dye (typically on the 5' end) and a quencher dye (typically on the 3' end). A region at each end of the molecular beacon probe is designed to be complementary to itself, so at low temperatures the ends anneal, creating a hairpin structure. This hairpin structure positions the two dyes in close proximity, quenching the fluorescence from the reporter dye. The central region of the probe is designed to be complementary to a region of a PCR amplification product. At higher temperatures, both the PCR amplification product and probe are single stranded. As the temperature of the PCR is lowered, the central region of the molecular beacon probe may bind to the PCR product and force the separation of the fluorescent reporter dye from the quenching dye. Without the quencher dye in close proximity, a light signal from the reporter dye can be detected. If no PCR amplification product is available for binding, the probe can re-anneal to itself, bringing the reporter dye and quencher dye into close proximity, thus preventing fluorescent signal.

Two or more molecular beacon probes with different reporter dyes may be used for detecting single nucleotide polymorphisms. For example, a first molecular beacon designed with a first reporter dye may be used to indicate the presence of a SNP and a second molecular beacon designed with a second reporter dye may be used to indicate the presence of the corresponding wild-type sequence; in this way, different signals from the first and/or second reporter dyes may be used to determine if a subject is heterozygous for a SNP, homozygous for a SNP, or homozygous wild-type at the corresponding DNA region. By selection of appropriate PCR temperatures and/or extension of the probe length, a molecular beacons may bind to a target PCR product when a nucleotide polymorphism is present but at a slight cost of reduced specificity. Molecular beacons advantageously do not require thermocycling, so temperature optimization of the PCR is simplified.

FRET Hybridization Probes

FRET hybridization probes, also referred to as LightCycler® probes, may also be used to detect a polymorphism of the present invention. FRET hybridization probes typically comprise two DNA probes designed to anneal next to each other in a head-to-tail configuration on the PCR product. Typically, the upstream probe has a fluorescent dye on the 3' end and the downstream probe has an acceptor dye on the 5' end. If both probes anneal to the target PCR product, fluorescence from the 3' dye can be absorbed by the adjacent acceptor dye on the 5' end of the second probe. As a result, the second dye is excited and can emit light at a third wavelength, which may be detected. If the two dyes do not come into close proximity in the absence of sufficient complimentary DNA, then FRET does not occur

between the two dyes. The 3' end of the second (downstream) probe may be phosphorylated to prevent it from being used as a primer by Taq during PCR amplification. The two probes may encompass a region of 40 to 50 DNA base pairs.

5 FRET hybridization probe technology permits melting curve analysis of the amplification product. If the temperature is slowly raised, probes annealing to the target PCR product will be reduced and the FRET signal will be lost. The temperature at which half the FRET signal is lost is referred to as the melting temperature of the probe system. A single nucleotide polymorphism in the target DNA under a hybridization FRET probe will still generate a signal, but the melting curve will display a lower T_m . The lowered T_m can
10 indicate the presence of a specific polymorphism. The target PCR product is detected and the altered T_m informs the user there is a difference in the sequence being detected. Like molecular beacons, there is not a specific thermocycling temperature requirement for FRET hybridization probes. Like molecular beacons, FRET hybridization probes have the advantage of being recycled or conserved during PCR temperature cycling, and a fluorescent
15 signal does not accumulate as PCR product accumulates after each PCR cycle.

Primer Extension

Primer extension is another technique which may be used according to the present invention. A primer and no more than three NTPs may be combined with a polymerase and
20 the target sequence, which serves as a template for amplification. By using less than all four NTPs, it is possible to omit one or more of the polymorphic nucleotides needed for incorporation at the polymorphic site. It is important for the practice of the present invention that the amplification be designed such that the omitted nucleotide(s) is(are) not required between the 3' end of the primer and the target polymorphism. The primer is then extended
25 by a nucleic acid polymerase, in a preferred embodiment by Taq polymerase. If the omitted NTP is required at the polymorphic site, the primer is extended up to the polymorphic site, at which point the polymerization ceases. However, if the omitted NTP is not required at the polymorphic site, the primer will be extended beyond the polymorphic site, creating a longer product. Detection of the extension products is based on, for example, separation by
30 size/length which will thereby reveal which polymorphism is present. For example, U.S. Ser. No. 10/407,846, which is hereby incorporated by reference, describes a form of primer extension.

RFLP

Restriction Fragment Length Polymorphism (RFLP) is a technique in which different DNA sequences may be differentiated by analysis of patterns derived from cleavage of that DNA. If two sequences differ in the distance between sites of cleavage of a particular restriction endonuclease, the length of the fragments produced will differ when the DNA is digested with a restriction enzyme. The similarity of the patterns generated can be used to differentiate species (and even strains) from one another.

Restriction endonucleases in turn are the enzymes that cleave DNA molecules at specific nucleotide sequences depending on the particular enzyme used. Enzyme recognition sites are usually 4 to 6 base pairs in length. Generally, the shorter the recognition sequence, the greater the number of fragments generated. If molecules differ in nucleotide sequence, fragments of different sizes may be generated. The fragments can be separated by gel electrophoresis. Restriction enzymes are isolated from a wide variety of bacterial genera and are thought to be part of the cell's defenses against invading bacterial viruses. Use of RFLP and restriction endonucleases in SNP analysis requires that the SNP affect cleavage of at least one restriction enzyme site.

Mass Spectrometry

Mass spectrometry may also be used to detect a polymorphism of the present invention. By exploiting the intrinsic properties of mass and charge, mass spectrometry (MS) can resolved and confidently identified a wide variety of complex compounds. Traditional quantitative MS has used electrospray ionization (ESI) followed by tandem MS (MS/MS) (Chen et al., 2001; Zhong et al., 2001; Wu et al., 2000) while newer quantitative methods are being developed using matrix assisted laser desorption/ionization (MALDI) followed by time of flight (TOF) MS (Bucknall et al., 2002; Mirgorodskaya et al., 2000; Gobom et al., 2000). Methods of mass spectroscopy that may be used with the present invention include: ESI, ESI tandem mass spectroscopy (ESI/MS/MS), Secondary ion mass spectroscopy (SIMS), Laser desorption mass spectroscopy (LD-MS), Laser Desorption Laser Photoionization Mass Spectroscopy (LDLPMS), and MALDI-TOF-MS.

Hybridization

There are a variety of ways by which one can assess genetic profiles, and many of these rely on nucleic acid hybridization. Hybridization is defined as the ability of a nucleic acid to selectively form duplex molecules with complementary stretches of DNAs and/or RNAs. Depending on the application envisioned, one would employ varying conditions of

hybridization to achieve varying degrees of selectivity of the probe or primers for the target sequence.

Typically, a probe or primer of between 13 and 100 nucleotides, preferably between 17 and 100 nucleotides in length up to 1-2 kilobases or more in length will allow the formation of a duplex molecule that is both stable and selective. Such fragments may be readily prepared, for example, by directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

For applications requiring high selectivity, one will typically desire to employ relatively high stringency conditions to form the hybrids. For example, relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50° C to about 70° C. Such high stringency conditions tolerate little, if any, mismatch between the probe or primers and the template or target strand and would be particularly suitable for isolating specific genes or for detecting specific mRNA transcripts. It is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide.

For certain applications, for example, lower stringency conditions may be used. Under these conditions, hybridization may occur even though the sequences of the hybridizing strands are not perfectly complementary, but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and/or decreasing temperature. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37° C. to about 55° C, while a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20° C to about 55° C Hybridization conditions can be readily manipulated depending on the desired results.

In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 1.0 mM dithiothreitol, at temperatures between approximately 20° C to about 37° C. Other hybridization conditions utilized could include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, at temperatures ranging from approximately 40° C to about 72° C.

In certain embodiments, it will be advantageous to employ nucleic acids of defined sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In preferred embodiments, one may desire to employ a

fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a detection means that is visibly or spectrophotometrically detectable, to identify specific hybridization with complementary nucleic acid containing samples.

In general, it is envisioned that the probes or primers described herein will be useful as reagents in solution hybridization, as in PCR, for detection of expression of corresponding genes, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under desired conditions. The conditions selected will depend on the particular circumstances (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, etc.). Optimization of hybridization conditions for the particular application of interest is well known to those of skill in the art.

After washing of the hybridized molecules to remove non-specifically bound probe molecules, hybridization is detected, and/or quantified, by determining the amount of bound label. Representative solid phase hybridization methods are disclosed in U.S. Pat. Nos. 5,843,663, 5,900,481 and 5,919,626. Other methods of hybridization that may be used in the practice of the present invention are disclosed in U.S. Pat. Nos. 5,849,481, 5,849,486 and 5,851,772. The relevant portions of these and other references identified in this section of the Specification are incorporated herein by reference.

Pharmacogenetics: Consequence for Drug Therapy

Carboxylesterase-1 is important for the function and metabolism of many known compounds in humans and non-human animals. Thus, the presence or absence of one or more polymorphisms of the present invention may be used to "individualize" or modify a therapy for a subject or patient based on the sensitivity of the subject to a therapeutic due to the presence or absence of a polymorphism of the present invention.

The hCES1 enzyme catalyzes the hydrolysis of drugs from numerous classes. The hydrolysis generally produces inactive metabolite(s) (e.g., MPH and cocaine). However, hCES1 is also involved in the generation of active metabolites (e.g., conversion of heroin to monoacetylmorphine and morphine) or the activation of prodrugs such as the angiotensin converting enzyme (ACE) inhibitors quinapril and benazepril. Although minor overlap exists between *CES1* and a related isoform *CES2*, only the *CES1* isoform mediates

transesterification reactions. hCES1 is the most abundant carboxylesterase expressed in the liver (~50-fold higher than hCES2), and contributes approximately 80% of total hepatic hydrolytic activity (Imai, 2006). Of the *CES1* isoforms identified *CES1A1* variants influence *CES1* hepatic hydrolytic activity more than *CES1A2* variants, because transcription of the *CES1A2* gene is substantially lower than that of the *CES1A1* (Fukami et al., 2008; Hosokawa et al., 2008).

In certain embodiments, evaluating the presence or absence of a polymorphism of the present invention may be used to individualize a therapy and/or determine the sensitivity of a subject to a compound. The compound may be a prodrug, an illicit drug, an opioid, a dopaminergic or noradrenergic drug, an ACE Inhibitor, an HMG-CoA reductase inhibitor or “statin”, an anesthetic agent, an antiviral drug, or anti-cancer drug, a toxin, a chemical warfare agent, or certain insecticides, (e.g., an organophosphate). Examples of known hCES1 substrates, including compounds of the types listed above, are shown in Table 2.

Table 2. Examples of known hCES1 substrates

lidocaine	cilazapril ¹	delapril ¹	imidapril ¹
cocaine ²	enalapril ¹	quinapril ¹	temocapril ¹
methylphenidate ²	benazepril	trandolapril	lovastatin ¹
oseltamivir ¹	meperidine ²	prasugrel	simvastatin
valacyclovir	capecitabine ^{1,2}	heroin	clopidogrel ²
sarin ⁵ , soman ⁵ , tabun ⁵	cholesterol ³	irinotecan (CPT-11) ^{1,4}	mycophenolate

1 = prodrug substrate requiring metabolic activation, 2 = agents subject to transesterification, 3 = endogenous compound, 4 = human CES-2 also contributes to biotransformation, 5 = chemical warfare agents

With regard to drug abuse, the existence of an unrecognized hCES1 deficiency could potentially lead to idiosyncratic toxicities and/or fatal exposures (e.g., misinterpreted as intentional or accidental drug overdoses on the basis of antemortem or postmortem blood concentrations). For example, the potential for the role of dysfunctional hCES1 variants in such inaccurate conclusions is highlighted by the observation of a slow metabolizer participating as a normal volunteer in a pharmacokinetic study of methylphenidate

metabolism and disposition. Following the administration of a single immediate-release weight-based dose of *dl*-methylphenidate, profound elevations in blood concentrations of methylphenidate were measured over time that were both unprecedented and far higher than any of 19 other study peers participating in the study. Furthermore, hydrolytic reactions can proceed on a stereoselective basis resulting in a distortion of the anticipated isomeric disposition of a racemic compound following the administration of a medication such as *dl*-methylphenidate and such an effect was observed in this case. The potential risk for individuals with deficient hCES1 activity who abuse methylphenidate would be expected to be amplified given the likelihood of the self-administration of doses larger than typically prescribed and at more frequent intervals. Additionally, the crushing of tablets and nasal insufflation (i.e. snorting) of the powder are well documented. Such routes of administration would be expected to deliver even greater amounts of the drug to the systemic circulation and ensuing difficulties with biodeactivation.

Methylphenidate is the most common pharmacologic agent used to treat attention-deficit hyperactivity disorder which afflicts school-age children with an estimated worldwide prevalence of 8-12%. Significant interindividual variability in MPH pharmacokinetics and pharmacodynamics is well recognized yet remains unexplained. The most common formulation of methylphenidate is the racemic mixture of *d*-threo-(R,R)- and *l*-threo-(S,S)-methylphenidate (MPH) enantiomer, with the *d*-isomer regarded as the active therapeutic isomer. The primary metabolic pathway governing the metabolism of MPH is rapid deesterification to the inactive metabolite, ritalinic acid. This process is mediated by hCES1. Additionally, Sun and coworkers demonstrated in vitro that this hydrolytic process is highly enantioselective in that the catalytic efficiency of hCES1 is up to 6-fold higher for *l*-MPH than *d*-MPH. Furthermore, pharmacokinetic studies that have measured both isomers have consistently shown that the *l*-isomer is present at only 1-15% of the blood concentration of *d*-MPH. Additionally, the plasma half-life ($t_{1/2}$) of *d*-MPH is markedly longer than that of *l*-MPH.

Based on these observations, the present invention permits one to establish a drug metabolism profile for each drug and *CES1* or variants thereof. By examining the *CES1* gene or protein of the subject involved, one can then predict which drugs will be effective in the subject (if at all), and at which doses.

Kits

The invention also provides kits for genotyping any one or more of a *CES1* isoform (*CES1A1*, *CES1A2*, and *CES1A3*). Such kits are useful for the diagnosis of a sequence alteration in *CES1* relative to wild-type *CES1* in a biological sample obtained from a subject.

5 Alternatively, the invention provides for selecting a drug treatment regimen or adjusting a dosage. In various embodiments, the kit includes at least one primer pair that identifies a *CES1* nucleic acid sequence (e.g., *CES1A1*), together with instructions for using the primers to genotype in a biological sample. In additional embodiments, the kit also includes instructions for selecting an appropriate therapy for a subject, monitoring drug therapy in a
10 subject, identifying a subject as responsive to drug therapy, or identifying a subject as sensitive to a drug. Advantageously, such testing is carried out prior to drug administration or after an adverse event associated with drug administration. Preferably, the primers are provided in combination with a thermostable DNA polymerase capable of long-range PCR amplification (e.g., a high density array). In yet another embodiment, the kit further
15 comprises a pair of primers capable of binding to and amplifying a reference sequence. In yet other embodiments, the kit comprises a sterile container which contains the primers; such containers can be boxes, ampules, bottles, vials, tubes, bags, pouches, blister-packs, or other suitable container form known in the art. Such containers can be made of plastic, glass, laminated paper, metal foil, or other materials suitable for holding nucleic acids.

20 The instructions will generally include information about the use of the compositions of the invention in genotyping a *CES1* gene isoform. In particular embodiments, the genotype identifies or characterizes a subject as having altered drug metabolism. In other embodiments, the instructions include at least one of the following: descriptions of the primer; methods for using the enclosed materials for the identification of a subject having
25 altered drug metabolism; precautions; warnings; indications; clinical or research studies; and/or references. The instructions may be printed directly on the container (when present), or as a label applied to the container, or as a separate sheet, pamphlet, card, or folder supplied in or with the container.

In various other embodiments, the kit includes reagents or components for genotyping
30 *CES1* in combination with reagents or components for the detection of a single nucleotide polymorphism (SNP) or variant of a gene encoding an enzyme involved in drug metabolism (e.g., Cytochrome P450 2D6, Cytochrome P450 2C19) or drug transporter (e.g., an ATP-binding cassette (ABC) transporters). The kits which contain reagents and components for determining a *CES1* genotype and for detecting variants in enzyme and transporters involved

in drug metabolism, are useful for guiding disease specific pharmacotherapies. For example, in the treatment of congestive heart failure (CHF), one or more drugs, including ACE inhibitors, the ionotropic drug digoxin, and beta-blockers may be prescribed depending on their predicted efficacy in a patient. The patient is evaluated for *CES1* expression or catalytic activity to predict the responsiveness of the patient to an ACE inhibitor that is a *CES1* substrate. Additionally, the patient is evaluated for one or more of P-glycoprotein SNPs, which predict the responsiveness of the patient to digoxin or CYP450 2D6 SNPs, which predict the responsiveness of the patient to beta-blockers (CYP450 2D6 SNPs). Such kits may contain one or more genomic tests of enzymes or drug transporters documented to have important SNPs. SNPs may be evaluated using a disease targeted panel of tests (e.g., a microarray). Such panels include commercially available microarrays for detecting one or more SNPs (e.g., AmpliChip® CYP450 Test; Roche). In other embodiments, the kit includes instructions for selecting one or more treatments based on the results of genotyping *CES1* and detecting one or more genetic variants in an enzyme involved in drug metabolism or drug transporter. Thus, testing performed on a patient using the kits of the invention may guide treatment selection specifically tailored to the individual.

The following examples are offered by way of illustration, not by way of limitation. While specific examples have been provided, the above description is illustrative and not restrictive. Any one or more of the features of the previously described embodiments can be combined in any manner with one or more features of any other embodiments in the present invention. Furthermore, many variations of the invention will become apparent to those skilled in the art upon review of the specification. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

It should be appreciated that the invention should not be construed to be limited to the examples that are now described; rather, the invention should be construed to include any and all applications provided herein and all equivalent variations within the skill of the ordinary artisan.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are well within the purview of the skilled artisan. Such techniques are explained fully in the literature, such as, “Molecular Cloning: A Laboratory Manual”, second edition (Sambrook, 1989);

“Oligonucleotide Synthesis” (Gait, 1984); “Animal Cell Culture” (Freshney, 1987); “Methods in Enzymology” “Handbook of Experimental Immunology” (Weir, 1996); “Gene Transfer Vectors for Mammalian Cells” (Miller and Calos, 1987); “Current Protocols in Molecular Biology” (Ausubel, 1987); “PCR: The Polymerase Chain Reaction”, (Mullis, 1994); “Current Protocols in Immunology” (Coligan, 1991). These techniques are applicable to the production of the polynucleotides and polypeptides of the invention, and, as such, may be considered in making and practicing the invention.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the assay, screening, and therapeutic methods of the invention, and are not intended to limit the scope of what the inventors regard as their invention.

EXAMPLES

Example 1. Genotyping of *CES1* isoforms identifies the presence of a variant allele in *CES1A1* linked to reduced drug metabolism.

The majority of hCES1 activity in the liver is due to the expression of the *CES1A1* gene. When a variant allele linked to altered drug metabolism is present in *CES1A1*, it has a greater effect on altering drug metabolism than when the variant allele is present in *CES1A2* or *CES1A3* isoforms. *CES1A1* is identical to *CES1A2* except in the promoter region and exon 1 (Fukami et al., 2008). Current Taqman[®]-based methods for detecting *CES1* variants are incapable of distinguishing variants in specific *CES1* isoforms due to the high similarity of their DNA sequences. Therefore, genotyping of *CES1* isoforms was used to distinguish in which *CES1* isoform a variant allele was present in an individual having reduced drug metabolism.

To develop a novel specific genotyping assay for *CES1A1* and *CES1A3/CES1A2* genes, a long-range PCR approach was used. Two sets of primers were designed in order to specifically amplify the fragments from *CES1A1* and *CES1A3/CES1A2* genes. A commercially available thermostable DNA polymerase was used (TaKaRa LA Taq[™]). The primer sequences and PCR reaction conditions are shown in Table 3. Both PCR reactions yielded ~14 kb amplification products, which span from exon 1 to exon 6 of *CES1A1* and *CES1A3/CES1A2* genes (Figure 1A). Bidirectional sequencing of the long-range PCR products revealed that specific amplification of *CES1A1* and *CES1A3/CES1A2* genes was achieved under the experimental conditions summarized in Table 3. The long-range PCR

products can be used as the templates for *CES1A1* and *CES1A3/CES1A2* genotyping (e.g., via direct DNA sequencing or Taqman[®] assay).

5 Table 3. Specific long-range PCR of *CES1A1* and *CES1A3/CES1A2* genes

Primers for the long-range PCR of *CES1A1* and *CES1A3/CES1A2* genes

Genes	Primers
<i>CES1A1</i>	Forward: 5'- TTC CAC GAT GTG CCG TGC CTT TA -3' Reverse: 5'- GGC ACA TAG GAG GAG TGT GGT CAC A -3'
<i>CES1A3/CES1A2</i>	Forward: 5'- TTC CAG GAT GTG GCT CCCTGC TCT TG -3' Reverse: 5'- GGC ACA TAG GAG GAG TGT GGT CAC A -3'

PCR mixture composition

10×LA PCR buffer II (Mg ²⁺ plus)	5 µl
dNTP mixture (2.5 mM each)	8 µl
genomic DNA	200 ng
TaKaRa LA Taq (5 units/µl)	1 µl
Forward primer	Final concentration 0.2 µM
Reverse primer	Final concentration 0.2 µM
	Add water up to 50 µl

PCR reaction conditions

94°C	1 min	30 ×
94°C	30 sec	
68°C	14 min	
72°C	10 min	
4°C	∞	

To determine whether the novel *CES1* SNP Gly143Glu (428G>A) allele was located in *CES1A1* or *CES1A3/CES1A2*, genotyping using the novel assay was performed on DNA samples collected from a poor metabolizer (PM) and the biological parents. Sequencing of the *CES1A1* PCR product from the poor metabolizer and the father demonstrated G and A peaks at position 428. In the mother, sequencing detected only an A peak at position 428, corresponding to the wild-type *CES1A1* allele. The results demonstrated that the PM and the father were heterozygous Gly143Glu/wild-type in *CES1A1*, and the mother was homozygous wild-type in *CES1A1* (Figure 1B). Additional sequencing revealed that all three subjects were wild-type for *CES1A2*, but not *CES1A3*.

Thus, the data indicated that in poor metabolizers SNP Gly143Glu was located in *CES1A1* and not in *CES1A3/CES1A2*. Because the majority of hCES1 in the liver is the product of *CES1A1* gene, the genotyping results validate clinical findings that the variant Gly143Glu has a profound impact on hCES1 functions.

Example 2. Genotyping of *CES1A1* and *CES1A3/CES1A2* distinguishes individuals homozygous or heterozygous for *CES1A1* variant alleles.

The *CES1* genotyping assay was used to genotype the *CES1A1* variant Gly143Glu in 107 saliva DNA samples collected from ADHD patients being treated with MPH. Among them, one homozygote (Figure 2) and three heterozygotes were indentified, while the others were wild-type (WT). This is the first instance that an individual homozygous Gly143Glu at *CES1A1* has been reported. By comparison, a previously developed Taqman[®] detection assay was employed to examine the DNA samples from selected subjects including the Gly143Glu homozygote, three heterozygotes, and ten randomly selected WT. The Taqman[®] assay was performed as previously described (Zhu et al., 2008). The Taqman[®] assay was not able to distinguish the homozygote from the other three heterozygotes, but grouped the homozygote with the Gly143Glu heterozygotes (Figure 3). The results provided direct evidence that non-isoform specific assays that do not discriminate among *CES1* isoforms, are not suitable for genotyping *CES1* genetic polymorphisms. In fact, screening of 1607 total subjects by non-isoform specific *CES1* assays in 3 recently published independent studies did not identify any Gly143Glu homozygotes (Zhu et al., 2008; Nemoda et al., 2009; Walter Soria et al., 2010). This result is likely due to the inability of these assays to distinguishing between *CES1A1* and *CES1A2*, and *CES1A3* genes. The discriminative *CES1* isoform genotyping assay described in the study is preferable for *CES1* pharmacogenetic studies in both clinical and basic research settings.

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

This application may be related to U.S. Patent Application No. 12/663,644, which is the U.S. national phase application, pursuant to 35 U.S.C. §371, of International Patent Application No.: PCT/US2008/0066280, filed June 9, 2008, which claims the benefit of U.S.

Provisional Application Ser. Nos. 60/942,818, 61/051,680, and 61/053,524, the disclosures of which are hereby incorporated herein in their entireties by reference.

All patents and publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent and publication was specifically

5 and individually indicated to be incorporated by reference.

References

The following documents are cited herein.

- 10 Bencharit S, Edwards CC, Morton CL, Howard-Williams EL, Kuhn P, Potter PM and Redinbo MR (2006) Multisite promiscuity in the processing of endogenous substrates by human carboxylesterase 1. *J Mol Biol* 363:201-214.
- Bencharit S, Morton CL, Xue Y, Potter PM and Redinbo MR (2003) Structural basis of heroin and cocaine metabolism by a promiscuous human drug-processing enzyme.
- 15 *Nat Struct Biol* 10:349-356.
- Brzezinski MR, Spink BJ, Dean RA, Berkman CE, Cashman JR and Bosron WF (1997) Human liver carboxylesterase hCE-1: binding specificity for cocaine, heroin, and their metabolites and analogs. *Drug Metab Dispos* 25:1089-1096.
- Fleming C D, Bencharit S, Edwards C C, et al. Structural Insights into Drug Processing by
- 20 Human Carboxylesterase 1: Tamoxifen, Mevastatin, and Inhibition by Benzil. *Journal of Molecular Biology* 2005; 352: 165-177.
- Fukami T, Nakajima M, Maruichi T, Takahashi S, Takamiya M, Aoki Y, McLeod HL and Yokoi T (2008) Structure and characterization of human carboxylesterase 1A1, 1A2, and 1A3 genes. *Pharmacogenet Genomics* 18:911-920.
- 25 Geshi E, Kimura T, Yoshimura M, Suzuki H, Koba S, Sakai T, Saito T, Koga A, Muramatsu M and Katagiri T (2005) A single nucleotide polymorphism in the carboxylesterase gene is associated with the responsiveness to imidapril medication and the promoter activity. *Hypertens Res* 28:719-725.
- Hosokawa M (2010) Are non-human primates useful experimental animals for pre-clinical
- 30 study? *Drug Metab Pharmacokinet* 25:221-222.
- Hosokawa M, Furihata T, Yaginuma Y, Yamamoto N, Koyano N, Fujii A, Nagahara Y, Satoh T and Chiba K (2007) Genomic structure and transcriptional regulation of the rat, mouse, and human carboxylesterase genes. *Drug Metab Rev* 39:1-15.

- Hosokawa M, Furihata T, Yaginuma Y, Yamamoto N, Watanabe N, Tsukada E, Ohhata Y, Kobayashi K, Satoh T and Chiba K (2008) Structural organization and characterization of the regulatory element of the human carboxylesterase (CES1A1 and CES1A2) genes. *Drug Metab Pharmacokinet* 23:73-84.
- 5 Imai T (2006) Human carboxylesterase isozymes: catalytic properties and rational drug design. *Drug Metab Pharmacokinet* 21:173-185.
- Imai T, Taketani M, Shii M, Hosokawa M and Chiba K (2006) Substrate specificity of carboxylesterase isozymes and their contribution to hydrolase activity in human liver and small intestine. *Drug Metab Dispos* 34:1734-1741.
- 10 Masson P, Froment M-T, Gillon E, et al. Hydrolysis of oxo- and thio-esters by human butyrylcholinesterase. *Biochimica et Biophysica Acta (BBA)--Proteins & Proteomics* 2007; 1774: 16-34.
- Nemoda Z, Angyal N, Tarnok Z, Gadoros J and Sasvari-Szekely M (2009) Carboxylesterase 1 gene polymorphism and methylphenidate response in ADHD. *Neuropharmacology* 15 57:731-733.
- Patrick KS, Straughn AB, Minhinnett RR, Yeatts SD, Herrin AE, DeVane CL, Malcolm R, Janis GC and Markowitz JS (2007) Influence of ethanol and gender on methylphenidate pharmacokinetics and pharmacodynamics. *Clin Pharmacol Ther* 81:346-353.
- 20 Ross MK and Crow JA (2007) Human carboxylesterases and their role in xenobiotic and endobiotic metabolism. *J Biochem Mol Toxicol* 21:187-196.
- Satoh T and Hosokawa M (1998) The mammalian carboxylesterases: from molecules to functions. *Annu Rev Pharmacol Toxicol* 38:257-288.
- Satoh T, Taylor P, Bosron WF, Sanghani SP, Hosokawa M and La Du BN (2002) Current progress on esterases: from molecular structure to function. *Drug Metab Dispos* 25 30:488-493.
- Shi D, Yang J, Yang D, LeCluyse EL, Black C, You L, Akhlaghi F and Yan B (2006) Anti-influenza prodrug oseltamivir is activated by carboxylesterase human carboxylesterase 1, and the activation is inhibited by antiplatelet agent clopidogrel. *J Pharmacol Exp Ther* 30 319:1477-1484.
- Sun Z, Murry DJ, Sanghani SP, Davis WI, Kedishvili NY, Zou Q, Hurley TD and Bosron WF (2004) Methylphenidate is stereoselectively hydrolyzed by human carboxylesterase CES1A1. *J Pharmacol Exp Ther* 310:469-476.

- Walter Soria N, Belaus A, Galvan C, Ana Pasquali M, Velez P, Del Carmen Montes C and Beltramo DM (2010) A Simple Allele-Specific Polymerase Chain Reaction Method to Detect the Gly143Glu Polymorphism in the Human Carboxylesterase 1 Gene: Importance of Genotyping for Pharmacogenetic Treatment. *Genet Test Mol Biomarkers*.
- 5 Yamada S, Richardson K, Tang M, Halaschek-Wiener J, Cook VJ, Fitzgerald JM, Elwood K, Marra F and Brooks-Wilson A (2010) Genetic variation in carboxylesterase genes and susceptibility to isoniazid-induced hepatotoxicity. *Pharmacogenomics J*.
- 10 Yang D, Pearce RE, Wang X, Gaedigk R, Wan YJ and Yan B (2009) Human carboxylesterases HCE1 and HCE2: ontogenic expression, inter-individual variability and differential hydrolysis of oseltamivir, aspirin, deltamethrin and permethrin. *Biochem Pharmacol* 77:238-247.
- 15 Yoshimura M, Kimura T, Ishii M, Ishii K, Matsuura T, Geshi E, Hosokawa M and Muramatsu M (2008) Functional polymorphisms in carboxylesterase1A2 (CES1A2) gene involves specific protein 1 (Sp1) binding sites. *Biochem Biophys Res Commun* 369:939-942.
- Zhu HJ, Appel DI, Jiang Y and Markowitz JS (2009a) Age- and sex-related expression and activity of carboxylesterase 1 and 2 in mouse and human liver. *Drug Metab Dispos* 37:1819-1825.
- 20 Zhu HJ, Appel DI, Johnson JA, Chavin KD and Markowitz JS (2009b) Role of carboxylesterase 1 and impact of natural genetic variants on the hydrolysis oftrandolapril. *Biochem Pharmacol* 77:1266-1272.
- Zhu HJ and Markowitz JS (2009) Activation of the antiviral prodrug oseltamivir is impaired by two newly identified carboxylesterase 1 variants. *Drug Metab Dispos* 37:264-267.
- 25 Zhu HJ, Patrick KS, Yuan HJ, Wang JS, Donovan JL, DeVane CL, Malcolm R, Johnson JA, Youngblood GL, Sweet DH, Langaee TY and Markowitz JS (2008) Two CES1 gene mutations lead to dysfunctional carboxylesterase 1 activity in man: clinical significance and molecular basis. *Am J Hum Genet* 82:1241-1248.

What is claimed is:

1. A method of genotyping a subject for *CES1*, the method comprising
amplifying a *CES1* nucleic acid in a biological sample from the subject by long range
5 polymerase chain reaction (PCR); and
detecting a variant allele of *CES1*, thereby genotyping the subject for *CES1*.

2. The method of claim 1, wherein *CES1* is an isoform selected from the group
consisting of *CES1A1*, *CES1A2*, *CES1A3*, and combinations thereof.

3. The method of any one of claims 1 and 2, wherein *CES1* is a *CES1A1* isoform.

4. The method of any one of claims 1-3, wherein the variant allele encodes a CES1
polypeptide having reduced or increased carboxylesterase activity or expression relative to a
15 reference.

5. The method of claim 4, wherein the method identifies at least one alteration in an
evolutionarily conserved residue of CES1.

6. The method of claim 5, wherein the conserved residue is in a triad residue, in the
active site, or in the oxyanion hole.

7. The method of claim 6, wherein the conserved residue serine 221 (S), glutamic
acid 354 (E), histidine 468, or Gly141-143.

8. The method of claim 5, wherein the alteration is in a residue that stabilizes
substrate-enzyme intermediates or that alters enzyme activity.

9. The method of claim 4, wherein the variant allele is selected from the group
30 consisting of Gly143Glu and Asp260Glu frameshift.

10. The method of claim 4, wherein the variant allele is one or more of an allele set
forth in Table 1.

11. The method of any one of claims 1-10, wherein the variant allele is indicated by the presence of a single nucleotide polymorphism (SNP).

12. The method of claim 11, wherein the SNP is selected from the group consisting of 428G>A and T891del.

13. The method of claim 11, wherein the SNP is one or more of a SNP set forth in Table 1.

14. The method of any one of claims 1-13, wherein analyzing comprises nucleic acid sequencing, allele-specific hybridization, allele-specific PCR, oligonucleotide microarray analysis, or mass spectrometry.

15. A method of genotyping a subject for *CES1*, the method comprising amplifying a *CES1* nucleic acid in a biological sample from the subject by long range polymerase chain reaction (PCR); and detecting a variant allele of *CES1*, thereby genotyping the subject for *CES1*, wherein the alteration is in a residue that stabilizes substrate-enzyme intermediates or that alters enzyme activity; and wherein analyzing comprises nucleic acid sequencing, allele-specific hybridization, allele-specific PCR, oligonucleotide microarray analysis, or mass spectrometry.

16. The method of any one of claims 1-15, wherein the method distinguishes a heterozygous genotype from a homozygous genotype.

17. The method of any one of claims 1-16, wherein the method further comprises assaying the carboxylesterase activity of the protein encoded by the variant allele of *CES1*.

18. The method of any one of claims 1-17, wherein the method further comprises assaying substrate hydrolysis for the variant allele of *CES1*.

19. The method of claim 18, wherein substrate hydrolysis is determined by contacting a drug with the protein and detecting the half-life ($t_{1/2}$), pharmacokinetics of the drug, or altered pharmacodynamic (PD) effect of the drug.

20. The method of any one of claims 1-19, wherein the method further comprises selecting an appropriate drug therapy for the subject, wherein a genotype comprising the presence of a variant allele having reduced carboxylesterase activity or expression relative to a reference identifies the subject as sensitive to the drug, and wherein a genotype that is homozygous wild-type identifies the subject as not sensitive to the drug.

21. The method of claim 20, wherein the subject is administered an effective amount of the drug or drug administration to the subject is discontinued.

22. A method of selecting an appropriate therapy for a subject, the method comprising genotyping *CESI* in a subject according to the method of claim 1, wherein a genotype comprising the presence of a variant allele having reduced carboxylesterase activity or expression relative to a reference indicates that drug therapy is not appropriate for the subject, and wherein a genotype that is homozygous wild-type indicates that drug therapy is appropriate for the subject.

23. A method of identifying a subject as responsive to drug therapy, the method comprising genotyping *CESI* in a subject according to the method of claim 1, wherein a genotype comprising the presence of a variant allele having reduced carboxylesterase activity or expression relative to a reference identifies the subject as not responsive to drug therapy, and wherein a genotype that is homozygous wild-type identifies the subject as responsive to drug therapy.

24. The method of any one of claims 22 or 23, wherein an appropriate therapy is selected for the subject, wherein a genotype comprising the presence of Gly143Glu and Asp260Glu frameshift indicates that drug therapy is not appropriate for the subject, and wherein a genotype that is homozygous wild-type indicates that drug therapy is appropriate for the subject.

25. The method of any one of claims 22-24, wherein the subject is administered an effective amount of the drug.

26. The method of any one of claims 22-25, wherein the drug is a prodrug that requires *CES1* activity for bioactivation.

27. The method of claim 26, wherein the prodrug is a compound selected from the list consisting of oseltamivir, cilazapril, enalapril, capecitabine, delapril, quinapril, imidapril, temocapril, and lovastatin.

28. The method of any one of claims 22-27, wherein the method further comprises selecting an alternative therapy.

29. A method of monitoring drug therapy in a subject, the method comprising genotyping *CES1* in a subject according to the method of claim 1, identifying a subject as having a variant allele of *CES1* according to the method of claim 1, and measuring plasma levels of the drug in the subject to determine an appropriate dosage.

30. The method of claim 29, wherein an appropriate therapy is selected for the subject.

31. The method of claim 30, wherein the subject is administered an effective amount of the drug.

32. A method of characterizing subject sensitivity to a drug, the method comprising genotyping *CES1* in a subject according to the method of claim 1, wherein a genotype comprising the presence of a variant allele having reduced carboxylesterase activity or expression relative to a reference identifies the subject as sensitive to the drug, and wherein a genotype that is homozygous wild-type identifies the subject as not sensitive to the drug.

33. The method of claim 32, wherein the method further comprises selecting an alternative therapy.

34. The method of claim 32 or 33, wherein the subject is identified as having an adverse reaction to a drug that requires *CES1* activity for bioinactivation.

35. The method of any one of claims 32-34, wherein the drug is a compound selected from the list consisting of cocaine, methylphenidate, meperidine, capecitabine, and clopidogrel.

5 36. The method of any one of claims 32-34, wherein the drug is a compound selected from the list consisting of an insecticide, an organophosphate insecticide, or paraoxon.

37. The method of any one of claims 22-36, wherein the subject is heterozygous or homozygous for a variant allele having reduced carboxylesterase activity or expression
10 relative to a reference.

38. The method of any one of claims 29-37, wherein the method further comprises decreasing a dosage of a drug administered to the subject or discontinuing treatment in the subject.

15 39. The method of any one of claims 22-38, wherein genotyping *CES1* in a subject comprises amplifying a *CES1* nucleic acid by long range polymerase chain reaction (PCR) and analyzing the sequence of the amplified nucleic acid for a variant allele of *CES1* relative to a wild-type reference sequence.

20 40. The method of claim 39, wherein *CES1* is an isoform selected from the group consisting of *CES1A1*, *CES1A2*, *CES1A3*, and combinations thereof .

41. The method of any one of claims 39 and 40, wherein *CES1* is a *CES1A1* isoform.

25 42. The method of any one of claims 22-41, wherein the variant allele encodes a *CES1* polypeptide having reduced carboxylesterase activity or expression relative to a reference.

30 43. The method of any one of claims 22-42, wherein the method identifies at least one alteration in an evolutionarily conserved residue of *CES1*.

44. The method of claim 43, wherein the conserved residue is in a triad residue, in the active site, or in the oxyanion hole.

45. The method of claim 44, wherein the conserved residue serine 221 (S), glutamic acid 354 (E), histidine 468, or Gly141-143.

5 46. The method of claim 43, wherein the alteration is in a residue that stabilizes substrate-enzyme intermediates or that alters enzyme activity.

47. The method of claim 42, wherein the variant allele is selected from the group consisting of Gly143Glu and Asp260Glu frameshift.

10

48. The method of claim 42, wherein the variant allele is one or more of an allele set forth in Table 1.

15 49. The method of any one of claims 22-48, wherein the variant allele is indicated by the presence of a single nucleotide polymorphism (SNP).

50. The method of claim 49, wherein the SNP is selected from the group consisting of 428G>A and T891del.

20 51. The method of claim 49, wherein the SNP is one or more of a SNP set forth in Table 1.

25 52. The method of any one of claims 22-51, wherein analyzing comprises nucleic acid sequencing, allele-specific hybridization, allele-specific PCR, oligonucleotide microarray analysis, or mass spectrometry.

30 53. The method of any one of claims 22-52, wherein the drug is a compound selected from the group consisting of a prodrug, an illicit drug, an opioid, a dopaminergic or noradrenergic drug, an ACE Inhibitor, a statin, an anesthetic, a toxin, a chemical warfare agent, an insecticide, (e.g., an organophosphate insecticide), an antiviral drug, and an anti-cancer drug.

54. The method of claim 53, wherein the drug is a compound selected from the group consisting of lidocaine, cilazapril, delapril, imidapril, cocaine, enalapril, quinapril,

temocapril, methylphenidate, benazepril, trandolapril, lovastatin, oseltamivir, meperidine, prasugrel, simvastatin, valacyclovir, capecitabine, heroin, clopidogrel, sarin, soman, tabun, cholesterol, irinotecan (CPT-11), and mycophenolate.

5 55. A method of genotyping a subject, the method comprising:
amplifying a *CES1* nucleic acid in a biological sample from the subject by long range
polymerase chain reaction (PCR); and

 detecting a variant allele of *CES1* relative to a wild-type reference sequence, wherein
the variant allele encodes a CES1 polypeptide having reduced carboxylesterase activity or
10 expression relative to a reference selected from the group consisting of Gly143Glu and
Asp260Glu frameshift.

 56. The method of claim 55, wherein the method involves selecting an appropriate
drug therapy for the subject, wherein a genotype comprising the presence of Gly143Glu and
15 Asp260Glu frameshift indicates that drug therapy is not appropriate for the subject, and
wherein a genotype that is homozygous wild-type indicates that drug therapy is appropriate
for the subject.

 57. The method of claim 56, wherein the subject is administered an effective amount
20 of the drug drug administration to the subject is discontinued.

 58. The method of any one of claims 55-57, wherein the variant allele is indicated by
the presence of a single nucleotide polymorphism (SNP) selected from the group consisting
of 428G>A and T891del.

25 59. The method of any one of claims 55-58, wherein analyzing comprises nucleic
acid sequencing, allele-specific hybridization, allele-specific PCR, oligonucleotide
microarray analysis, or mass spectrometry.

30 60. A kit for genotyping a CES1 isoform in a subject comprising a set of nucleic acid
probes, wherein the nucleic acid probes can selectively bind to and amplify a nucleic acid at a
CES1 isoform locus.

 61. The kit of claim 50, comprising a nucleic acid probe set forth in Table 3.

62. The kit of any one of claims 60 and 61, further comprising instructions for selecting an appropriate therapy for a subject, identifying a subject as responsive to drug therapy, monitoring drug therapy in a subject, or characterizing subject sensitivity to a drug based on the genotype of a CES1 isoform of the subject.

63. The kit of claim 62, wherein the CES1 isoform comprises *CES1A1*.

64. The kit of claim 63, wherein the CES1 isoform further comprises one or more of *CES1A2* and *CES1A3*.

65. The kit of any one of claims 61-64, further comprising a microarray for detecting an SNP in a gene encoding an enzyme involved in drug metabolism or a drug transporter.

66. The kit of claim 65, wherein the enzyme is one or more of cytochrome P450 2D6 and cytochrome P450 2C19.

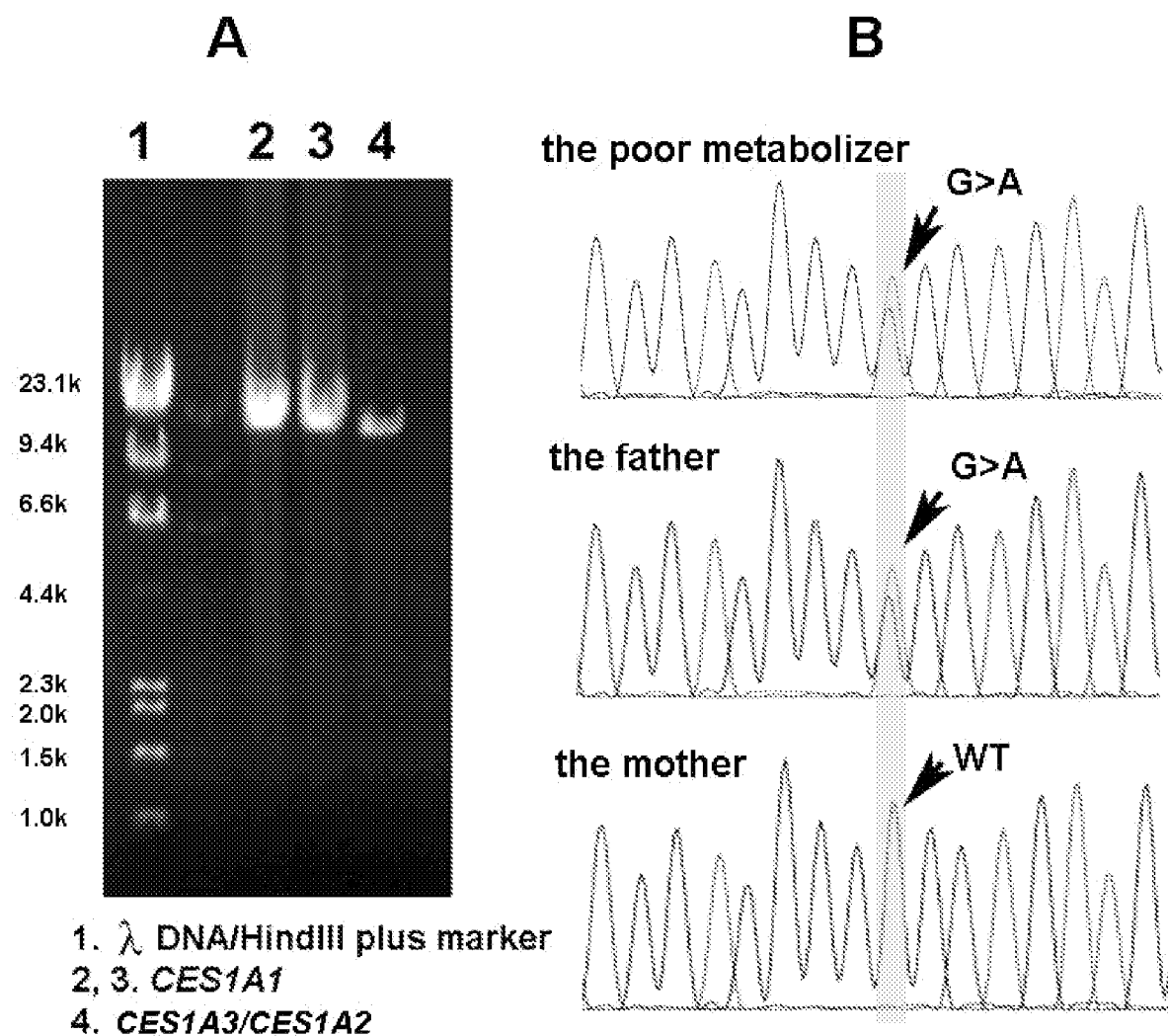
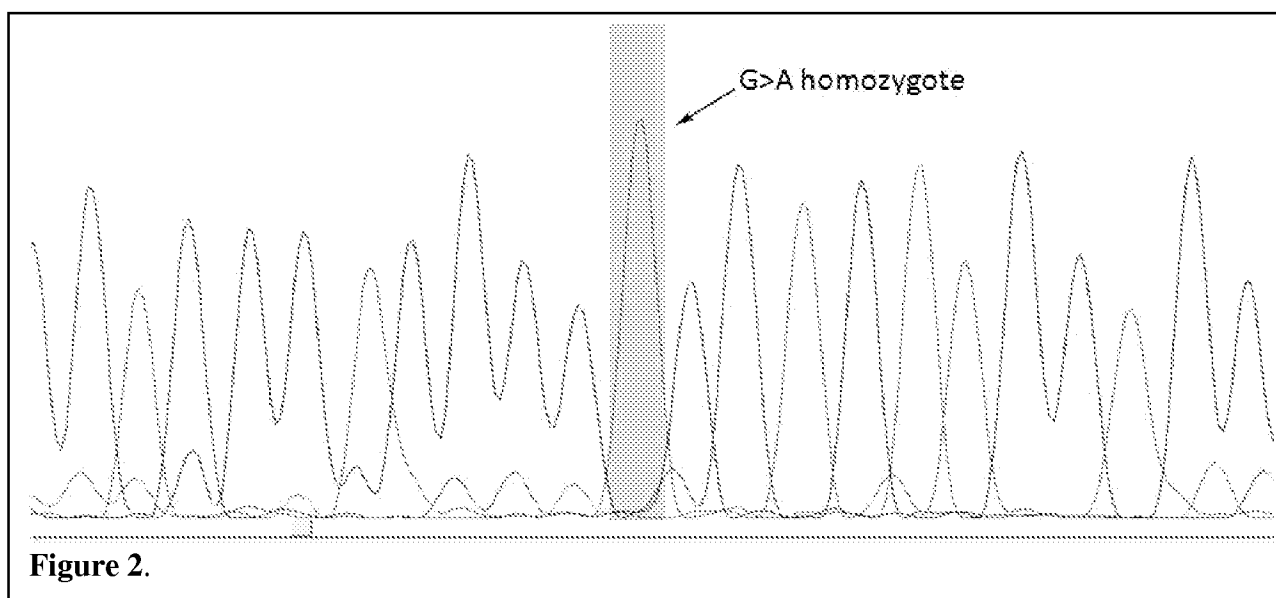


Figure 1.



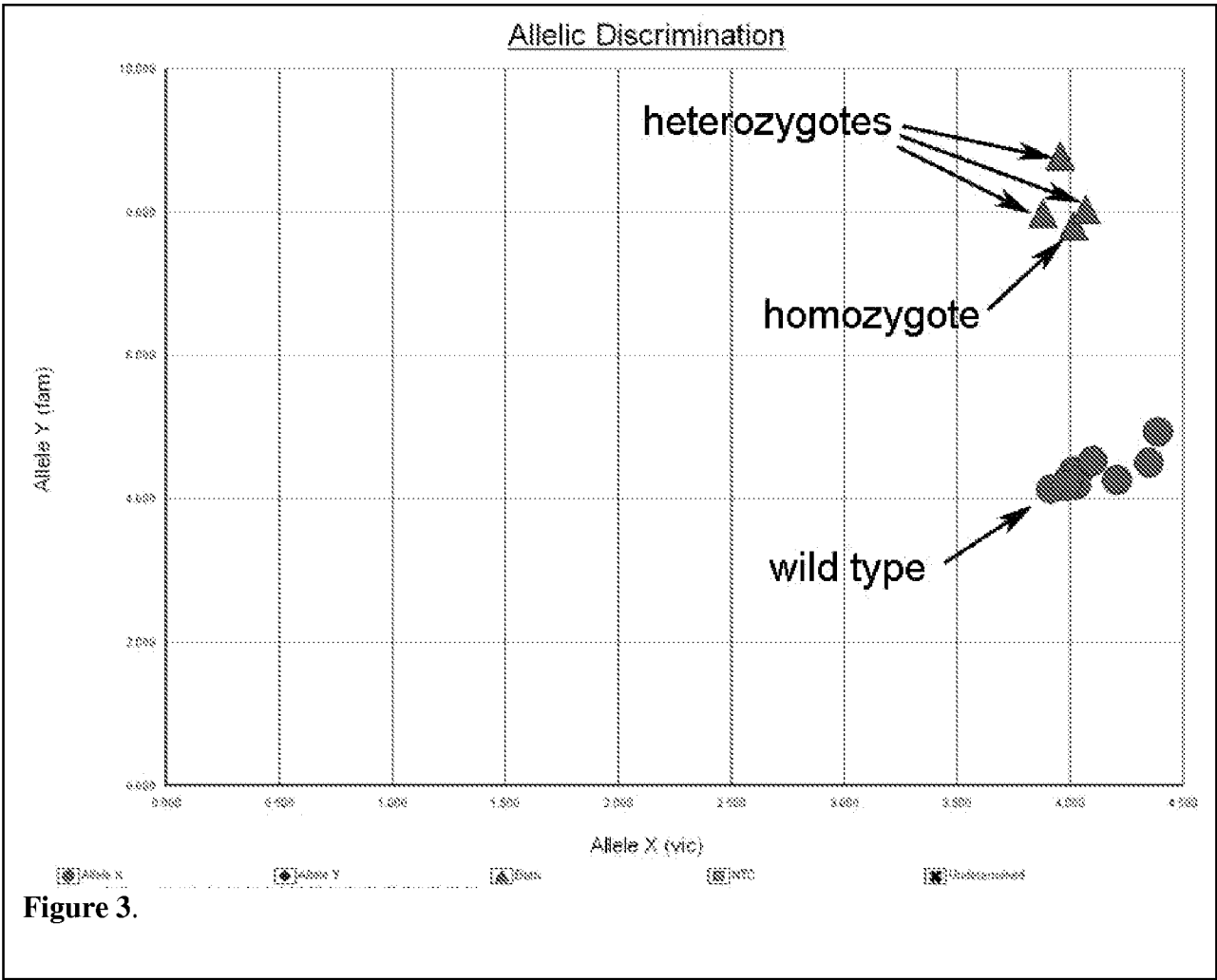


Figure 4

CES1 polypeptide

MWLRAFILATLSASAAGHP	SSPPVVDTVHGKVLGKFVSL	EGFAQPVAIFLGIPFAKPPL
GPLRFTPPQPAEPWSFVKNA	TSYPPMCTQDPKAGQLLSEL	FTNRKENIPLKLSEDCLYLN
IYTPADLTCKNRLPVMVWIH	GGGLMVGAASTYDGLALAAH	ENVVVVTIQYRLGIWGFFST
GDEHSRGNWGHLDQVAALRW	VQDNIASEFGNPGSVTIFGE	SAGGESVSVLVLSPLAKNLF
HRAISESGVALTSVLVKKGD	VKPLAEQIAITAGCKTTTSA	VMVHCLRQKTEEEELLETTLK
MKFLSLDLQGDPRESQPLL	TVIDGMLLLKTPEELQAERN	FHTVPYVMVGINKQEFGLIP
MQLMSYPLSEGQLDQKTAMS	LLWKSYPVLCIAKELIPEAT	EKYLGGTDDTVKKKDLFLDL
IADVMFGVPSVIVARNHRDA	GAPTYMYEFQYRPSFSSDMK	PKTVIGDHGDELFVFGAPF
LKEGASEEEIRLSKMVMKFW	ANFARNGNPNGEGLPHWPEY	NQKEGYLQIGANTQAAQKLK
DKEVAFWTNLFAKKAVEKPP	QTEHIEL	

Figure 4 continued

LOCUS 2011	AB119997_hCES1A1_c			30295 bp ds-DNA	linear	14-APR-
1	gcacagtccc	tctgaactgc	acagagacct	cgcaggcccc	gagaactgtc	gcccttccac
61	gatgtggctc	cgtgccttta	tcctggccac	tctctctgct	tccgcgggctt	ggggtgagtc
121	cttctgaaat	caaatatgcg	gggcactttt	tgaaatcctt	gttctggggcc	gaggtgggcg
181	cagatgcgta	gaaaggcaaa	gacacaacag	gtccgggtct	gcggcggggc	acgacttcca
241	tattaggact	tggacttggg	gcagctgagc	cgggccagcc	tggcccgcgc	ggagacgcag
301	ggagggtgaa	catcctgaga	tcgccctgta	ctgctcgccc	aaatgcaggc	cgacagccgg
361	cccacctcag	gccgcacaca	cgagaaatth	aggcagctca	gtttgctgtc	ccttaattag
421	tccccttacg	gctggatcac	ggagagcggc	tcaacataat	attaagagca	tcgactctgg
481	agccagagt	cgtgggttcg	agtcctccgc	gggcgcttcc	tcgctgtgtg	tccttaggca
541	agggactgcg	ctattccgtg	cttgggttac	ccccctctctg	aaacagaggt	ggcagtgtctg
601	cctacctcac	tgggctttgg	tgaggaggaa	atgggtggct	gtttctaaag	cactcttggc
661	agcgcctggt	atgtagaaa	cacgaaaaag	ggatggataa	ataaaggcgc	tgtttaaaaa
721	cgtagtgtat	gaatagaaaa	tttgagctgt	ggtcaggcca	ctatatcagg	acctcttgcc
781	attaaaaaga	cagtttctta	ctcatgttct	caacaggagg	gggacaccag	tgcgcggggc
841	acaaggggaa	gcagcagagc	ctggggaggg	gacagggcgg	atctgaggcc	aaggggcttt
901	tactgtggct	tccaaggata	gaagaggaaa	ggcagggcag	gcaggctgag	gactggcaag
961	ttgggatact	tcaggggctc	tggggcacag	gggctgttcc	taattgtttt	gtacctagcc
1021	ctggcatgat	tagggcaggt	gggtagtggc	cagggcatgg	caaccgatc	aaagaggcgg
1081	ttggagtgtg	ggctccaggt	tggctggacc	ctgggaggga	cagtttctcc	aaggccaaca
1141	aggcctcaga	tgccagaaca	tcagaatata	gaaagcaaaa	cagatggtac	atagaggggt
1201	gcagcgaggg	gcctcgaaat	acccccgtgc	cattttatag	acagagaaac	tgagactcac
1261	acagggaaat	gtacttgccc	agggccaacc	cgaaaatcac	gcctaggtct	gtctctgggt
1321	cctgaatcct	gacagcccag	ccccgtctgc	atcctgtctt	ccttctggca	gcctgacact
1381	gcgagagctc	tggaaactgtg	gggaagggca	gagaagagag	gaagcctatt	tggaaactgaa
1441	tgttccttta	cctctgggta	ctcctggcaa	tgccatccat	ttccttcctt	aggaagacca
1501	tacaggctgc	ctccacttcc	tcagtttctt	cctctgtaaa	atgggcaggt	tgatgcctac
1561	ccagcaaggc	ttatttgaag	atggatttaa	acaaactgga	atgcagcct	ggcacatagt
1621	aggtgcccac	catgttcgtg	cttcataata	catcagcagc	aactggctct	gttctcacac
1681	ttgtttctgc	ctcccccat	tcaccacatt	tttcaccctt	ggcaatctgg	ctgccatcca
1741	cctcatcctg	ttgaagggaa	tccttgagct	gagcatgacc	agcccagtg	tttcttctct
1801	gacttgactc	tcctccacct	ccctgaagtc	ctcagcaatc	ttagctgggt	cctcgccctc
1861	ttgctccctg	ggcttcaaag	acatgggaac	gttctctctc	ccctattctg	atggatcttt
1921	ggctctggcc	gtagtgggtc	ctgtgtttga	cctttgtttt	ttttctctt	ccctctcaca
1981	ctggcatctt	caatgaagcc	atacacatcc	accgtatgta	ccctcacagt	ccaggagaag
2041	attctgcctt	gctccctcac	caggcttctt	tccagctttg	aagtgcgctc	atcctgtggc
2101	atcatcttgc	ctggatatcc	cactcacatc	tccaccgagc	ctgtgtaaac	agggatccag
2161	cattacaagc	cagctctttg	gtcttctttc	ccttccattt	ccccacctc	tggtcaggga
2221	ccaagtcttt	tcttttcctt	tcttttcttt	ccctccctcc	ctcactccca	cttgctggcc
2281	tgcctgcctg	cctgcctgcc	tgcctgcctt	ccttccctcc	ttccttccct	ccttccctcc
2341	ttccttccct	ccttccctcc	ttccttcttt	ccttccctcc	ttccttccct	ccttccctcc
2401	atccttccat	ctttcccttt	cttcttctct	tcttctctct	ctctggaggga	gccaggatta
2461	caggctcatg	ctaccacacc	cagctaattt	ttctattttt	agtagagagg	gggtttcacc
2521	atgttggcca	ggctgggtct	gaactcctga	cctcaaatga	tctgcctgcc	ttggcctccc
2581	aaagtgtcgg	gattacaggt	gatccactgt	gcctgggtct	caagtctttc	aaaatacctt
2641	ttagacatca	tcataatttt	acataatttt	tcttttttta	tttaaaattt	taaaataatt
2701	aagatggggg	ggaggatttc	aggtgaagac	cggaaatccc	atacagtgct	gaaaagggtg
2761	aaatcaccaa	gaaaacaggc	cacatgtaat	gccaccaccc	attgctatgt	cgataaacat
2821	gcttccaagg	aagttcatct	cagtattatt	tgttaattagc	aacaacaaca	aaagaaaaga
2881	agctaaatat	tgacattgca	gggttggcta	ttaataagat	gagacatgga	tgctgttgtg
2941	taactgggct	ttccataatg	ggtttccatt	gtgattggat	attccccctt	aagaccaaatt
3001	ctgaattgtt	ctgttctgct	ctagccccac	actttgaggg	agggtgttgac	gtctgtgtctg
3061	tatccaaagc	agagcagctg	ggaggacaag	aggctggaaa	tcatttcccta	ggagggaagc
3121	tcagagaaac	tgaggtgact	ggagccagat	aggacctgag	gggtgggcag	caagatcaag
3181	tgtctgaaag	gccacaggct	gcagatcaga	tcaggccagt	atagtcctcat	gacaggtaga

Figure 4 continued

```

3241 accagaactg atgcgagggc tgtgggaaat gaatttcage tcagtgagga aacattgtcc
3301 gtgctgcaag aatggagcag acctgccag ggccaccaga agaattcttc ttgacctgat
3361 cacatccttt atctggaacg gttcctgctg cttattttgtc cattaaagaa aactcaagcg
3421 cttagcctgg cattctaagc cctccctgac atgtgcggtc tcattccagg aatattttcc
3481 actcccccca gactccacct ggtgcatgtg ggccatgcc ccttccccag gcaggctctg
3541 cattctgctg ctctgagcct cagcaaattc ccttctctct gctgtgtacc actctctgct
3601 tctcatctcc cctcattcag tcttccccct cgcgaagtcc cagctcctcc atgatgcccg
3661 caatcagaag gccttgcccc ctcttctaata tcctagactg ccttatgggt acctctccca
3721 ggacatggcc acttccttcc cagctgtage ccagggtgta tatgtgaccc ttcctataa
3781 tggtcagtgc ttgtgagaat atgggacacc cctttttcat cctctcatcc agcatgcggc
3841 tcagtgccag gattctaata gataaatgtt tcagagact tctaaggggg aagctaaatg
3901 tcttgttctt tcctagtagc tctccttctt gcatttattt ttagctggat gtttttatgc
3961 ctccaattct agtttgccca ttaaaccagc tcaatgggtt agggaggaca ttgatcgta
4021 tccctattgt acatcaagag aaactgaggg ctagaggggt taggtgactt atttaaggtc
4081 actcacttag aaagcggcaa actccacctg gaatctgggt ccagcctttt gcctctgatg
4141 catcctgatt tattctccat gtccagcagg gcatccgtcc tcgccacctg tgggtggcac
4201 cgtgcatggc aaagtgtgtg ggaagtctgt cagcttagaa ggatttgcac agcctgtggc
4261 cattttcctg ggaatccctt ttgccaaagg gcctcttggg cccctgaggt ttactccacc
4321 gcagcctgca gaaccatgga gctttgtgaa gaatgccacc tcgtaccctc ctatgtaagc
4381 tgcgcatgt gtcttgggg atgtttacct caaagtgatg caggaaggag tcaaggcagt
4441 cccctgatgg gctgatcctt tgctctggac tccttaagat cttttagat ccttaagaac
4501 attccagaac tctcacagca ttctggagtc cattatttaa cacatgttta ttgagcacct
4561 gctgtgtgcc aggcgtgttc tgggttcttg ggatccatta gtgaaaagc aaagatccct
4621 aggttcctgg agcgtgcatt ctagcagggg gagacagata agaacaatat gcaaattact
4681 ccgtggatta tgtgatatga tagaacagga aaagtgcctt ggaaacaaag aagcagttaa
4741 gcaggtgagg gagatcagaa ggaaggatcc catgaattct gctgtcttga attttgacat
4801 agtcggggag aggagtgggc caggaggtgt gtgggtggca agtgtactga ggtgtccact
4861 cggcaagtgc tcaactctca ggtgtgccta ttctgagga gaacaatgga aggtgtgtcc
4921 attcacctg gccaaagtgg gaagaaaagc ccaaagggtc taattggcct caccatcact
4981 cccagggtg ggaaggacct tacagacatc agagcagacc ctgctcatct cagcagccag
5041 gtgttcggag gctttccagc agcccaccag tggcgcctt cttttcttcc cttttttaaa
5101 aaataatggt tactgttttt attttctgag tgcaaagtga atatatactt attacagaaa
5161 atttagggaa tacagagaaa ggaaagtaag aaaattaaga tcacatacaa tattccacca
5221 cccagagaga acaactatca gtattttaga tggtaaagta ttatgatgta atatattggt
5281 tgtattccag tttttttgcc tgttgatata ctttctttct ttaaaaggga atgctatatt
5341 gagataatth ggatgcctgc tttgtctatt tataagtata tcatgagcat aattccatgc
5401 cattaaatag cacagcctgg tattaanaagc tgcattacat tgggtattga accataatta
5461 atttcccaa gctttttattt ggggtcatta agttatatat gatgttttta ataataaac
5521 atcatgaatg tgaaacattt atgcacattc atgaatagtt taggataact tctcagaaat
5581 gacatttctg agtcaaagag tatatatthc ttgtttgatt tccctcccaa aatattccaa
5641 ctacactcct cccattgacg tatgaacatc tccatttggc agactcttgg caatgcttgg
5701 gattattaaa aatatttttt gctattgtat aggcaaaata ataataacaa taccagcaat
5761 attattttcc tctttaattt gcatttatct gaagaccagt gagttctttt ttcctttctt
5821 cacattggaa tgtattatta ttattattat tattattatt attattgtta ttcctcttgc
5881 acttgggtgat cttaggaaaa tttcccagga tggaaatcaag aagtcactct tgcctgctct
5941 gattctgtca cttgagagct gtgtcatccc agagtctgag tgccctcagt tccctcatgta
6001 cctagtgggg gcggtgatga ctgccttttt gtacattgag gtctgggact gttgcttata
6061 atatacttgt gtataagcac cacatggcat ggatctacct ggctatgtaa ccttggtcag
6121 gttctctaac ctgcctaggg ttcagttttc tcatctgtaa aatggagata atagaaccta
6181 cctaagggga ctagtctgag gcttaaatga gttaatatat aaaaggfact taaaacagtg
6241 cctggaacac tgggagaact aaataactat taccagtgtt catctatgtt gcatccttga
6301 gatctttcat cagggaacaa caacagcaac tgccctgggag aggcagggtg tagaatcttg
6361 gcatgtcagg gctgcaagac tcatgtagaa atcaactccg tcacttagac ctggcggtt
6421 gggacgcaga gtgtaaaagg aattcaccca aggtcacccc ataaatgcat gggagagaca
6481 aaactagagc acaattctcc tgatgttgtc ccacaaagat atccccctg agctattgtg
6541 agaaatgtca cctcccaggg aggatcagtg tattagtggg ctccccctct tgatgctggg
6601 agttccaaag gctctggaaa gggaagaggt gtgaagccct tctcactctg cttgggtctta

```

Figure 4 continued

```

6661 ggagacctta gtgagtccca ggcggcccccac ctgaagcccc tgatagcctc ctacccacta
6721 caatgtcgtg agtctgtaac tatcaagtc atttccaggc ttaaaccctcc cagtccaagg
6781 cgctgccgta gaaaacacct cgcttagatc tggctgaact tcaggggttc ttctttcccc
6841 ctccccaggt gcacccaaga tcccaaggcg gggcagttac tctcagagct atttacaac
6901 cgaaaggaga acattcctct caagctttct gaagactgtc tttacctcaa tatttact
6961 cctgctgact tgaccaagaa aaacaggctg cgggtaagtt gtgggacccc ctgggtcaagg
7021 cgtgtcagtg ccagtacccc gcatcttctg gggcagaatg gaagggagtg aaggcagctt
7081 ggggagggag ctgcacaggc caggattgtt tcagagaccc cagggcctcc acaggaaaca
7141 ctggttttcc acacatcagt ttacctcgt gacataggca ctctggggga gtttgttttg
7201 catctttaat gaacatctta attgttccaa gttccctcat gctaatagaa ggatgagaat
7261 gattactttt agttactaga ttaaaaatct gataccaaga ttgggtaaga catgattttc
7321 aatgttagtg gtggaaagaa cttagagaaa agccaacacc ccaactcttc cttttcaaat
7381 agagaaactg agggccagac aggggaagggg aggtcaccga gtagggtagt ttgggtcaa
7441 gtcttctgat cttcacatca gtgcttttga catggagtag ggcattgagc ttgtatgtgg
7501 cctttctcca gcacagcaac attcatctgg ggttaactttc actgtccagg tacggatggt
7561 gggggcggag cctggacaat cccagaagct cttccttttg cctttggcca aatagctggg
7621 atttcttttt gtacctaat tccctcccta attgtcttgg caaaggctga tgacccagtg
7681 agtcagaatt tcccaaataa atgaagggca ggacatgggg cagcccgcga gctgcagggt
7741 ggaggggtgg gaggggtctgc tggcaaagga ggcacccagg gtgaaaaaac aagggtgtt
7801 gtgtctagcc agtctcccag gggcacctg gcatgggtgc taggagtggt ggagcaggtc
7861 tgcgaacgtg gaactagaag ttgggtctac atgtgctttg tgagacacag gaagtgaag
7921 aaagggtttac tcattcattc attcattcat tcattcattc attcattctt ctattgtgta
7981 ctcagtgtcc catatacaca gttagacctg gggcttcagg gaaaggagga acatgaatcc
8041 actccacaca gcccctgcct tcaaggatct tgccttctgg gctgtagaga catatcctcc
8101 cttcagcact ttattgcagg ccagccctca ggctcagttc tcacacaaag atgaatgcac
8161 actcaatccc tccccaggag gcaggcaggg attagagttg tgggagaaac aagggtctg
8221 ggaacctgga gtcaggagaa cctccaggaa tggccacagg ctctagaact acacggtctg
8281 ggttttaaate ctggttctgc ttttgactag ctgagagact tgggagtttc ttaacttctc
8341 taagcctctg tttcctcaaa atgggaatca tcacatatga tcagtcaagg gtgaccacta
8401 agcgtgatct cagtccttgg ctaaaaagggc agaggcaggc tcataccctt gccctaacag
8461 tggactcact gcactgtggc cggcgcctgc agtgtgctgg gcagctcaac atgtggactc
8521 tggaggcaga ctgcctgggt taaagctctg ctgctaccac ttttgagttg taggaccttg
8581 ggcaggttaa gtgcctccat tttccacct ggataatggg acattattag tacctacca
8641 atgtatagtt tattttactg tgcttagaac agcacttggg acatagttta tgttacgagt
8701 tcttttatat ttcataaata atattattac aatatcatta ttttagccac tctcacttct
8761 tccccactat gtggactttt aaaatgcaag aatcatgcat gagggaattg atccactctc
8821 atcctccagt attagcagag tccctggcac atgctaagtt ctgcacagtg ttagtccatt
8881 gactgaaggc atcctgaagt ggggtgggtt taggctgggt tctgcagggt ggataggatt
8941 tggggatgca gaggcataag gaggcatttg aggaagagac accatcacag gcacgtgtgc
9001 ttctcgagac aatcattcat gcacagagaa gtttacttca ccaagagata gtttaactgag
9061 atgaagctga gatacaaggg ctgccccatg ggtacgctga gtgcataaat agtctaggct
9121 tgaggggtgat gggagtgtcc tcccgaagag gacatcctta gcctttgcta tgcccatgat
9181 gatgttctca gcatcaagag cctttcgggg aggaggcact ggaggctggg tagaaaggaa
9241 gggacagcca ggggtatgtgc agtaggggtg tggcttgggt ccagtttggg agaattatgc
9301 aggagagaga ttgccttttg caaagttgct gggagctggg gaactttgga tagcatcctc
9361 cttcagttgt aaccaacagg taggtcacca aacaagacat cttctgagtg ctccctggct
9421 ctgtgtctg taactgggtat gttgcttctt ctctgccag gtgatgggtg ggatccacgg
9481 aggggggctg atgggtgggt cggcataaac ctatgatggg ctggcccttg ctgccatga
9541 aaacgtgggt gtgggtacca ttcaatatcg cctgggcac tggggattct tcaggtaaga
9601 aatcgactc tctcactgc actttggccc ccagaacgag gatgctagga ccagctctg
9661 gtcatgccag cctcagggg agcttagcta ggttccacag taaggcatcc aagccccttc
9721 gtaattggac actacctacc ctctcactac ccagccactc atccacttgc ctctgagctt
9781 ttgcatgtgc tgttccctct gcctggaatg cttattctat ccgaagaaca cctcttcac
9841 ctgcaagacc cagccccag ttacctccac tgaaagactc atctcctctt cacctatgtt
9901 tctccagcac attgcttttc tctgtcatga cacttagcag tcagcccagc aggtgaagggt
9961 tgacacatct gacttctcat ggaaaggagg gaaggggcag agtcacagag ggagctcagg
10021 gcatgttgtg ggtgaaggaa tgggggctgc agatgggtgg gtgtgggagc atctgacccc

```

Figure 4 continued

```

10081 ttccctgtggg agtggttggga cccactcagc tcattttctg cttgaaatth ggcaagtgtg
10141 ggaagcaagg atggagtctt cccatcaggt ccccatgggg acaggagtca gatctccttc
10201 aaggcgcaac tgctatgaac cttttctaa gttctcctcc atccatgcc a gtcagggtttc
10261 ttccctgaaca cgcactcag ttggtctcat tgcggtcac aatgacctcc cctgctctc
10321 caccacacct ctgaggacct cccagcagca gcagatggcc actcagcctc tgtgacatgc
10381 ttccctccca gccctttcc tggctcctcc tccctcctgt ccacaggcct ctcttgtctc
10441 cctctccctc cttactaatt cccagtgate ccaggaggca gccagagccc ctggttccag
10501 cccacctcac ctttgaagcc cacaggcatc actcactccc tccctcccac ccttggtgtg
10561 ctaaagggtg cccactctt cccagggaca agttgaatcc ttgactccca ctgcccctcc
10621 ccaacagaga ccccttctc ctcccaccac gcctccactc agcttctgct gccacatcct
10681 cccttgggct caggagaagg ccttgtggac agcccggact gtctcttttc tcacaacctc
10741 cttggtctgc ccttcagaat cccactcagag tccctactt tgcacctctt ccttggtctc
10801 tagccagccc aaagctctct gtctgtcacc ttgattatga tctttaacac tttcagatcc
10861 ctgcccacttt tatataaagc tgagttgcac agtagccaga gtgactcact aacacatgca
10921 ttgaatcatg cgatgcctct gcacaaaacc ctgcaggggc tcccatgct ccttaccatg
10981 acagccgaag ttcttgccag gccccagggc cctctctagc tgcacctcct gactcacact
11041 ttctttcact gggccccagg ccttttgca ttgctttttc ttctgctgg aattcctttc
11101 tttgatatt tacaaggttc atgtgctcat tgtgcttgt ttatgaaggc ctttgctctc
11161 aaaacctcaa atgtcactgc ctccaaaaag ccttccctga ttgggatctg aaatggcacc
11221 cctgtccca ctctgtttt tttccttact ccattgtatt tacctcagt g cactttgact
11281 acctgacatc attatattct ctctatgtat ctgtttatct gcttgtgatc tgtagtggga
11341 attcagtgtc atcacagggt ttgctgttt tgcctcaga tgtgctctag ctcataagac
11401 agcgtcaggg atacagtaga tgctcagtaa atagttgcca gttgtgtgaa tgtagaacca
11461 tacatcacca caatgctgta ctaatgaaag cagggtctca gagatgctga agcccagcct
11521 tattcttaag ggttcaactga gaacccttag ccccatctgt ggtcctgaag gtcctgcactg
11581 acatctctgc tccccacct caacctgttc tcttccctac agcacagggg atgaacacag
11641 ccgggggaac tggggtcacc tggaccagg gtgctgccc cgtggtgctc aggcacacat
11701 tgccagcttt ggagggaacc caggctctgt gacctcttt ggagagtcag cgggaggaga
11761 aagtgtctct gttcttgtga gttttcctgt caccaggccc aaccccacgc ttgatgtgat
11821 gctgatgaga cctcttggac acctgtcaat ccagctatgg atacctctga ttacaatacc
11881 tgaattcagg actggcccta ctcacagccc agcatcagg ggtagaaagc taaagaccag
11941 ctctccttgg ccccaggaa gctcagagct cagcttagtg tccggggcc atatgaatct
12001 ccagttgcag cctgcaatgg tggcattctc ctctccagc ttggttgagc tctctctctc
12061 tctctctctt tctctctctc cttcacactc attttaacat aaggactcac atcccttctc
12121 ttgggaagta ggaataggca taaattatga gtaagggcag gcagagagaa gaggtggaca
12181 gaggtcattg tttggctaaa ccagacctac atgtggagg gacatggaca ctgagcaggc
12241 tgggaattcc tctggggtgg tctgatggct tgtccatgcc caagaaggag gcgacagtct
12301 ttgtgactgt ggggtccagt gggctagggg aggcaggcag aggaaggagg gatggagccg
12361 cggataggga gggatggggc aggggttgtg ggtcatagac acaaccttgg gtgtgagggg
12421 tccctgctggg attggggatt ggggtcagga gacactgggg gatctgggat gaaaaccag
12481 atgagagggt ctgggagctg taggaagact tccacctcct tgagggtggc agagggtcag
12541 cccactactg gattcctcag tccctgttg gttttatagt ggagtagatc tagcctggaa
12601 tagcgagtga gtcactgacc ccaactcctga gcatgaactc tccctcccct cactctgctg
12661 tcagggtttt tctccatttg ccaagaacct cttccacagg gccatttctg agagtggcgt
12721 ggccctcact tctgttctgg tgaagaaagg tgatgtcaag cccttggctg aggtaggctc
12781 ccggttggtc cgtctccggc tggacacccc cactccttg gctctatgct cctgaatcct
12841 cagggatctc tcttgtggtc ggttgtagct aatgttctcc tagaatcact gaggcacca
12901 tggctgagca ggaagggcga ggagacacct tgatcagcgt ccagtttca cagccaggca
12961 aaccgacaca gggcttgga catggcccta gcagccagta cagtgccta tctgtgacca cactcctcct
13021 gactccagct cctgggtgcc atgttgggca gtgatgggtg cttgtgtctc tctgggtctg
13081 cctaattggc ggttaagtga acaaccagga tttgaaagca gacaaggaaa ttcaggattc
13141 catgctctct atcccagctc cactcacat ttcttgacac attcacttgg agatgattat
13201 gtccatttca atggagataa ggtagcagag atgagagatt acataattga cccatgttac
13261 aattgagatt agtaacagag gcaacgctcc gtgggacctg gaaccacac ccacgggtct
13321 acagtatctt ctgctgctcc ctgccactag tacaagttgg gcttgggata gaatgccact
13381 ttccctcttt atggaggga gggatgtcgc tcttgaactc tgttgttgcc tgtgatcttt

```

Figure 4 continued

```

13501 gcagcaaatt gctatcactg ctgggtgcaa aaccaccacc tctgctgtca tgggttcaactg
13561 cctgcgacag aagacggaag aggagctctt ggagacgaca ttgaaaatgg taggttgccct
13621 gttcccgtag cccaaaccct gtaaaacttg tcccagactt cttcattttca gctgtcctct
13681 tgcccaggga cagtttcctg ggacaatttc tcaactctca gtatctgaat ggggaatctg
13741 atttgtcctt ttttattgta aaatgccaca aatgtaaaaa aagaaagtca aaaaataaca
13801 tgataaaaaa ttgtgtaaca atccccagtc ctgaccaaaa tttaatatth tgccattctt
13861 gtttttcagat gtatttttaa gaaattaaat gttacatatt cccaggggac accatcatcc
13921 tgaatttggt ggatggagat attaaagtca aagattttca gaaagatgtc acaattttatc
13981 ttggttgact tagaaactgt ctgtattaga cctagtgggtg gtccagtttt ctagattttc
14041 tgagactgac tcagttgtca tcctaggata gtcateccatc caccatttcg ttaatccgtc
14101 tatctatgat tgtcttatcc atctatctgt tcaactaattc atccaattca ctcatatctt
14161 tttatcaatc taatgtcaac ctattcataa ctttgtattt atctattttc aatccattca
14221 ccattccatg atcatccagc caccctatat tccaactcca tcaccattc cccaaaaatc
14281 aacaatccaa ttatcgctg tctgtagtgt ttcaaccatc tttcatgta tccattcaat
14341 tcaactgtac tccatgtatt gaccaactcc atccatccct ccattgatcc atccatccat
14401 ccatgctaaa tatgtagggg tgggtgttag aggtagcaaa acagacatga agtggacata
14461 gtccctgctc tcaaggaact atccaaagag aaatacattc atatacttcg cagggttgagt
14521 atggtggcag cacagagggg aagcattcat tgtagtatac agggatgctt cacagagaag
14581 gcggaggagc cagatttgag accagacagt ggaattcagg agttcatgcc cttaatcctg
14641 actccacctt atctttcctg agaaattcag ctttcagatc attgtgccc ttttaacaga
14701 ggtaacaaag acagataaat tttgtaattg ccccatgtca cagttcaaat tagtcacaga
14761 ggcaacactg taggaccag aaccgacacc tgcgggtcta gagtatcttc tgccttcctt
14821 gctccctgcc atcctcacat attaggactt ggggacctt aggattgttg gatggccatg
14881 gcagtctctc atcacaggga agaccaggag gacaaacacc cagatgacac agcaccagg
14941 gccattggga actattccct ttgaggggga agagtgtgaa ttccaagctc aggagtagcc
15001 tggactctct ccttggaact gaggetactc ctggatccca gggetggcct ctaccactgg
15061 actctgcttg tgtttccatg agttgagtc aatgtggtat tgggtgcttag gtcttggtc
15121 aggetctaag tgaccaaagt gttcaaggaa tgaaaacaca actgttttct aacgcccga
15181 aggaggcaaa tattccagca attctgcatg tggcatcaga ggaccagct taaaaggga
15241 gagttgagtc tttgggaaaa cccatcccta agatcctaga acattctttt gagtttttct
15301 gagatcttgt ggaagcacct ccatgattac gcgtgcctc cagtacacac acatgcagac
15361 acacagacaa acacacacac acacacacac aaagagccac acacagacac actcatacac
15421 acacagagac acacacagac acccacacac aaagagccac acacagacac actcatacac
15481 acacacacac acactcctgg ctagtgggac ttcaatcctt aatggaggga cctgcactgg
15541 ttacatcctg agcacaggtc agatctggga actacagtgc aacactacag tcaactgcaat
15601 cttggtgaac acacacaaa gagaagcatg gggtaggatg agactcagtg agcttgctg
15661 agtcagggtt ttctaattgg agatgaggaa actctcccaa gtacctgaa cagaaaaaga
15721 aatgtgttgc aatagcatcg tcgtgtaaac tgaaaatctc aggagttgat ggcttcaggc
15781 atggctgggt ccagatgatc ataggatatt attgagaatc taccattctc tgccttcag
15841 ctttgacat ctctgtgttg gcttcaactc taggcagata tatctcctgg gaaggttaa
15901 gacaacagca ctctagactt atgtcctatc tgcttaaaaa cctcttttct aaacaattcc
15961 tgcaaaagtt ttggggtaaa ctctattgga ttgacatgag tgtatgccc tccctgaact
16021 aatattcatg ttcagggaga tggagcaccc caggtcacat atagatatac gaattcacgg
16081 agtgatgcgg gaagaacctg acacctctgt tgcaccaactc accagctca gtgttctct
16141 ggtaagcctc tcaccacat cctctgcctt tgtcttcaca gaaattctta tctctggact
16201 tacagggaga cccagagag gtaaggacct tttgtttctg gattacgggt tttgagttt
16261 agcacttta agctccaatt aactgtgagt gaaagaatca tctcttgggt aattatagta
16321 actcctgcgt gtttgtttct gaggccaga gagggttagt gactcacctg ggtaacacag
16381 ccaggaagac tagtggctgg cctggaacct atttctctga ctccagtc agtgctcca
16441 ggcacacct ctgtgtgcc tgggtctgc cactccct tcttttttac atttctgct
16501 cccaaaatg gcaactgagg aggaggtga aacagaacct tccacaatca ggaggcagag
16561 ataacagcga tgattagcca aggagagggg aagcctaaat ctcaagtc aa ggacactcgt
16621 ctctccag cacacaggaa actccaacag tattctccca atctctcat cccactccc
16681 accccacct cccagtgggt tgacagtttc tgggtgacatc acctctgacg aatcttaca
16741 tctgtctctc tctgtgctt cctggagat cacagcactc tctaaatgg tcatgggcgg
16801 agtacatgag attcattcag caaagactta ataaacactt cctatgtgcc aggacttggt
16861 taggagctgg ggatataaca gtgaaaaaag aaagactccc taccctcttg ggacctcca

```

Figure 4 continued

```

16921 ttacagtgcc aggacagtgc aataaaaaag caaatcaagt atgtattccg tttgatgggtg
16981 tgcagcatta ggggatagag tttgtgggtg gcttcaattt aagtagcagg atcagggaat
17041 ggttcaatga gaaggtggca tttgagccaa gctctggagg aggcaggaag ccagctgtca
17101 ggaaacctgg agaagcctat tccaggcaga gggaaacagtc actgcaaaga ccctgacaga
17161 agggggccagt ctggctggag aggaccgagt gtggggacag actggcttga ggctaaagag
17221 gtaatgggca gtgggagaga gatcaattaa agtacttgga tttggattca gagcaagatg
17281 ggaagccttt ggaggttttg aacagaggag tcacatgatc ttagatctca caaagggtctc
17341 tgtctctggt gttcaaatag actgtaggaa aggggcaaag tggatgcagt tgaccagctg
17401 ggcagccact gcattgccat aatccaggca agggctcctg gctgcttaga cagggtctgtg
17461 gcagtgggga tgggaggagt agttggagtc tggatatctt tgaaggaata gctgacagga
17521 tttgctgatg gacaggatgc aaggggtaag aaacggggag gagtggagga cgccctgagt
17581 tttctgacac aacaagggtg tgcccacaca gggcagccac ttgcaaactc taagtggaga
17641 acaatcactt gtcatgtttg ctgatggcag tcttcacaga tggctgtcac ccaactgagg
17701 tgctcccgtc tcttgccctc acttttagcc tcttctactt cacttctctc cccatttcc
17761 atgtctcatg gtgcagccag actggccttc cttaacagaa ttagagcacg tcacccccac
17821 ttcttaaaag cctgaccttc cttatgagac aaatcccctt cctccatgg agttagtctc
17881 tcaaaacaca cctgcgtgca cactgtgtgc cagactgtgc agaaggtagg ctctgcacct
17941 gtccctccct ctgaccctg cccacagca tttctccaag ccaagaactc ttggttgttt
18001 ctggaattta ccctgcacag tgacacctat gtcccttttg tcatgcggtt cccactgttt
18061 agaaattctc actccttttc tcacctggca ataccattca ctgccagag cccaccgcag
18121 gcaggcgggc tctcttccac gtagtctgcc ctgcatcttc cttccagagg ggctgctcct
18181 cttgctggcc tcccacagca gccctgectg aactgcacag cctcttaagg aggctggact
18241 tgcaccccaa tcctgttctt cattgtcagc catattggat taggatctga tcttgtccag
18301 ctaggtctgg tccacttttc agaacatctc cccaggcag tccgaactag ctcgtcattc
18361 attggctttt gtagagcaga ccaaagggtc cagaagccat aggggtctcaa aggctaggaa
18421 ttgtccagtt ccactctgata tccagaaggg aaatacagcc catggggcgt ggaactggga
18481 agattccaca gaggactggt gtctatgaga gggaccacac gcaaccttgc taacaagtaa
18541 tacctaacag ttattcagcc tttccagaca ccagggtgctg tgctaagtat tttacatgtg
18601 ttcagccatt taatcctcac aataactcac ctcttagatg agaaaactga ggcccaaaca
18661 ggtgaaactc ccagccagta agtagcagag ccaggcttca gatgcagtta atctggtttc
18721 aaagtccatg ctctatctta gcaacatcca cgctctatcg tagcaactac accattgtga
18781 ctaaatatga aatataactg taccagaac cagctgcctt gatggcaaca catgagttgg
18841 gctctctcta atctgtgacc atataaaaaat tattcatcaa aggtgaaacc taaaattaag
18901 acatggatca atatactgtg agttaatcac tgattctttt actcatgatt ttctctctag
18961 tagggaatca tgtcctagtc taggtcctt gagtgaaccag ggttcggtac ctctcaaag
19021 ccaccctggg atcaataaca gctcttggtt aaatacagat agatagcata atctcttctc
19081 cattaatcat ggattctgtg tttataaatt cacctgcttg ctaaaatgtc cttataacct
19141 ccaaatcaat actggtggca ctttcatggt cacacacagg caggtacaga gcaggggaag
19201 ctggatttgc atgatgggca tgttccatct gagatccggc gaggcaacct ctctcttctt
19261 gtttcatctc tgatgtttac aagtgtcctc ttcgcggtct atttagtgcc gtgtgttctg
19321 catttttgtg ctttttgtag tgatgtcact gtttagagcg aaccaagag tagcgctgct
19381 gtctggcggt cctgagctaa acaggctgtg ctgtgcctta cggggaagggt gcctgtgtga
19441 gacaagcttg gatgaggcag gagctgcagt gctgctggcc gtgcatttgg tgttgaaaaa
19501 tccacataat agcacattca gaaaaaggag aaggaaattg gccgattcat acatgaggct
19561 gtgatagaaa gggctaaagt aacacctatt gtgtgggatg gagcctgga agtctctttc
19621 aacagaaacc cacacataaa agaaggttga tgaaaatgtt gtggccagag gtttgcaaaa
19681 acctaaccac gtatttcccc cttgagtgat gacctgtat tctctaattc agtactcacg
19741 ggggctttat cagaagcaac tgccaggact accgaaaatc aattgtggac acacagagga
19801 aaatgcatgc acttcaaatt gaaatatata gacatgtaat tttacataga taatacatgt
19861 atttcaatat gagtgaagat ttcttctctc cctcatcacc tgagcagtggt cagcctctgc
19921 ctgaagttct gccctgagt catggagacc taccctccta agttaggatc ctgagcatgt
19981 gcagccatgg cgcattggca tgccggtcta tggtaactgg ctacacctca gacaggcaga
20041 gttggggaca tgggtgtgac tagggaggga atcacagggt ctgcaactgt ctctcccag
20101 agtcaacccc ttctgggcac tgtgattgat gggatgctgc tgctgaaaac acctgaagag
20161 cttcaagctg aaaggaattt ccacactgtc cctacatgg tcggaattaa caagcaggag
20221 tttggctggt tgattccaat ggtgagaagg cacatgtctg ttggaggggac tcaaccaccc
20281 ccatcagccc tcctgtctct ggtcccagga atgcttccct ctccagctg cactctgagt

```

Figure 4 continued

```

20341 cctgggcagc tgttcccttc agcaggagtt aacatttcca cggaaatctt ctccaggagg
20401 gcacagtttt caccggggta tcagggccct gggattccct ctcccagac tctcttattg
20461 tcttcctatt gaagaatcct gaagtgtctg tgctgggagg gccagtagag aaaataactc
20521 tgcagatggg aaaacggagg gggcctagaa gggggaagga gtggaggagt cgggagaagc
20581 agtcgggggc ttcttggtt cccacctcgg gctgggtgtg cgtaatccc ctgaaaatca
20641 ctcatgctct tcatcacctc attgaaaata tccaaagaaa gtggcttctt cttgagttac
20701 cattgacttc tagggaacag cagacatcag cgtctactaa gagagtggat ggtcagagga
20761 agatgaggat ggaaaagcta cctaatcggg tgctatgctc aatacctggt aacaaaataa
20821 tctatacagc aaactccaga gacccaaatt acccatgata aaacctgcac agttaccccc
20881 aaacctaaaa taaaagttag aaacaaataa atggaaaaag taattgagtt ctaaaagtaa
20941 agaagtgttc tccttccagt tgaagataaa caaaactgac cctggttaaa gcagtaggac
21001 agatttattc agtaatatga ttgcaatagg ggaaatagtc cagcgtggag tggactcaca
21061 tccctgaaac aaaagcctga tgacgtttca taggccaggt gtgccaaggg aaacactgtg
21121 gtctatgggg agaggcttgg ttcattgtac tagaccact ggatgtcttc atcctggctc
21181 acctggggca gaaacaaact cctcttatct ttaggacagg aagccatgca ttagaccgca
21241 ggaagaaccc actgaagtca ggctctactc tttcgacagg gactgggaag atcaggagtg
21301 agctccctga atgtttgcat tgcaaagaga aggtctcag gtcccaagg aaatgggggt
21361 ggggtgtaga ttcacatccc aaagggcaga gaaagtgttt atcattgcag gcttctaaag
21421 taactgctct agttgggttc ggggacctgg ctgcttgtct ccaggttttg gctggagcag
21481 agtaaattgt cctggcaggg ttgagccttc ccaggcagtt atgttaagga tactggggtc
21541 attctagggg cacagcctta agctgctaga agcaatgcta gagattgggtc gagtccctga
21601 gtgcagaggt ttgggaagag ttgttatgtg cgtagagatg tgcagttcgc cctcagaag
21661 gatttcagag gttataccaa agaccaaagc cacgaagaca ggatttaagg atttcagtgt
21721 catctgtgac aggtgggggt actaggaggg tctacatcct ctccccaga gccctctatt
21781 ccacactggc ctggcagaca ccttgcatct ccattgtcact gacctgta gggatcccag
21841 ccacagacac acacacacac acgcgcgcac acacacacac tgcaaacac tgccacagct
21901 tctctgaaat tttgtttttt gttttttttt tttgtaaact ttgcaatgct tgagtttctg
21961 gatataagcc catttgattc attgatttca cccatttcag taaagactca tgcccttaa
22021 agccccata attagaggac tttcagactg cattggtttg gttggttgg cagtttgttt
22081 cttatcatta attcccaatg attcataaat gctgaacttt tttttttttt ttagcagttg
22141 atgagctatc cactctccga agggcaactg gaccagaaga cagccatgtc actcctgtgg
22201 aagtcctatc cccttgttgt aagagtctag gaatcatggg aattggctga gaccagaga
22261 gggacaagga cttgccc aaa tcatagagca attaaatggc agaatgaaga ctggggcctc
22321 agggtttccc ccatctttcc acctttccca cttcattttc caccctagcg gggagttgca
22381 cagggttat ggggttcacc attgaggcag gactttctgg tgggctggag aagctgcac
22441 gctcaccggg ggctgggtgt cactttttga tctatttcag tgcattgcta aggaactgat
22501 tccagaagcc actgagaaat acttaggagg aacagacgac actgtcaaaa agaaagacct
22561 gttcctggac ttgatagcag atgtgatgtt tgggtgtcca tctgtgattg tggcccgga
22621 ccacagaggt gagtccaga ggtcgaacgg gagggacaca aacccacga gcttctgtat
22681 ctgaccacc tacctcccag catagacaga catggaaaca gctgaggggtc aggcctcaa
22741 ggctgattcc atatggcatg ggatgaggtg ctgtcttatt ttggtttatg ccagcagaag
22801 actgtgagaa aaaaaaatcc agaggcaagt ggttttttta aggtgatgat ccagataac
22861 agcagtaggg aggtggggaa gtgagagagg aaagagatgg aatccaatcc aagatgcgtt
22921 gtcaaggagg tatccactgt ggacaactgg agcggggaaa ctcaaggaaa caccaagaat
22981 acagggtca ggggcatccc atctgagtgg caaggagacc gggatattat tcaccagctt
23041 ctacttgtca ttggttgaag gctcttgcaa ggaattgttt attccctgtg acttcgacct
23101 gctgcacaca agggcagagt agtctcttgt gaccagacaa aggcctcagg cctggagctg
23161 cagatgctgg aggtggaagt caggctggca tgcccaggaa gaggggatat aggtgagacc
23221 ctggcagcat ctgctgcaag ggctccatgc ctggagtatt tgtgggacat agggctgctg
23281 taggacaaat attgaaagat gagtctgtga agggttttta ggtctgagcc agctgcaaac
23341 aagtttgtta agtattggag ctggataaaa agtgggagcc ctgcatggga tccagaggat
23401 tgtccaggac caagaaggcc agagagagga gccaggggcg gtgtgggaga ggtgggtctg
23461 agaatggcag tgtctgaggg ttggtttttat gcatctcagt gcgtgcagta gcatgcaagc
23521 acagggcaag ggcagggtct atgcaggacc tgctaccact gacccaaggc tgtgtgggtg
23581 gggcatgacg ttagggtatgt gtgtgctcac ctagagagga gtgtctgttc ctgctgcca
23641 gctagaggag attcacagga ttccttttct tccagctctc attcgcctag tctgcaagct
23701 aggaaatgtc ctctctctaa ccatgctgga agggttttta aatttagaac ttcataagca

```

Figure 4 continued

```

23761 tgataaaaac tagctaacgc ttaactagca ggtgcoctga cacacctttg cacaggaagg
23821 ggcaggtgct cataaccctc atttcttcac agggtcctga taaaaacctt taaccaagac
23881 cactggagtt gaatctgttc tttcttaaac ttgcccttgc cctatgctct gcgtctgaac
23941 tatatataga gttcccatct ccattttctg atggtttgag caatactgaa cttatgtttt
24001 tcatctgggt tcccatgact ttaagttcag ggcttttttt tttttttttt tttttgccag
24061 gactacacag attttatcta ctttcactta ttgttgtctc actctttttg ataaaaatac
24121 atttctgaat ataaaaattaa tatgtttaa atagaaact ttcaaaaata tggaaacaca
24181 aaaagaagaa actatgtatc actcatttcc actacctgga ggcaagtgtc ataaattttt
24241 tgtgttttcg tctaaagatt ttccatgtcc tatatacttt aaaaatcaat tttatatattt
24301 acttttgttc atagaacact atgttgtgag gattttctca atgttgatac aaactcatca
24361 aagcattcct tattccatga aatggctgtg ttaaatagac tttaatattc atatcctatt
24421 agatgtttag cctggctact cacattctga tgctacaata acattgagct cagcctaata
24481 ttcataagtc tctaaacaca tgtctaataa tttcttagg attattttta gaagtaaata
24541 actagttgg acaatgcaaa tgttttagaa ctttgggcaa atattgtaaa attgcattcc
24601 actgcataag aatgtggatt tcttaattct ctttctcatt tgggatgcta agattaaaaa
24661 ttagcataac acttccagtt tcataagcct tccccaccaa caagaagcac atcagcctgt
24721 tttaatctgc atgtcctcca tgactagccc tctcattgtc tctgcttctc ttattatgtc
24781 taattgtcat tacccaaata tgtgcgtac tttcctctga agacttgtgt tgtgattgta
24841 agaattgtat aaatatcaat ggcttttagc aaagagtttt acaaagtacg ttttgggcag
24901 catgctgcct ataataagtt ctatcattca caaaagctct tcttctatg gttggctctg
24961 agtttctttg gaaaatgttc gaggggcaaa ggcaagggtc aacctcctt ttagcaagtt
25021 ttgtttgacc ttgaagcagc catttaacag gtggattaca gagccacaga aggatggcat
25081 cttcccaggt gcgcttcctg ctgagccgag ggtgcagtgc aggaggtgca caaagatgag
25141 tgaatagatt attcttcatt ctcacattga aatgacatgt gcgggtgggg catgagggtta
25201 gggatgcgtg agctccatt gttaaactct tttgtgaacg aattttta at gatgttaaga
25261 aatgatggct ggaggcgtg gctcttgcca acactctcag cattttggga ggctgagggtg
25321 ggaggattat gagagcctag gatttcaaga tcagcctggg cagcatagtg agacctgtc
25381 tctacaaaac aattttaaag ggaagtgatt acagtgtact acaaaatacc atatatataa
25441 tcatccccag ttaaatacat aaaagagaga tatgctaaaa taaaaagaa gcattctctg
25501 agtggtagag ttatgagtga tactttaaat ttgattgtta aaattgattt ttaaattata
25561 taggatattg aaaaatttta gaaggaaata caaaaatgc atggcccttg cctcaggtag
25621 tggaaattat agtgatatct agttttgctt tatgcgttct catatttttt tttcttctaa
25681 ttattcctaa ttgtttttat tttccaaatt tcttactgta cccaaagggg aggcatttgt
25741 aaattttctc caccagcctg gcattgccct gactctttcc gtcttagtgt cacttctcag
25801 tgggtatcat gaccagtgc aagtgtgag gcacagggcc aggcggccca cgaggtagcc
25861 ctgaaggggc agcgatctgt ggctccaatg agccaatcat tgggaaactg gcatcttct
25921 ggggctggga caggagagct ttgttgggtg gagaggaggg cagagccgaa ggagagaggg
25981 caggggagag gcaagggtcac cctgtctcag cgccttcttt tctctgccag ggctttactg
26041 ccttaacacc ctaacactcc caggccaggc aggcagcaag acacatgggg acagtgagct
26101 ataaagcagt gagtcccgca ggttccctgag acacaagagg cagattgttc cctctctgca
26161 tatgtttgta tttaaatagc ctcatgttca gacaagcaac ttggatttgc ttcagtcttc
26221 tcaggatggg tgagaaaaac actggaaatg ctcaaaacga ggcagaacag aagacacctg
26281 tcatggtgga gatggcatcc aagtgaatga acgcgagtct gttatcaatg aattatatcc
26341 ggtgtaacca gatgtagaga gattgtcttg tgtgagaatg caccagcagg caaagggtt
26401 tacgaggcag agtcagggtc cctgcaccgg cagctgtttt aaatgaattg acccagggtt
26461 tccacattca ttcattaaac atttgttcat tgaacaaata tttggctcta attccgcttc
26521 cacctcagac agctatgcag cccggagcac atcactgagc tttcctgaga ctctgtttct
26581 ctactgtgct gacagccaga ggtttgctgc agagctggag acctggtag cctccaggtg
26641 tgtgtctggc aacatgaact gtgcatgaat ccttgtcttc tcagatgctg gagcaccac
26701 ctacatgtat gagtttcagt accgtccaag cttctcatca gacatgaaac ccaagacggt
26761 gataggagac cacggggatg agctcttctc cgtctttggg gccccatttt taaaaggtaa
26821 tgctccttcc tgtctgtgag ctaggagtta gagacttggg ttccagtctt ggtgtgggtg
26881 ccttcagcaa ttttgtcttc tctgtgtctt agtttctctga tttgtataa cagggttgaa
26941 tgtcactggg ttgtccgccc gcttctctcc actctaacac atgtgaggag ggatgtgtat
27001 tcatgtgcc tctgcacaga cgctcagaat cctgctgtgt gacatcaaga cagggtgggtg
27061 tgggggtctc ctgcaggatt cactatgaca ggggtgaaat tcaactgggt gttccccggc
27121 ttcctcccgc tctaacacac ctgaggagg atgtgtatca tgtacctct gccacagaca

```

Figure 4 continued

```

27181 ctcagaatcc tgggctgtgt catcaacaag ataggcggtg ggacacatgc ggggccctgc
27241 tctaggttga gaagccacca atcaatttcc ctctcccagg agcgctgaga cagtgagtga
27301 gggtgagatt agaattgggtg ttaaccctaa gtcagggtcta atgtgtgcct gtttgtcttt
27361 ctgtctgtct ctaacaaaaa ctcataaacc ttacatttta gtacatttaa taattgttga
27421 tttagaattc acaggatgta tgaatttgga acaggagaag agaataattt ttatatgggt
27481 tatagtcagc aagggtggcca tcccacaggc tgggaagggt cctctggcca agcccagaaa
27541 cgggctcatt ggaggagggg ttgggtagga gctttatgtg aacggattgg ctaaacatac
27601 atgttcaaca ggttacaggg ggagcaatgg atattcatga agacagtcct gacacatgtg
27661 tattaacaa acatgtatgt aacatggccc atgttcacct ggtggaggag acctaattt
27721 taaatgtatt acaattaggg cctgtaagtc caaaggctct ctcaggacac gaagctcagc
27781 aagtgaggag ctctctgtaca cgggccaggc ccagtcctat gctgggtgat ttcttatctg
27841 gagaaagtta ctgaaatcag tctcttatgc aatcaaagca gtagttaaagt ctggcaagtc
27901 agggttctgt ttgtctgcgt atcccagctg cagccgttgt tattgttttc cttttgctta
27961 gctccaggcc agtgctgggt tagctgctag agaaagagaa gcacctcgtg gcagtgagaa
28021 caaagtggat tctttttttt ttttattatt atactttaag ttttagggta catgtgcaca
28081 ttgtgcagggt tagttacata tgtatacatg tgccgtgctg gtgcgctgca cccactaact
28141 cgtcatctag cattaggtat atctcccagt gctatccctc cccctcccc ccacccacc
28201 acagtcccca gagtgtgata ttccccttcc tgtgtccatg tgatctcatt gttcaattcc
28261 cacctatgag tgagaatatg cgggtgtttg tttttgttc ttgcgatagt ttactgagaa
28321 tgatggtttc caatttcctc catgtcccta caaaggacat gaactcatca ttttttatgg
28381 ctgcatagta ttccacgggt tatatgtgcc acattttctt aatccagtc atcattgttg
28441 gacatttggg ttggttccaa gtctttgcta ttgtgaataa tgccgcaata aacatacgtg
28501 tgcatgtgtc tttatagcag catgatttat agtcatttgg gtatataccc agtaatggga
28561 tggctgggtc aaatggtatt tctagttcta gatccctgag gaatcgccac actgacttcc
28621 acaatggttg aactagttaa cagtcccacc aacagtgtaa aagtgttcc atttctccac
28681 atcctctcca gcacctgttg tttcctgact ttttagagat gcaggectga acccttgccc
28741 gacgtggcct taggtctcgt ttttaatttg gtgtcttatt gccacaaaga gtgtgttctg
28801 tcagaatgat gaccttcatt ttattgctga tgcgtggtcg gtggtgtcta aatcacaaaa
28861 gggaggaggt attatgaggc gtgtctgacc tccgtgctcg tcccaggcag gaacagagtt
28921 gaaggttttt cgaggctccc cttagccag agagggtctg ttcagtcagt tgggggtgtc
28981 aggattttat ttttagttta cattatacaa tagtttaaaa aatcctttca cgggtgtctc
29041 taagtcttaa ctctcagcag ttacagatca gatgggactc accccatggc tgctcagaac
29101 gcaccagcgc ctctgggcat ctactgtgct atgcttaggc gccttgcggc tctgttgttt
29161 ttcagaatgt gaattgtggg tgtgtgctgg ggaggagaa aagacccaag agagagcaac
29221 cctggaaagt ggggttgcgt tagaggaaac cctgggggtg ggcatttcc gtgacctga
29281 gcggggagga ggcaggcgag ggtcacagga gacaaaggca gccagccaat ggaatgacgc
29341 atgcccattg ctgagcccta ggtcagatat gctgaaagg aaagtacaca caaatccacc
29401 acgatcttct aaaagtgcag acctcagctc agtgggtgtg gggcatgtgc caagactatg
29461 tttccagcag actcgcatgt gatgttgacg ctgcagtcga cggaccacac tttgagtaac
29521 aaaggctttt tttcttcttc cccacagagg gtgcctcaga agaggagatc agacttagca
29581 agatggtgat gaaattctgg gccaaacttg ctcgcaatgg gtgaggctgg tggcaaagac
29641 agagcacggc tgggtgagggt ggggggagg gcatgcctat tgggaagggg cagcttctaa
29701 ggttctagt atcaaaactc tgacctgtg accatagcac tctgataatg agagctctct
29761 gcaaacggag aggcgcgccc tggagatagt gagctctcca tctctggagg tatacaagcc
29821 tcttgacaga gataacttgg gcatcctcac acatctctga agattgttgg gaacacacag
29881 cagctttggg gcaattctaa ttgattctgt ttccagaaac cccaatgggg aagggtgcc
29941 cactggcca gactacaacc agaaggagg gtatctgcag attggtgcca acaccaggc
30001 ggcccagaag ctgaaggaca aagaagtagc tttctggacc aacctcttg ccaagaaggc
30061 agtgagaag ccacccaga cagaacacat agagctgtga atgaagatcc agccggcctt
30121 gggagcctgg aggagcaaa actgggtct tttgcgaaag ggattgcagg ttcagaaggc
30181 atcttaccat ggctggggaa ttgtctgggt gtggggggca ggggacagag gccatgaagg
30241 agcaagtttt gtatttgtga cctcagcttt gggaataaag gatcttttga aggcc

```

Figure 4 continued

LOCUS AB119998_hCES1A2_c 30275 bp ds-DNA linear 14-APR-2011
 DEFINITION
 ACCESSION
 VERSION
 SOURCE
 ORGANISM
 COMMENT
 COMMENT ApEinfo:methylated:1
 ORIGIN

```

1 acagtcacctc tgagctgcac ggagacctcg caggcccccg gaactgtcgc ccttccagga
61 tgtggctccc tgctcttgct ctggccactc tcgctgcttc cgcggcttgg ggtgagtcct
121 tctgaagtca aatatgcggg gcactttttg aaatccttgt tctgggccga actgggcgca
181 gatgcgtaga aaggcaaaga cacaaaaggt ccggctccgt ggcggggcgc gacctccgta
241 cttggaattg gacttgagac agctgagccc agccagcctg gccgcgcgcg agacacaggg
301 aggggtgaaca tgctgagacc gccctgcact gcttgcccca agaaggccgg cagctggccc
361 acctcactcc gcacacaaga gaaatttagg cagctcagtt tattgtccca taattcgtcc
421 cgttatggct cgataacgga tagcggtcca ggagaatatt aagagcatcg actctggagc
481 cagagtgcct ggggttcgagt ccccgctggg cgcttctctc ctgtgtgtac ttaggcaaga
541 gactgagcta ttctgtgctt gggttacccc ctctctgaaa cagaggtggc agtgctacct
601 acctcactgg gctttggtga ggaggaaatg ggtgggtgtt tctaaagcac tcttggaagc
661 gcctggtagt tagaaagcat gaaaaagggg tggataaatt taaaggcgct gtttaaaaac
721 gtagtgtagt aatagaaaat ttgagccgtg gtcaggccac tatatcagga cctcttgcca
781 ttaaaaagac agtttcttac tcatgttctc aacaggaggg ggacaccagt gcgcggggca
841 caaggggaag cagcagagcc tggggagggg acagggcgga tctgaggcca aggggctttt
901 actgtggctt ccaaggatag aagaggaaag gcagggcagg caggetgagg actggcaagt
961 tgggatactt caggggctct ggggcacagg ggtgttctct aattgttttg tacctagccc
1021 tggcatgatt agggcaggtg ggtagtggcc agggcatggc aaccgatca aagaggcagt
1081 tggggtgtgg gctctggggg ggttgagcgc tgggagggac agtttctcca aggtcaacaa
1141 ggctcagat gccagagtat cagaatacag aaaacaaaac acatggtaca tagaggggtg
1201 cagcgagggg cctcgaaata ccccatgcc attttataga cagagaaact gagactcaca
1261 cagggaaatg tacttgccca gggccaacc ctaaatcatg cctaggtctg tctctgggtc
1321 ctgaatcctg acagcccagc cccgtctgca tcctgtcttc cttctggcag cctgacactg
1381 cgagagctct ggaactgtgg ggaagggcag agaagagagg aagcctattt ggaactgaat
1441 gttcctttac ctctggttac tctggcaat gccatccatt tcttctctta ggaagaccat
1501 acaggctgcc tccacttctc cagtttcttc ctctgtaaaa tgggcatggt gatgcctacc
1561 cagcaaggct tatttgaaga tggattttaa caaactggaa tgcagcctg gcacatagta
1621 ggtgccacc atgttcgtgc ttcatataac atcagcagca actggtcttg ttctcacact
1681 tgtttctgcc tccccctatt caccacttt ttacccttg gcaatctggc tgccatccac
1741 ctcatcctgt tgaagggaat ccctgagctg agcatgacca gccagtggt ttcttctctg
1801 acttgactct cctccacctc cctgaagtc tcagcaatct tagctgggtc ctgcacctct
1861 tgctccctgg gcttcaaaga catgggaacg ttctctctcc cctattctga tggatctttg
1921 gtcttgccg tagtggtccc tgtgtttgac ctttgttttt ttttctcttc cctctcacac
1981 tggcatcttc aatgaagcca tacacatcca ccgtatgtac cctcacagtc caggagaaga
2041 ttctgccctg ctccctcacc aggttctctt ccagctttga agtgcgctca tctgtggca
2101 tcatcttgcc tggatatccc actcacatct ccaccgagcc tgtgtaaaca gggatccagc
2161 attacaagcc agctcttttg tcttcttttc cttccatttc cccacctct ggtcagggac
2221 caagtctttt ctttctcttt ctttcttttc cctccctccc tcaactccac ttgctgctct
2281 gcctgcctgc ctgctgcctt gcctgccttc cttccttctt tcttctcttc cttcttctct
2341 tcttctcttc cttcttctct tcttctcttc tttccttctt tcttctcttc cttctctctt
2401 tcttctcttc ctttctctct tctctctctc tggaggagcc aggattacag gctcatgcta
2461 ccacaccag ctaatttttc tattttcagt agagaggggg tttcaccatg ttggccaggc
2521 tggctcttgaa ctctgacct caaatgatct gcctgccttg gcctcccaaa gtgctgggat
2581 tacaggtgat ccaactgtgc tggctctcaa gtctttcaaa atacctttta gacatcatca
2641 tatttctaca tattttttct ttttttattt aaaattttta aataattaag atggggggga
2701 ggatttcagg tgaagaccgg aaatcccata cagtgtctgaa aagggtgaaa tcaccaagaa
2761 aacaggccac atgtaatgcc accaccatt gctatgtcga taaacatgct tccaaggaag

```

Figure 4 continued

```

2821 ttcatctcag tattatttgt aattagcaac aacaacaaaa gaaaagaagc taaatattga
2881 cattgcaggg ttggctatta ataagatgag acatggatgc tgttgtgtaa ctgggctttc
2941 cataatgggt ttccattgtg attggatatt cccctttaag accaaatctg aattgttctg
3001 ttctgctcta gcccacact ttgagggagg tgttgacgtc ttgtctgtat ccaaagcaga
3061 gcagctggga ggacaagagg ctggaaatca tttcctagga gggaagctca gagaaactga
3121 ggtgactgga gccagatagg acctgagggg tgggcagcaa gatcaagtgt ctgaaaggcc
3181 acaggctgca gatcagatca ggccagtata gtcccatgac aggtagaacc agaactgatg
3241 cggaggctgt gggaaatgaa tttcagctca gtgaggaaac attgtccgtg ctgcaagaat
3301 ggagcagacc tgcccagggc caccagaaga atcttccttg acctgatcac atcctttatc
3361 tggaacgggt cctgctgctt atttgtccat taaagaaaaa tcaagcgctt agcctggcat
3421 tctaagccct cctgacatg tgcggtctca ttccaggaat attttccact cccccagac
3481 tccacctggt gcatgtgggc catgccccct tccccaggca ggctctgcat tctgctgctc
3541 tgagcctcag caaattccct tctctctgct gtgtaccact ctctgcttct catctccctc
3601 cattcagtc cccctgccg caagtcccag ctccctccatg atgcccgaat tcagaaggcc
3661 ttgccccctc ttctaattcc tagactgcct tatgggtacc tctcccagga catggccact
3721 tccttcccag ctgtagcccc aggtgtatat gtgacccttc cctataatgg tcagtgcctg
3781 tgagaatatg ggacaccctc ttttcactct ctcatccagc atgcggtcca gtgccaggat
3841 tctaattgat aaatgtttcc agagacttct aagggggaag ctaaatgtct tgttctttcc
3901 tagtagctct cttctcttgc tttattttta gctggatgtt tttatgcctc caattctagt
3961 ttgcccatta aaccagctca atgggttagg gaggacattg atcgatccct ctattgtaca
4021 tcaagagaaa ctgaggccta gaggggttag gtgacttatt taaggctact cacttagaaa
4081 gcggcaaact ccacctggaa tctgggtcca gccttttgcc tctgatgcat cctgatttat
4141 tctccatgtc cagcagggca tccgtcctcg ccacctgtgg tggacaccgt gcatggcaaa
4201 gtgctgggga agttcgtcag cttagaagga tttgcacagc ctgtggccat tttcctggga
4261 atcccttttg ccaagccgcc tcttggaacc ctgagggtta ctccaccgca gcctgcagaa
4321 ccatggagct ttgtgaagaa tgccacctcg taccctccta tgtaagetgc ggcatgtgtc
4381 cttggggatg tttacctcaa agtgatgcag gaaggagtca aggcagtcct ctgatgggct
4441 gatcctttgc tctggactcc ttaagatctt tgtagatcct taagaacatt ccagaactct
4501 cacagcattc tggagtccat tatttaacac atgtttattg agcacctgct gtgtgccagg
4561 cgtgttctgg gttcttggga tccattagtg aaaaagcaaa gatccctagg ttccctggagc
4621 gtgcattcta gcagggggag acagataaga acaatatgca aattactccg tggattatgt
4681 gatagtaga aacaggaaaa gtgccttggga aacaaagaag cagttaagca ggtgaggag
4741 atcagaagga aggatcccat gaattctgct gtcttgaatt ttgacatagt cggggagagg
4801 agtgggccag gaggtgtgtg ggtggcaagt gtactgaggt gtccactcgg caagtgtcca
4861 actctcaggt gtgcctattc ctgaggagaa caatggaagg tgtgtccatt caccctggcc
4921 aagctgggaa gaaaagccca aaggttctaa ttggcctcac catcactccc cagggtagga
4981 aggacctac agacatcaga gcagacctg ctcatctcag cagccagggtg ttccggaggct
5041 ttccagcagc ccaccagtgg ccgccttctt ttcttccctt ttttaaaaaa taatgtttac
5101 tgtttttatt ttctgagtgc aaatgtaata tatacttatt acagaaaatt taggaaatac
5161 agagaaagga aagtaagaaa attaagatca catacaatat tccaccacc agagagaaca
5221 actatcagta ttttagatgg taaagtatta tgatgtaata tattgtttgt attccagttt
5281 ttttgctgtg tgatatactt tctttcttta aaagggaatg ctatattgag ataatttgga
5341 tgctgtcttt gctcatttat aagtatatca tgagcataat tccatgccat taaatagcac
5401 agcctggtat taaaagctgc attacattgg gtattgaacc ataattaatt tccccagct
5461 tttatttggg gtcattaaagt tatatatgat gtttttaata atataacatc atgaatgtga
5521 aacatttatg cacattcatg aatagtttag gataacttct cagaaatgac atttctgagt
5581 caaagagtat atatttcttg tttgatttcc ctcccaaaat attccaacta cactcctccc
5641 attgacgtat gaacatctcc atttggcaga ctcttggcaa tgcttgggat tattaaaaat
5701 atttttggct attgtatagg caaaataata ataacaatac cagcaatatt attttctct
5761 ttaatttgca tttatctgaa gaccagttag tcttttttcc ctttcttcac attggaatgt
5821 attattatta ttattattat tattattatt gttattcctc ttgcacttgg tgatcttagg
5881 aaaatttccc aggatggaat caagaagtca tctttgctgg ctctgattct gtcacttgag
5941 agctgtgtca tcccagagtc tgagtgcctc agtttccctc tgtacctagt gggggcgggtg
6001 atgactgcct tttggtacat tgaggtctgg gactgttgct tataatatac ttgtgtataa
6061 gcaccacatg gcatggatct acctggctat gtaaccttgg tcaggttctc taacctgcct
6121 aggttccagt tttctcatct gtaaaatgga gataatagaa cctacctaat gggactagtc
6181 tgaggcttaa atgagttaat atacaaaagg tacttaaaac agtgccctgga acactgggag

```

Figure 4 continued

```

6241 aactaaataa ctattaccag tgttcatcta tgttgcatac ttgagatctt tcatcagggc
6301 aacacaacag caactgcctg ggagaggcag gtggtagaat cttggcatgt cagggtgca
6361 agactcatgt agaaatcact ccgctcactt agacctggcg gattgggacg cagagtgtaa
6421 aaggaattca cccaaggtca ccccataaat gcatgggaga gacaaaacta gagcacaatt
6481 ctctgatgt tgtcccacaa agatatcacc cctgagctat tgtgagaaat gtcacctccc
6541 agggaggatc agtgtattag tggctctccc tccttgatgc tgggagttcc aaaggctctg
6601 gaaagggaag aggtgtgaag cccttctcac tctgcttggg cttaggagac cttagttagt
6661 cccagcgccc ccacctgaag cccctgatag cctcctaccc actacaatgt cgtgagtctg
6721 taactatcaa gtccatttcc aggettataac ctcccagtc caggcgctgc cgtagaaaac
6781 acctcgctta gatctggctg aacttcaggg gttcttcttt cccctcccc aggtgcaccc
6841 aagatcccaa ggcggggcag ttactctcag agctattttac aaaccgaaag gagaacattc
6901 ctctcaagct ttctgaagac tgtctttacc tcaatatttta cactcctgct gacttgacca
6961 agaaaaacag gctgccggtg agttgtggga cccctgggtc aaggcgtgtc agtgccagta
7021 ccccgcatct tctggggcag aatggaaggg agtgaaggca gcttggggag ggagctgcac
7081 agggcaggat tgtttcagag accccagggc ctccacagga aacactgggt ttccacacat
7141 cagttttacc tcgtgacata ggcactctgg gggagtttgt tttgcatctt taatgaacat
7201 cttaattggt ccaagttccc tcatgctaata agaaggatga gaatgattac ttttagttac
7261 tagattaaaa atctgatacc aagattgggt aagacatgat tttcaatgtt agtgggtgaa
7321 agaacttaga gaaaagccaa ccccccaact ctctcttttc aaatagagaa actgaggccc
7381 agacagggaa ggggaggtca cccagtaggg tagttttggc tcaagtcttc tgatcttcac
7441 atcagtgtct ttgacatgga gtagggcatt gagcttgtat gtggcctttc tccagcacag
7501 caacattcat ctggggtaac tttcactgtc caggtagcga tgttgggggc ggagcctgga
7561 caatcccaga agctcttcct tttgcctttg gccaaatagc tgggatttct ttttgtacct
7621 aatttccctc cctaattgtc ttggcaaagg ctgatgaccc agtgagtcag aatttcccaa
7681 ataaatgaag ggcaggacat ggggcagccc gccagctgca ggttggaggg tggggagggt
7741 ctgctggcaa aggaggcacc cagggtgaaa aaacaagggg tgttgtgtct agccagtctc
7801 ccaggggcac cgtggcatgg tgccataggag tgggggagca ggtctgcgaa cgtggaacta
7861 gaagttggtc ctacatgtgc tttgtgagac acagggaagt agagaaaggt ttactcattc
7921 attcattcat tcattcattc attcattctt ctatttgtta ctcagtgcct catatacaca
7981 gttagacctg gggcttcagg gaaaggagga acatgaatcc actccacaca gcccctgct
8041 tcaaggatct tgccctctgg gctgtagaga catatcctcc cttcagcact tttatgcag
8101 ccagccctca ggctcagtc tccacaaaag atgaatgcac actcaatccc tcccagag
8161 gcaggcaggg attagagttg tgggagaaac aagggtctct ggaacctgga gtcaggagaa
8221 cctccaggaa tggccacagg ctctagaact acacggtctg ggtttaaatc ctggttctgc
8281 ttttgactag ctgagagact tgggagtttc ttaacttctc taagcctctg tttcctcaa
8341 atgggaatca tcacatatga tcagtcaagg gtgaccacta agcgtgatct cagtccctgg
8401 ctaaaaaggc agaggcaggc tcataccctt gccctaacag tggactcact gcaactgtgg
8461 cggcgctgc agtgtgctgg gcagctcaac atgtggactc tggaggcaga ctgcctgggt
8521 taaagtcctg ctgctaccac ttttgagttg taggaccttg ggcaggttaa gtgcctccat
8581 tttccacctt ggataatggg acattatcag tacctacca atgtatagtt tattttactg
8641 tgcttagaac agcacttggg acatagttta tgttacgagt tcttttatat tcatgaata
8701 atattattac aatatcatta ttttagccac tctcacttct tccccactat gtggactttt
8761 aaaatgcaag aatcatgcat gaggaattgt atccactctc atcctccagt attagcagag
8821 tccctggcac atgctaagtt ctgcacagtg ttagtccatt gactgaaggc atcctgaagt
8881 ggggtgggtt taggctgggt tctgcagggt ggataggatt tggggatgca gaggcataag
8941 gaggcatttg aggaagagac accatcacag gcacgtgtgc ttctcgagac aatcattcat
9001 gcacagagaa gtttacttca ccaagagata gtttaactgag atgaagctga gatacaagg
9061 ctgccccatg ggtacgtgga gtgcatgaat agtctaggct tgaggggtgat gggagtgtcc
9121 tcccgaagag gacatcctta gcctttgcta tgcccatgat gatgttctca gcatcaagag
9181 cctttcgggg aggaggcact ggaggctggg tagaaaggaa gggacagcca gggatgtgc
9241 agtagggtgg tggttgggt cccagttggg agaattatgc aggagagaga ttgccttttg
9301 caaagttgct gggagctggg gaactttgga tagcatctc cttcagttgt aaccaacagg
9361 taggtcacca aacaagacat cttctgagtg ctccctgggt ctgtgtcctg taactgggat
9421 gttgtcttct ctctgccag gtgatggtgt ggatccacgg aggggggctg atggtgggtg
9481 cggcatcaac ctatgatggg ctggcccttg ctgccatga aaacgtgggt gtggtgacca
9541 ttcaatatcg cctgggcac tggggattct tcaggtaaga aatcggaact tctcactgc
9601 actttggccc ccagaacgag gatgctagga ccagctctg gtcatgccag cctcagggg

```

Figure 4 continued

```

9661 agcttagcta ggttccacag taaggcatcc aagccccttc gtaattggac actacctacc
9721 ctctcactac ccagccactc atccacttgc ctctgagctt ttgcatgtgc tgttccctct
9781 gcctggaatg cttattctat ccgaagaaca cctcttcac cgtcaagacc cagccccag
9841 ttacctccac tgaaagactc atctcctctt cactatgtt tctccagcac attgcttttc
9901 tctgtcatga cacttagcag tcagcccagc aggtgaaggt tgacacatct gacttctcat
9961 ggaaaggagg gaaggggcag agtcacagag ggagctcagg gcatgttggt ggtgaaggaa
10021 tgggggctgc agatgggtgg gtgtgggagc atctgacccc ttctgtggg agtgttggga
10081 cccactcagc tcattttctg cttgaaattt ggcaagtgtg ggaagcaagg atggagtctt
10141 cccatcaggt ctccatgggg acaggagtca gatctccttc aaggcgcaac tgctatgaac
10201 cttttctaa gttctcctcc atccatgcc agtcaggttc ttctgaaca cgccactcag
10261 ttggtctcat tgcggtcac aatgacctcc cctgtctct caccacacct ctcaggacct
10321 cccagcagca gcagatggcc actcagcctc tgtgacatgc ttcttccca gcccctttcc
10381 tggctcctcc tcctcctgt ccacaggcct ctcttgtctc cctctcctc cttactaatt
10441 ccagtgatc ccaggagca gccagagccc ctgtttccag ccacacctac ctttgaagcc
10501 cacaggcatc actcactccc tcctccccac cctgggtgt ctaaagggtg cccactctt
10561 cccagggaca agttgaatcc ttgactccca ctgcccctcc ccaacagaga cccccttctc
10621 ctcccaccac gcctccactc agcttctgct gccacatcct cccttgggtc caggagaagg
10681 ccttgtggac agcccggact gtctcttttc tcacaacct cttggtgctg ccttcagaat
10741 ccaactcagag tccctactt tgcacctctt cctgggtcc tagccagcc aaagctctct
10801 gtctgtcacc ttgattatga tctttaacac tttcagatcc ctgccacttt tatataaagc
10861 tgagttgcac agtagccaga gtgatcactt aacacatgca ttgaatcatg cgatgcctct
10921 gcacaaaacc ctgcaggggc ttcccatgct ccttaccatg acagccgaag ttcttgccag
10981 gccccagggc cctctctagc tgcacctcct gactcacact ttctttcact gggccccagg
11041 ccctttgcac ttgctttttc ttctgcctgg aattcctttc tttggatatt tacaagggtc
11101 atgtgctcat tgtgcttggt ttatgaaggt ctttgcctcc aaaacctcaa atgtcactgc
11161 ctccaaaaag cttccctga ttgggatctg aaatggcacc ccctgtccca ctctgtttt
11221 tttccttact ccattgtatt tacctcagt gactttgact acctgacatc atttatattc
11281 ctctatgtat ctgtttatct gcttgtgatc tgtagtggga attcagtgtc atcacagggt
11341 ttgctgttt tgctctcaga tgtgctctag ctcataagac agcgtcaggc atacagtaga
11401 tgctcagtaa atagttgcca gttgtgtgaa tgtagaacca tacatcacca caatgctgta
11461 ctaatgaaag caggggtctc gagatgctga agcccagcct tattcttaag ggttactga
11521 gaacccctag cccatctgt ggtcctgaag gtctctctgc acatctctgc tccccacct
11581 caacctgttc tcttctcac agcacagggg atgaacacag ccgggggaac tggggtcacc
11641 tggaccaggt ggctgccctg cgctgggtcc aggacaacat tgccagcttt ggagggaacc
11701 caggctctgt gaccatcttt ggagagtcat cgggaggaga aagtgtctct gttcttgtga
11761 gttttcctgt caccaggccc aacccacgc ttgatgtgat gctgatgaga cctcttgac
11821 acctgtcaat ccagctatgg atacctctga ttacaatacc tgaattcagg actggcccta
11881 ctcacagccc agcatcagg ggtagaaagc taaagaccag ctctccttgg ccccaggaa
11941 gctcagagct cagcttagtg tctggggcc atatgaatct ccagttgcag cctgcaatgg
12001 tggcattctc ctctccagc ttggttgagc tctctctctc tctctctctt tctctctctc
12061 cttcacactc attttaacat aaggactcac atcccttctc ttgggaagta ggaataggca
12121 taaattatga gtaagggcag gcagagagaa gaggtggaca gaggtcattg tttggctaaa
12181 ccagacctac atgtggagg gacatggaca ctgagcaggc tgggaattcc tctggggtgg
12241 tctgatggct tgtccatgcc caagaaggag ggcacagctt ttgtgactgt ggggtccagt
12301 gggctagggg aggcaggcag aggaaggagg gatggagccg cggataggga gggatggggc
12361 aggggttgtg ggtcatagac acaaccttgg gtgtgagggg tctgtctggg attggggatt
12421 gggttcagga gacactgggg gatctgggat gaaaaccag atgagagggt ctgggagctg
12481 taggaagact tccacctcct tgaggtgggc agaggtcag ccactactg gattcctcag
12541 tcccggttg gttttatagt ggagttagat tagcctggaa tagcagtga gtcactgacc
12601 ccaactcctga gcatgaactc tcctccctc cactctgtg tcaggtttg tctccattgg
12661 ccaagaacct cttccaccgg gccatttctg agagtggcgt ggccctcact tctgttctgg
12721 tgaagaaagg tgatgtcaag cccttggctg aggtaggctt ccggtggta cgtctccggc
12781 tggacacccc cacctccttg gctctatgct cctgaatcct cagggatctc tcttgtggtc
12841 ggttgtagct aatgttctcc tagaatcact gaggcaccaa tggctgagca ggaaggcgga
12901 ggagacacct tgatcagcgt cccagtttca cagccaggca aaccgacaca gggcttggaa
12961 gggatttgcc aagggcagca ggtgatcagg gcagagctgg gactccagct catggcccta
13021 gcagccagta cagtgcctca tctgtgacca cactcctcct atgtgccagg gcctgggtgc

```

Figure 4 continued

```

13081 atgttgggca gtgatggtgt cttgtgtctc tctgggtctg cctaattggct ggtaagtgga
13141 acaaccagga tttgaaagca gacaaggaaa ttcaggattc catgctctct atcccagctc
13201 cacctcacat ttcttgacac attcaccttg agatgattat gtccatttca atggagataa
13261 ggtagcagag atgagagatt acataattga cccatgttac aattgagatt agtaacagag
13321 gcaacgctcc gtgggacctg gaaccacac ccacgggtct acagtatctt ctgctgctcc
13381 ctgccactag tacaagttgg gcttgggata gaatgccact ttcctctttg atggagggaa
13441 gggatgtcgc tcttgaactc tgttgttgcc tgtgatcttt gcagcaaatt gctatcactg
13501 ctgggtgcaa aaccaccacc tctgctgtca tggttcactg cctgcgacag aagacggaag
13561 aggagctctt ggagacgaca ttgaaaatgg taggttgcoo gtccccgtag cccaaaccct
13621 gtaaaccttg tcccagactt cttcatttca gctgtcctct tgcccaggga cagtttcctg
13681 ggacaatttc tcaactctca gtatctgaat ggggaatctg atttgtcctt ttttattgta
13741 aaatgccaca aatgtaaaaa aagaaagtca aaaaataaca tgataaaaaa ttgtgtaaca
13801 atccccagtc ctgaccaaaa tttaatattt tgccattctt gttttcagat gtatttttaa
13861 gaaattaaat gttacatatt cccaggggac accatcatcc tgaatttggg ggtaggagat
13921 attagctca aagattttca gaaagatgtc acaatttata ttgggtgact tagaaaactgt
13981 ctgtattaga cctagtgggt gtccagtttt ctagattttc tgagactgac tcagttgtca
14041 tcctaggata gtcattccat caccatttgc ttaatccgtc tatctatgat tgtcttatcc
14101 atctatctgt tcactaattc atccaattca ctcatatctt tttatcaatc taatgtcaac
14161 ctattcataa ctttgtattt atctattttc aatccattca ccattcatgg atcatccagc
14221 caccttatat ctcaactcca tccccattc ctccaaaatc aacaatccaa ttatcgctg
14281 tctgctagtt ttcacccatc tattcatgta tccattcaat tcaactgtac tccatgtatt
14341 gaccaactcc atccatccct ccattgatcc atccatccat ccattgctaaa tatgtagggg
14401 tgggtgttag aggtagcaaa acagacatga agtggacata gtccctgctc tcaaggaact
14461 atccaaagag aaatacatte atatacttcg cagggttgagt atgggtggcag cacagagggg
14521 aagcattcat tgtagtaatc agggatgctt cacagagaag gcggaggagc cagatttgag
14581 accagacagt ggaattcagg agttcatgcc cttaatccct actccacctt atctttcctg
14641 agaaattcag ctttcagatc attgtgcccc ttttaacaga ggtaacaaag acagataaat
14701 tttgtaattg ccccatgtca cagttcaaat tagtcacaga ggcaacactg taggaccag
14761 aaccgacacc tgcggtctca gagtatcttc tgccctccct gctccctgcc atcctcacat
14821 attaggactt ggggaccttt aggattgttg gatggccatg gcagtctctc atcacaggga
14881 agaccaggag gacaaacacc cagatgacac agcaccaggg gccattggga actattccct
14941 ttgaggggga agagtgtgaa ttccaagctc aggagtagec tggactctct ccttggacct
15001 gaggctactc ctggatccca gggtggcct ctaccactgg actctgcttg tgtttccatg
15061 agttgagtc aatgtggtat tgggtgcttag gtcttggtc aggtctctaa tgaccaaaag
15121 gttcaaggaa tgaaaacaca actgttttct aacgcccggg aggaggcaaa tattccagca
15181 attctgcatg tggcatcaga ggaccagct taaaaggga gagttgagtc tttgggaaaa
15241 cccatcccta agatcctaga acattctttt gagtttttct gagatcttgt ggaagcacct
15301 ccatgattac gcgctgcctc cagtacacac acatgcagac acacagacaa acacacacac
15361 acacagacac acacacagag aaacacatac agacacacac acacagagac acacacagac
15421 acccacacac aaagagccac acacagacac actcatacac acacacacac aactcctgg
15481 ctagtgggac ttcaatcctt aatggaggga cctgcactgg ttacatcctg agcacaggtc
15541 agatctggga actacagtgc aacactacag tcaactgcaat cttggtgaac acacacaaaa
15601 gagaagcatg gggtaggatg agactcagtg agcttgctg agtcagggtt ttctaattgg
15661 agatgaggaa actctcccaa gtacctgaa cagaaaaaga aatgtgttg aatagcatcg
15721 tcgtgtaaac tgaaaatctc aggagtgtat ggcttcaggc atgggtgggt ccagatgatc
15781 ataggatatt attgagaatc taccattctc tgtccttcag ctttgcacat ctctgtgttg
15841 gcttactct taggcagata tatctcctgg gaagggttaa gacaacagca ctctagactt
15901 atgtcctatc tgcttaaaaa cctcttttct aaacaattcc tgcaaaagtt ttggggtaaa
15961 ctctattgga ttgacatgag tgtatgccc tccctgaact aatattcatg ttcagggaga
16021 tggagcaccg caggtcacat atagatatac gaattcacgg agtgatgcgg gaagaacctg
16081 acacctctgt tgccccactc acccagctca gtgttctcct ggtaagcctc tcaccacat
16141 cctctgcctt tgtcttcaca gaaattctta tctctggact tacagggaga cccagagag
16201 gtaaggacct tttgtttctg gattacgggt tttgagtcct agcaccttta agctccaatt
16261 aactgtgagt gaaagaatca tctctgggt aattatagta actcctgcgt gtttgtttct
16321 gagggccaga gaggggtagt gactcacctg ggtaacacag ccaggaagac tagtggctgg
16381 cctggaacct attttctga ctcccagtc agtgctccca ggcacacct ctgtgtgccc
16441 tgggctctgc ccactccct tcttttttac attttctgct ccccaaatg gcactgtagg

```

Figure 4 continued

```

16501 aggaggctga aacagaacct tccacaatca ggaggcagag ataacagega tgattagcca
16561 aggagagggg aagcctaaat ctcagtcctc ggacactcgt ctctctccag cacacaggaa
16621 actccaacag tattctccca atctctctcat cccactccc acccccacct cccagtgggt
16681 tgacagtttc tgggtgacatc acctctgacg aatcttaciaa tctgtctctc tctgtctgct
16741 ccctggagat cacagcactc tctaaatgg tcatgggcgg agtacatgag attcattcag
16801 caaagactta ataaacactt cctatgtgcc aggacttgtt taggagctgg ggatataaca
16861 gtgaaaaaag aaagactccc taccctcttg ggaccttcca ttacagtgcc aggacagtgc
16921 aataaaaaag caaatcaagt atgtattccg tttgatgggtg tgcagcatta ggggatagag
16981 tttgtgggtg gcttcaattt aagtagcagg atcagggaat ggttcaatga gaaggtggca
17041 tttgagccaa gctctggagg aggcaggaag ccagctgtca ggaaacctgg agaagcctat
17101 tccaggcaga ggggaacagtc actgcaaaga cctgacaga aggggccagt ctggctggag
17161 aggaccgagt gtggggacag actggcttga ggctaaagag gtaatgggca gtgggagaga
17221 gatcaattaa agtacttgga tttggattca gagcaagatg ggaagccttt ggaggttttg
17281 aacagaggag tcacatgac ttagatctca caaaggctct tgtctctggt gttcaaatag
17341 actgtaggaa aggggcaaag tggatgcagt tgaccagctg ggcagccact gcattgccat
17401 aatccaggca agggctcctg gctgcttaga cagggtgtg gcagtgggga tgggaggagt
17461 agttggagtc tggatatctt tgaaggaaata gctgacagga tttgctgatg gacaggatgc
17521 aaggggtaag aaacggggag gagtggagga cgccctgagt tttctgacac aacaaggttg
17581 tgccacaca gggcagccac ttgcaaactc taagtggaga acaatcactt gtcattgttg
17641 ctgatggcag tcttccagca tggctgtcac ccaactgagg tgcctccgtc tcttgctca
17701 acttttagcc tcttctactt cacttctctc cccatttcc atgtctcatg gtgcagccag
17761 actggccttc cttaacagaa ttagagcacg tcacccccc tctttaaag cctgaccttc
17821 cttatgagac aaatcccctt ccctccatgg agttagtcct tcaaacaca cctgcgtgca
17881 cactgtgtgc cagactgtgc agaaggtagg ctctgcacct gtccctccct ctagacctg
17941 cccacagca tttctccaag ccaagaactc ttggttggtt ctggaattta ccctgcacag
18001 tgacacctat gtccttttgc tcatgcggtt cccactggtt agaaattctc actcctttcc
18061 tcacctggca ataccattca ctgcccagag cccaccgcag gcaggcgcg tctcttccac
18121 gtagtctgcc ctgcatcttc ctccagagg ggtgctcct cttgctggcc tcccacagca
18181 gcctgctg aactgcacag cctcttaagg aggtggact tgcacccaa tctgttctt
18241 cattgtcagc catattggat taggatctga tctgttccag ctaggctggt tccacttttc
18301 agaactctc cccaggcag tccgaactag ctgctatcc attggctttc gttaggcaga
18361 ccaaaggctc cagaagccat agggctctca aggctaggaa ttgtccagtt ccactgata
18421 tccagaaggg aaatacagcc catggggcgt ggaactggga agattccaca gaggactggt
18481 gtctatgaga gggaccacaca gcaaccttgc taacaagtaa tacctaacag ttattcagcc
18541 tttccagaca ccagggtgctg tgetaagtat tttacatgtg ttcagccatt taatcctcac
18601 aataactcac ctcttagatg agaaaactga ggcccaaaca ggtgaaactc ccagccagta
18661 agtagcagag ccaggcttca gatgcagtta atctgggttc aaagtccatg ctctatctta
18721 gcaacatcca cgctctatcg tagcaactac accattgtga ctaaatatga aatataactg
18781 taccagaaac cagctgccct gatggcaaca catgagttgg gctctctcta atctgtgacc
18841 atataaaaat tattcatcaa aggtgaaacc taaaattaag acatggatca atatactgtg
18901 agttaatcac tgattctttt actcatgatt ttctctctag tagggaatca tgtcctagtc
18961 taggctcctt gagtgaccag gggtcggtag ctctcaaaag ccaccctgg atcaataaca
19021 gctcttgttt aaatacagat agatagcata atctcttcc cattaatcat ggattctgtg
19081 tttataaatt cacctgcttg ctaaaatgtc cttataaacc ccaaataat actggtggca
19141 ctttcatggt cacacacagg caggtacaga gcagggggag ctggatttgc atgatgggca
19201 tgttccatct gagatccggc gaggcaacct ctctcttctt gtttcatctc tgatgtttac
19261 aagtgtctc ttcgcggtct atttagtgcc gtgtgttctg catttttgtg cttttttag
19321 tgatgtcact gtttagagcg aacccaagag tagcgtgct gtctggcgtt cctgagctaa
19381 acaggctgtg ctgtgcctta cggggaaggt gcctgtgtga gacaagcttg gatgaggcag
19441 gagctgcagt gctgctggcc gtgcatttgg tgttgaaaaa tccacataat agcacattca
19501 gaaaaaggag aaggaaattg gccgattcat acatgaggct gtgatagaaa gggctaaagt
19561 aacacctatt gtgtgggatg gacgctgga agtctcttcc aacagaaacc cacacataaa
19621 agaaggttga tgaaaatgtt gtggccagag gtttgcaaaa acctaaacca gtatttcccc
19681 cttgagtgat gacctgtat tctctaattc agtactcac ggggctttat cagaagcaac
19741 tgccaggact accgaaaatc aattgtggac acacagagga aaatgcatgc acttcaaatt
19801 gaaatatata gacatgtaat tttacataga taatacatgt atttcaatat gagtgaagat
19861 ttcttctct cctcatcacc tgagcagtg cagcctctgc ctgaagttct gccctgagt

```

Figure 4 continued

```

19921 catggagacc taccccccta agttaggatc ctgagcatgt gcagccatgg cgcattggcca
19981 tgccgggtcta tgggtactggt ctaccctca gacaggcaga gttggggaca tgggtgtgac
20041 tagggaggga atcacagggtg ctgcaactgt ctctccccag agtcaacccc ttctgggcac
20101 tgtgattgat gggatgctgc tgctgaaac acctgaagag cttcaagctg aaaggaattt
20161 ccacactgtc ccctacatgg tcggaattaa caagcaggag tttggctggt tgattccaat
20221 ggtgagaagg cacatgtctg ttggaggggac tcaaccaccc ccatacagccc tcctgtctct
20281 ggtcccagga atgcttccct ctccgagctg cactctgagt cctgggcagc tgttcccttc
20341 agcaggagtt aacatttcca cggaaatctt ctccaggagg gcacagtttt caccggggta
20401 tcagggccct gggattccct ctcccagac tctcttattg tcttctatt gaagaatcct
20461 gaagtgtctg tgctgggagg gccagtagag aaaataactc tgcagatggg aaaacggagg
20521 gggcctagaa gggggaagga gtggaggagt cgggagaagc agtcgggggc ttctggctt
20581 cccacctcgg gctgggtgtg cgtaattccc ctgaaaatca ctcatgctct tcatcacctc
20641 attgaaaata tccaaagaaa gtggcttctt cttgagttac cattgacttc tagggaacag
20701 cagacatcag cgtctactaa gagagtggat ggtcagagga agatgaggat ggaaaagcta
20761 cctaatacgg tgctatgctc aatacctggg aacaaaataa tctatacagc aaactccaga
20821 gacccaaatt acccatgatc aaacctgcac agttaccccc aaacctaaaa taaaagttgg
20881 aaacaaataa atggaaaaag taattgagtt ctaaaagtaa agaagtgttc tccttccagt
20941 tgaagataaa caaaactgac cctggttaaa gcagtaggag agatttattc agtaatatga
21001 ttgcaatagg ggaaatagtc cagcgtggag tggactcaca tcctgaaac aaaagcctga
21061 tgacgtttca taggccaggt gtgccagggt aaacactgtg gtctatgggg agaggcttgg
21121 ttcatgtgac tagaccacct ggatgtcttc atcctggctc acctggggca gaaacaaact
21181 cctcttatct ttaggacagg aagccatgca ttagaccgca ggaagaacct actgaagtca
21241 ggctctactc ttctgacagg gactgggaag atcaggagtg agtccctga atgtttgcat
21301 tgcaaagaga aggtctcag gtcccaagg aatgggggtt ggggtggtaga ttcacatccc
21361 aaagggcaga gaaagtgttt atcattgcag gcttctaaag taactgctct agttgggttc
21421 ggggacctgg ctgcttgtct ccagggtttt gctggagcag agtaaattgt cctggcaggg
21481 ttgagccttc ccaggcagtt atgttaagga tactggggtc attctagggg cacagcctta
21541 agctgctaga agcaatgcta gagattgggtc gagtccctga gtgcagaggt ttgggaagag
21601 ttgttatgtg cgtagagatg tgcagttcgc cctcagaag gatttcagag gttataccaa
21661 agaccaaagc cacgaagaca ggatttaagg atttcagtgt catctgtgac aggtgggggt
21721 actagggagg tctacatcct ctcccaga gccctctatt ccacactggc ctggcagaca
21781 ccttgcatct ccatgtcact gaccactgta gggatcccag ccacagacac acacacacac
21841 acgcgcgcac acacacacac tgcaaaacac tgccacagct tctctgaaat tttgtttttt
21901 gttttttttt tttgtaaact ttgcaatgct tgagtttctg gatataagcc catttgattc
21961 attgatttca cccatttcag taaagactca tgcccttaa agcccccata attagaggac
22021 tttcagactg cattgggttg gttgggttgg cagtttgttt cttatcatta attcccaatg
22081 attcataaat gctgaacttt tttttttttt ttagcagttg atgagctatc cactctccga
22141 agggcaactg gaccagaaga cagccatgtc actcctgtgg aagtccatc ccctgttgt
22201 aagagtctag gaatcatggg aattggctga gaccagaga gggacaagga cttgccc aaa
22261 tcatagagca attaaatggc agaataaga ctggggcctc aggggttccc ccactttcc
22321 acctttccca cttcattttc caccctagcg gggagttgca cagggttat ggggttcacc
22381 attgaggcag gactttctgg tgggctggag aagctgcac gctcaccgg ggctggtggt
22441 cactttttga tctatttcag tgcattgcta aggaactgat tccagaagcc actgagaaat
22501 acttaggagg aacagacgac actgtcaaaa agaaagacct gttcctggac ttgatagcag
22561 atgtgatgtt tgggtgtcca tctgtgattg tgcccggaa ccacagaggt gagtccaga
22621 ggtcgaacgg gagggacaca aacccacga gcttctgtat ctgaccacac tacctcccag
22681 catagacaga catggaaaca gctgagggtc aggcctcaa ggctgattcc atatggcatg
22741 ggatgaggtg ctgtcttatt ttgggttatg ccagcagaag actgtgagaa aaaaaaatcc
22801 agaggcaagt ggttttttta aggtgatgat ccagataac agcagtaggg aggtggggaa
22861 gtgagagagg aaagagatgg aatccaatcc aagatgcgtt gtcaaggag tatccactgt
22921 ggacaactgg agcggggaaa ctcaaggaaa caccaagaat acagggtca ggggcatccc
22981 atctgagtgg caaggagacc gggatattat tcaccagctt ctacttgtca ttggttgaag
23041 gctcttgcaa ggaattgttt attcctgtg acttcgacct gctgcacaca agggcagagt
23101 agtctcttgt gaccagacaa aggcctcagg cctggagctg cagatgctgg aggtggaagt
23161 caggctggca tgcccaggaa gaggggatat aggtgagacc ctggcagcat ctgctgcaag
23221 ggctccatgc ctggagttat tgtgggacat agggctgctg taggacaaat attgaaagat
23281 gagtctgga agggttttaa ggtctgagcc agctgcaaac aagtttgta agtattggag

```

Figure 4 continued

```

23341 ctggataaaa agtgggagcc ctgcatggga tccagaggat tgtccaggac caagaaggcc
23401 agagagagga gccagggcgg gtgtgggaga ggtggtctgt agaatggcag tgtctgaggg
23461 tgggtttttat gcatctcagt gcgtgcagta gcatgcaagc acagggcaag ggcagggctct
23521 atgcaggacc tgctaccact gacccaaggc tgtgtgggtg gggcatgacg ttagggatgt
23581 gtgtgctcac cttagagagga gtgtctgttc ctgcgtgcc a gctagaggag attcacagga
23641 ttccttttct tccagctctc attcgcttag tctgcaagct aggaaatgtc ctctctctaa
23701 ccatgctgga agggttttta aatttagaac ttcataagca tgataaaaac tagctaaccg
23761 ttaactagca ggtgccctga cacacctttg cacaggaagg ggcaggtgct cataaccctc
23821 atttcttcac aggttcctga taaaaacctt taaccaagac cactggagtt gaatctgttc
23881 tttcttaaac ttgcccttgc cctatgctct gcgtctgaac tatatataga gttcccatct
23941 ccattttctg atggtttgag caatactgaa cttatgtttt tcatctgggt tcccatgact
24001 ttaagttcag ggcttttttt tttttttttt tttttgccag gactacacag attttatcta
24061 ctttcactta ttgttgtctc actctttttg ataaaataac atttctgaat ataaaattaa
24121 tatgtttaat tatagaaact ttcaaaaata tggaaacaca aaaagaagaa actatgtatc
24181 actcattttc actacctgga ggcaagtgtc ataaattttt tgtgttttcg tctaaagatt
24241 ttccatgtcc tatatacttt aaaaatcaat tttatatatt acttttgttc atagaacact
24301 atgttgtgag gattttctca atgttgatac aaactcatca aagcattcct tattccatga
24361 aatggctgtg ttaaatagac tttaatatcc atatcctatt agatgtttag cctggctact
24421 cacattctga tgctacaata acattgagct cagcctaatt ttcataagtc tctaaacaca
24481 tgtctaataa ttttcttagg attattttta gaagtaaata actaggttgg acaatgcaaa
24541 tgttttagaa ctttgggcaa atattgtaaa attgcattcc actgcataag aatgtggatt
24601 tcttaattct ctttctcatt tgggatgcta agattaaaaa ttagcataac acttccagtt
24661 tcataagcct tccccaccaa caagaagcac atcagcctgt ttaatctgc atgtcctcca
24721 tgactagccc tctcattgtc tctgcttctc ttattatgtc taattgtcat tacccaaatac
24781 tgtgcgctac tttcctctga agacttggtg tgtgattgta agaattgtat aaatatcaat
24841 ggcttttagc aaagagtttt acaaagtacg ttttgggcag catgctgcct ataataagtt
24901 ctatcattca caaaagctct ttcttctatg gttggctctg agtttctttg gaaaatgttc
24961 gaggggcaaa ggcaaggtea aacctcctt ttagcaagtt ttgtttgacc ttgaagcagc
25021 catttaacag gtggattaca gagccacaga aggatggcat ctcccaggt gcgcttctctg
25081 ctgagccgag ggtgcagtgc aggaggctga caaagatgag tgaatagatt attcttcatt
25141 ctcacatgt aatgacatgt gcgggtgggg catgaggtta gggatgcgtg agctcccatt
25201 gttaaatcct tttgtgaacg aatttttaat gatgttaaga aatgatggct ggaggcgtg
25261 gctcttgcca acactctcag cattttggga ggctgaggtg ggaggattat gagagcctag
25321 gatttcaaga tcagcctggg cagcatagtg agacctgtc tctacaaaac aatttaaaag
25381 ggaagtgatt acagtgtact acaaaaatacc atataataa tcatccccag ttaaatacat
25441 aaaagagaga tatgctaaaa taaaaaagaa gcattctctg agtggtagag ttatgagtga
25501 tacttttaat ttgattgtta aaattgattt ttaaattata taggatatgg aaaaatttta
25561 gaaggaaata caaaaaatgc atggcccttg cctcaggtag tggattatg agtgatatct
25621 agttttgctt tatgcgtttc catatttttg tttcttctaa ttattcctaa ttgtttttat
25681 tttccaaatt tcttactgta cccaaagggg aggacattgt aaattttctc caccagcctg
25741 gcattgccct gactctttcc gtcttagtgt cacttctcag tgggtatcat gaccagtgca
25801 aagtgtctgag gcacagggcc aggcggccca cgaggtagcc ctgaagggggc agcgatctgt
25861 ggctccaatg agccaatcat tgggaaactg gcattctcct ggggctggga caggagagct
25921 ttgttggtgg gagaggaggg cagagccgaa ggagagaggg caggggagag gcaaggtcac
25981 ccttgctcag cgcttctttt tctctgccag ggctttactg ccttaacacc ctaacactcc
26041 caggccaggc aggcagcaag acacatgggg acagtgagct ataaagcagt gagtgccgca
26101 ggttcttgag acacaagagg cagattgttc cctctctgca tatgtttgta tttaaatagc
26161 ctcacgttca gacaagcaac ttggatttgc ttcagtcttc tcaggatggg tgagaaaaac
26221 actggaatg ctcaaaacga ggcagaacag aagacacctg tcatggtgga gatggcatcc
26281 aagtgaatga acgcgagtct gttatcaatg aattatatcc ggtgtaacca gatgtagaga
26341 gattgtcttg tgtgagaatg caccagcagg caaaggggtt tacgaggcag agtcagggtc
26401 cctgcaccgg cagctgtttt aaatgaattg acccagggtt tccacattca ttcattaaac
26461 atttgttcat tgaacaaata tttggctcta attccgcttc cacctcagac agctatgcag
26521 cccggagcac atcactgagc tttcctgaga ctctgtttct ctactgtgct gacagccaga
26581 ggtttgctgc agagctggag accttggtag cctccagggt tgtgtctggc aacatgaact
26641 gtgcatgaat ccttgtcttc tcagatgctg gagcaccac ctacatgtat gagtctcagt
26701 accgtccaag cttctcatca gacatgaac ccaagacggt gataggagac cacggggatg

```

Figure 4 continued

```

26761 agctctttctc cgtcttttggg gccccatttt taaaaggtaa tgtctcttcc tgtctgtgag
26821 ctaggagttta gagacttggg ttccagtctt ggtgtggtgg ccttcagcaa ttttgcctc
26881 tcctggtctt agtttcctga ttctgtataa cagggttgaa tgtcactggg ttgtccgccg
26941 gcttcctccc actctaacac atgtgaggag ggatgtgtat tcatgtgccc tctgcacaga
27001 cgctcagaat cctgctgtgt gacatcaaga cagggtggtg tgggggtctc ctgcaggatt
27061 cactatgaca gggttgaatg tcaactgggt gtccccggc ttcctccgc tctaacacac
27121 ctgagggagg atgtgtatca tgtaccctct gccacagaca ctcagaatcc tgggctgtgt
27181 catcaacaag ataggcgggt ggacacatgc ggggccctgc tctaggttga gaagccacca
27241 atcaatttcc ctctcccagg agcgtgaga cagtgagtga gggtgagatt agaattggtg
27301 ttaaccctaa gtcagggtcta atgtgtgctt gtttgcctt ctgtctgtct ctaacaaaaa
27361 ctcataaacc tttaacattta gtacatttaa taattgttga tttagaattc acaggatgta
27421 tgaatttggg acaggagaag agaataattt ttatatggtt tatagtcagc aagggtggcca
27481 tcccacaggc tgggaagggt cctctggcca agcccagaaa cgggctcatt ggaggagggg
27541 ttgggtagga gctttatgtg aacggatttg cttaacatac atgttcaaca gggtacaggg
27601 ggagcaatgg atattcatga agacagtcct gacacatgtg tattaacaaa acatgtatgt
27661 aacatggccc atgttcacct ggtggtggag acctaatatt taaatgtatt acaattaggg
27721 cctgtaagtc caaaggctct ctccaggcac gaagctcagc aagtgaggag cttctgtaca
27781 ccggccagggt ccagtcctat gctggtgatg ttcttatctg gagaaagtta ctgaaatcag
27841 tctcttatgc aatcaaagca gtagttaagt ctggcaagtc agggttctgt ttgtctgcgt
27901 atcccagctg cagccgttgt tattgttttc cttttgctta gctccaggcc agtgctggtt
27961 tagctgctag agaaagagaa gcacctcgtg gcagtgagaa caaagtggat tctttttttt
28021 ttttattatt atactttaag ttttagggta catgtgcaca ttgtgcagggt tagttacata
28081 tgtatacatg tgccgtgctg gtgcgctgca cccactaact cgtcatctag cattaggtat
28141 atctcccagt gctatccctc cccctcccc ccacccacc acagtcccca gagtgtgata
28201 ttcccccttc tgtgtccatg tgatctcatt gttcaattcc cacctatgag tgagaatatg
28261 cgggtgtttg tttttgttc ttgcgatagt ttactgagaa tgatggtttc caatttcac
28321 catgtcccta caaaggacat gaactcatca ttttttatgg ctgcatagta ttccacggtg
28381 tatatgtgcc acattttctt aatccagctc atcattgttg gacatttggg ttggttccaa
28441 gtctttgcta ttgtgaataa tgccgcaata aacatacgtg tgcatgtgtc tttatagcag
28501 catgatttat agtcatttgg gtatataccc agtaatggga tggctgggtc aaatggtatt
28561 tctagttcta gatccctgag gaatgcgcc actgacttcc acaatggttg aactagtta
28621 cagtcccacc aacagtgtaa aagtgttcc atttctccac atcctctcca gcacctgttg
28681 tttcctgact ttttagagat gcaggcctga acccttgccc gacgtggcct taggtctcgt
28741 ttttaatttg gtgtcttatt gccacaaaga gtgtgttctg tcagaatgat gaccttcatt
28801 ttattgctga tgctggtccg gtggtgtcta aatcacaaaa gggaggaggt attatgaggc
28861 gtgtctgacc tcctgtccgt tcccaggcag gaacagagtt gaagggtttt cgaggctccc
28921 cttagccag agaggtctg ttccagtcagt tgggggtgtc aggattttat ttttagttta
28981 cattatacaa tagtttaaaa aatcctttca cgggtgtctc taagtcttaa ctctcagcag
29041 ttacagatca gatgggactc accccatggc tgctcagaac gcaccagcgc ctctgggcat
29101 ctactgtgc atgcttaggc gccttgccgc tctgttgttt ttcagaatgt gaattgtggg
29161 tgtgtgctgg ggaggagaa aagacccaag agagagcaac cctggaaagt ggggtgtctg
29221 tagaggaaac cctggggtgg ggcatttcc gtgacctga gcggggagga ggcaggcgag
29281 ggtcacagga gacaaaggca gccagccaat ggaatgacgc atgccattg ctgagcccta
29341 ggtcagatat gctgaaagggt aaagtacaca caaatccacc acgatcttct aaaagtgcag
29401 acctcagctc agtgggtgtg gggcatgtgc caagactatg ttccagcag actcgcagt
29461 gatgttgacg ctgcagtcca cggaccacac tttgagtaac aaaggctttt tttcttcttc
29521 cccacagagg gtgcctcaga agaggagatc agacttagca agatggtgat gaaattctgg
29581 gccaaacttg ctgcgaatgg gtgaggctgg tggcaaagac agagcacggc tggtagagggt
29641 gggggcggg gcatgcctat tgggaagggg cagcttctaa ggttctagt atcaaaactc
29701 tgacctgtg accatagcac tctgataatg agagctctct gcaaaccggag aggcgcgcc
29761 tggagatagt gagctctcca tctctggagg tatacaagcc tcttgacaga gataacttgg
29821 gcatcctcac acatctctga agattgttgg gaacacacag cagctttggg gcaattctaa
29881 ttgattctgt ttccagaaac cccaatgggg aagggtgcc ccaactggcc gagtacaacc
29941 agaaggaagg gtatctgcag attggtgcca acaccaggc ggcccagaag ctgaaggaca
30001 aagaagtagc tttctggacc aacctctttg ccaagaaggc agtgagaaag ccacccaga
30061 cagaacacat agagctgtga atgaagatcc agccggcctt gggagcctgg aggagcaaag
30121 actggggtct tttgcgaaag ggattgcagg ttcagaaggc atcttaccat ggctggggaa

```

Figure 4 continued

```
30181 ttgtctggtg gtggggggca ggggacagag gccatgaagg agcaagtttt gtatttgtga
30241 cctcagcttt gggaataaaag gatcttttga aggcc
```

Figure 4 continued

LOCUS 2011	CES1A3 33573 bp ds-DNA linear 14-APR-				
1	acagtcacctc	tgagctgcac	ggagacctcg	caggccccccg	gaactgtcgc ccttccagga
61	tgtggctccc	tgtctctgtc	ctggccactc	tcgctgcttc	cgcggcttgg ggtgagtcct
121	tctgaagtca	aatatgcggg	gcactttttg	aaatccttgt	tctgggccga actgggcgca
181	gatgcgtaga	aaggcaaaga	cacaaaaggt	ccggctccgt	ggcggggcgc gacctccgta
241	cttgaattg	gacttgagac	agctgagccc	agccagcctg	gcccgcgcgg agacacaggg
301	aggggtgaaca	tgctgagacc	gccctgcact	gcttgccccca	agaaggccgg cagccggcccc
361	acctcactcc	gcacacaaga	gaaatttagg	cagctcagtt	tattgtccca taattcgctcc
421	cgttatggct	cgataacgga	tagcggctca	ggagaatatt	aagagcatcg actctggagc
481	cagagtgcct	gggttcgagt	ccccgctggg	cgcttccctcg	ctgtgtgtac ttaggcaaga
541	gactgagcta	ttctgtgctt	gggttaccce	ctctctgaaa	cagaggtggc agtgctacct
601	acctcactgg	gctttgggtg	ggaggaaatg	ggtgggttgt	tctaaagcac tcttggagac
661	gcctgggtatg	tagaaaagcat	gaaaaagggg	tggataaatt	taaaggcgcg gtttaaaaat
721	gtagtgtatg	aatagaaaat	tcgagctgtg	gtcaggccac	tatatcagga cctcttgcca
781	ttaaaaagac	agtttcttac	tcatgttctc	aacaggaggg	ggacaccagt gcgcggggca
841	caaggggaag	cagcagagcc	tggggagggg	acagggcgga	tctgaggcca aggggctttt
901	actgtggctt	ccaaggatag	aagaggaaag	gcagggcagg	caggctgagg actggcaagt
961	tgggatactt	caggggctct	ggggcacagg	ggctgttcc	aattgttttg tacctagccc
1021	tggcatgatt	agggcaggtg	ggtagtggcc	agggcatggc	aaccgcgatc aagaggcggt
1081	tggagtgtgg	gctccaggtt	ggctggacce	tgagagggag	tttctccaag gtcaacaagg
1141	cctcagatgc	cagagcatca	gaatacagaa	aacaaaacac	atggtacata gaggggtgca
1201	gcgaggggccc	tcgaaatacc	cccatgccat	tttatagaca	tagaaactga aactcacaca
1261	gggaaatgta	cttgcccagg	gtcaaccgca	aaatcacgcc	taggtctgtc tctgggtcct
1321	gaatcctgac	agcccagccc	cgtctgcate	ctgtcttcc	tctggcagcc tgacactgcg
1381	agagctctgg	aactgtgggg	aagggcagag	gaagagagga	agcctatttg gaactgaatg
1441	ttcctttacc	tctggttact	cctggcaatg	ccatccattt	ccttcccttag gaagaccata
1501	caggtgcct	ccacttctc	agtttctctc	tctgtaaaat	gggcatgttg atgcctacct
1561	agcaaggctt	atttgaagat	taatttaaaa	aactggaatg	catgcctggc acatagtagg
1621	tgcccaccat	gttcgtgctt	catataacat	cagcagcaac	tggctctgtt ctcacacttg
1681	tttctgcctc	ccccatttca	cccacttttt	cacccttggc	aatctggctg ccattccacct
1741	catcctgttg	aagggaatcc	ctgagctgag	catgaccagc	ccagtggttt cttctctgac
1801	ttgactctcc	tccacctccc	tgaagtcctc	agcaatctta	gttgggttcc tggcctcttc
1861	ctccctgggc	ttccaggacg	tgggaacgtt	ctctccctgc	tactctgatg gatctttcgt
1921	cttggctgta	gtggtccctg	tgtttgatct	ttgttttttt	ctcttccctc tcacactggc
1981	atcttcaatg	aagccataca	aatccatggt	gtgtatgctc	actgtctagg agaacattct
2041	gccctgctcc	cccaccaggc	ttccttccag	ccttgaagtg	cactcgtcct gtggcgctcat
2101	cttgccatga	tatcccactg	acatctccac	cgagcctatg	taaacaggga tccaacatca
2161	caagccaggt	ctttgatctt	ctttcccttc	catttccccc	cacgtctagt cagggacaga
2221	gtctttttct	tttcttttct	tttctttccc	ttcctccctc	gtcctgctt gcatgcctgc
2281	ctgcctgcct	tccttccatc	cttcccttcc	tccttccctc	cttcccttcc tccttccctc
2341	cttcccttcc	tccttccctc	tttcccttcc	ttccttccct	cttctctgtc tccagaggag
2401	ccaggattac	aggctcatgc	taccacgccc	agctaatttt	tgtattttca gtagagaggg
2461	gtttttacca	tgttggccag	gctggctctg	aactcctaac	ctcaaagat ctgcctgcct
2521	tggccttcca	aagtgcctgg	attacagggt	agccactgtg	cctggctctc aagtctttca
2581	aaataccttt	tagacatcat	catatttcta	catatttttt	tcttttttta tttaaaattt
2641	taaaatagtt	aagatcggtt	gggaagattt	caggtgaaga	ccagaaatcc catacagtgc
2701	tgaaaagggtg	gaaatcacca	agaaaacagg	ccacatgtaa	tgccaccacc cattgctatg
2761	tcgataaaca	tgcttccaag	gaagttcatc	tcagtattgt	ttgtaattag caacaacaaa
2821	agaaaagaag	ctaaatattg	acattgcagt	gttggctatt	aataagatga gacatggatg
2881	ctgttggtga	actgggcttt	gcataaggag	tttccattgt	gattggatat tcccctctaa
2941	gaccaaactc	gaattgttct	gttctgttct	agccccacac	tttgaggggag gtgttaacat
3001	cttgtcttta	tccaaagcag	agcagctggg	aggacaagag	gctggaaatc atttccatagg
3061	aggggaagctc	agacaaactg	aggtgactgg	agccagatag	gacctgaggg gtgggcagca
3121	agatcaagtg	tctgaaaggc	cacaggctgc	agatcagatc	aggccagtat agtcccatga
3181	caggtagaac	cagaactgat	gcggaggctg	tgggaaatga	atttcagctc agtgaggaaa

Figure 4 continued

```

3241 cattgtctgt gctgaaagaa tggagcagac ctgcccaggg ccaccagaag aatcttcctt
3301 gacctgatca catcctttat ctggaacggt tcctgctgct tatttgtctg taaagaaaac
3361 tcaagcgctt agcctggcat tctaagccct ccctgacatg cgcggtctca ttccaggaat
3421 attttccact cccccagac tccacctggt gcatgtgggc catgccccct tccccaggca
3481 ggctctgcat tctgctgctc tgagcctcag caaatccct tctctctgct gtgtaccact
3541 ctctgcttct catctcccct cattcaatcc tcctctactg caagtccag ctcctccatg
3601 atgcccgcaa tcagaaggcc ttgccccctc ttctaattcc tagactgctg tatgggtacc
3661 tctcccagga cacggccact tccttcccag ctgtagcccc aggtgtatat gtgaccttc
3721 cctatgatgg tcagtgcctg tgagaatatg ggtcaccct tgttcatect ctcatccagc
3781 atgcggtcca gtgccaggat tctaattggat aaatgtttcc agagacttct aagggggaag
3841 ctaaagtctc tgttctttcc tagtagctat ccttcttgca tttatttttg gctggatgtt
3901 tttttgcctc caattctaatt ttgctcctta aaccagcttg atgagttagg aaggacattg
3961 atccctcatc cctattgtac atcaagagaa actgaggcct agagggttta ggtgacttat
4021 ttaaggtcac tcacttagag agtgggaaac ccacctgga atctgggtcc agccttttgc
4081 ctctgatgca tcctgatttc ttctccatat ccagcagggc atctgtcctc gccacctttg
4141 gtggacaccc tgcattggcaa agtggtgggg aagttcgtca gcttagaagg atttgcacag
4201 cctgtggcgg ttttctggg aatccctttt gccaaagcgc ctcttggaac cctgaggttt
4261 actctaccac agcctgcaga gccatggaac tttgtgaaga atgccacctc gtacctcct
4321 atgtaagcca ggggtggctg tggcatgtgt ccttggggat gttcgctca cagtgatgca
4381 ggaaggagtc aaggcagtc cctgatgggc tgatcctttg ctctggaatc cttaatgaca
4441 ttgtagatcc ttaagaacat tccagaactc tcacagcatt ctggagtcca ttatttaaca
4501 cgtttattga gcacctactg tgtgccaggc atgttctggg ttcttgggat ccattagtga
4561 acaaagcgaa gatccctagg ttcatggagc gtgcattcta gcagctggac agataaaaac
4621 agcatgcaaa ttactccatg cgtcatgtga tatgttagaa cggaataagt gccttggaaa
4681 caaagaagca gttgagcagg tgagggagtt caaaaggaag gatctcatgc attcctctgt
4741 cttgaacttt cacatagttg gggagagggc tgggccagga agtactgagg ggtccacttg
4801 gcaagtgtc aactctcagg tgtgccaatt cctgaggaga aaaatggcag gtgtgcccat
4861 tcacctggc caatctggga agaaaagccc aaaggttcta attggcctca tcaccactcc
4921 ctagggtacc aaggctcccta caggcatcag agcagacctc gctcatcttg gcagccaggt
4981 gttcgagggc tttccagcag cccgcctctg gctgccttct tttctttgct attttttaaa
5041 aaataatggt tactgttttt attttctgat tgcaaatgta atatacatt attacagaaa
5101 attcaggaaa tacagagaag gaaaagtaag aaaattaaaa tcacatacaa tatcccacca
5161 cccaaagaga acaactatca atattttaga tagcaaagtt ttatgatata atattttggt
5221 tgtattccag tttttttgcc tgttgcata ctctcttctc ttaaaaggga atgctatatt
5281 gagataatth ggatgcctgc tttgttcatt ataagtaatc atgagcatca ttccatgcc
5341 ttaaatagca cagattggca ttaacagctg ctttacattg tattgaacca taattaatct
5401 ccacaagctt ttacttgggg tcattaagtg acatatgatg ttttcaataa tataacatca
5461 tgaatgtgaa gccttatgca catatatgaa tatttttagga tagctgctca gaaacgacat
5521 ttctgagtca aagagtatat acctcttggt tgatttccct cccaaaatat tccaactaca
5581 ctctcccat tgatgtatga acatctccat ttggcagact cttggcaatg cttgggatta
5641 ttaaaaaatg tttttggcta ttggataggc aaaataataa taacaataac agcaatatta
5701 ttttctctct taatttgcat ttatctgaag actagtgaag ttgagttctt ttcctttct
5761 tcacattgga atgtcttttt ttttattatt attgttggtt tgcctcttgc acttggtgat
5821 ctgtaggaaa atttcccagg ttggaaatca agaagtcac tttgctggct ctgattctgt
5881 cacttgacag ctgtgtcatc ccagagctcg agtgcctcag tttcctcatg tacctagtgg
5941 gggcggtgat gactgccttt tggtagattg aggtctggga ctgttgctta taatatactt
6001 gtgtataagc accacatggc atggatcttc ttggctatgt gatgttggtc aggttctcta
6061 acctctctag gcttcagttt tctcatctgt aaaatggaaa taatagaacc tacctaaggg
6121 gactgttctg aggcctaaat gagttaatat acaaaaggta cttaaaacag tgcctggaac
6181 actgggagaa ctcaataact gttaccaatg ttcatctatg ttgtatcctc aagatcttct
6241 atcagggcaa ccacagcagc acttgccctg gataggcgga tggtagaatc ttggcatgtc
6301 agggctgcaa gacctatgta gaaatcactg ctctcactta gacctggcag atttgggacc
6361 cagaatgtga aaggaattca ccaagggtca tcccgtaaat gtgtgggaga gacaaaacta
6421 gagcacagtt ctctgatgt tgtcccaaaa agattttacc cctaagacat tgcaagaaat
6481 gtcacctccc agggaagatc gttgtattgg ttttcttccc tccttgatgc tgggagtc
6541 gaaggctcca ggaagggaag aggtgtgaag cccttctcac tctgcttgggt cttaggagac
6601 cttagttagt cccagcacc ccacctgaag ccactgatag cctcctaccc actacaatgt

```

Figure 4 continued

```

6661 cgtgagtctg taaatatcaa gtccatttcc aggettgaac ctcccagccc aaggcactgc
6721 cctagaacac acctcgctta gatctggctg aacttcaggg gttcttcttt cccctcccc
6781 aggttcaccc aagatcccaa ggcggggcag ttaatctcag agctatttac aaactgaaaa
6841 gagaacattc ctctcaagct ttctgaagac tgtctttacc tcaatattta cactcctgct
6901 gacttgacca agaaaaacag gctgccggta agccatggga tccctgggtc aaggcgtgtc
6961 agcgccagta ccctgcatct tctggggcag aatggaaggg agtgaaggca gcttggggag
7021 ggagctgcac aggccaggat tgtttcagag accccagggc ctccacagga aacactgggt
7081 ttccacacct cagtttcccc tcatgacaca gggactctgg gggagtgtgc tgtgcatctt
7141 taatgaacat cttaattgtt ccaagttccc tcatgctaata agaaggatga taatgattac
7201 ttttagttac tagattaaaa atctgatata cacagtgggt aagacatgat tttcaataat
7261 gttagtgggt gaaggaactt agagaaaagc caacacccca actcttcttt ttc aaataga
7321 gaaactgagg ccagacagg gaaggagagg tcacacaata gggtagtttt ggctcaagtc
7381 ttctgactct caaatcagtg ctttcgacag ggattagggc attgagcttt tatgtggcct
7441 tctccagta atcaacattc atctggggta actttcactg tccaggtaca ggtgttgagg
7501 gaagaacatg ggacaatccc agaagcattt ccttttgctt ttggccaaat atctgggatt
7561 ttttttacct aatttccctc cttaattgtc ttggcaacgg ctgatgaccc agtgagttag
7621 catttcccaa ataaatgaag ggcgggagat ggggcaggct gccagctgca ggttggaggg
7681 tggggagggt ctgctgtcaa aggaggccac tccaggggtg aaaaagaagg ggtgttgtgt
7741 ccagccagtc tcccaggggc accgtggcat ggtgcttagg agtgggggag caggtctgca
7801 aacatgggac tagaagttgg ccctacgtgt actttgt aag acaaagg aag tgggagaaac
7861 gtttactcat tcattcattc attcattcat tcattcttct attgtgtact cagtccccca
7921 tatacacagt tagaccaggg gcttcaggga aaggaggaac atgaatccac tccacgcagc
7981 ccctgccttc aaagatcttg ccatctgggc tgtagagaca tattctctct tcagcagggt
8041 attgcaggcc agcccttggg ctacgccttc acacaaagat gaatgcacac tcagtccttc
8101 cccaggaggc agccatggat tagatgtgtg ggagaaacag gggctctggg aacctggagg
8161 caggagaacc tctaggaatg gccacagggt ctagaactac acagtctagg tttaaatcct
8221 ggttctgctt ttgagtagct gggcatttct taacttttct aagcctctgt ttcctcaaaa
8281 tgagagtcac cacatatggg cagtcaaggg tgaccactaa gcttgatctc ggacctgggt
8341 taagtaggca gaggcaggct catacccttg ccctaacagt ggactcactg cactgtggcc
8401 ggcgctgca gtatgctggg cagctcaaca tgtgaactct ggagccagac cgctgggtt
8461 aaagtcctgc tgetaccact tttgagttgt aggaccttgg gcagggttaag tgctccatt
8521 ttcccacctg gagaatgcca cagtattagt atctaccaa tggatagttt attttactgt
8581 gcttagaaca gcacttggga catagtttat gttacgagtt cttttatatt tcatgaataa
8641 tattattaca atatcattat tttagccact ctcaattctt cccactatg tggactttta
8701 aaatgcaaga atcatgcatg aggaattgta tccactctca tccctcagta ttagcagagt
8761 ccctggcact tgctaagttc tgcacagtgt tagtccattg actgaaggca tccctgaagt
8821 ggtgggtttt aggttgggtt ctgcagggtg gataggattt ggggatgcag aggcataagg
8881 aggcatttga ggaagagaca ccatcacagg cacgtgtgct tctcgagaca atcattcatg
8941 cacagagaag tttacttcac caagagatag ttaactgaga tgaagctgag atacaagggc
9001 tgccccatgg gtacgctgag tgcataata gtccaggctt gaggggtgat ggagtggggg
9061 ggagaagagg acatccttag ctttgctat gcccatgat atgttctcag catcaagagc
9121 ctttcgggga ggaggcactg gaggttgggt agaaagg aag ggacagccag ggtatgtgca
9181 gtaggggtgt ggcttgggtc ccagttggga gaattatgca ggagagagat tgccttttgc
9241 aaagttgctg ggagctggtg aactttgggt agcatcctcc ttcagttgta accaacaggt
9301 aggtcaccaa acaagacatc ttctgagtgc tccctggctc tgtgtcctgt aactggtagt
9361 ttgcttctc tctgccagg tgatggtgtg gatccacgga ggggggctga tgggtgggtg
9421 ggcataacc tacgatgggc tggcccttgc tgcccatgaa aacgtgggtg tggtgacct
9481 tcaatatcgc ctgggcatct ggggattctt caggtaagaa attggactct cctcactgca
9541 ctttgcccc cagaacgagg atgctaggac ccagctcttg tcatgccagc cctcagggga
9601 gcttagctag gttccacagt aaggcatcca agcccttca taattggaca ctacctacc
9661 tctcactacc cagccactca tccacttgcc tctgagcttt tgcagtgct gttccctctg
9721 cctggaatgc ttgttctatc cgaagaacac ctcttcatcc tgcagacccc agccccaggt
9781 tacctccact gaaagactca tctcctcttc acctatgttt ctccagcaca ttgcttttct
9841 ctgtcatgac acttagcagg cagcccagca ggtgaagggt gacacatctg acttctcatg
9901 gaaaggaggg aaggggcaga gtcacagagg gagctcagg catgttgtgg gtgaagg aat
9961 ggggcctgca gatggtgggg tgtgggagca tctgaccct tctgtagga gtgttgggac
10021 ccactcacc cattttctgc tcgaaatttg gcaagcgcg gaagcaagga tggagtcttc

```

Figure 4 continued

```

10081 cccatcaggt ctccatgggg acaggagtea gatctccttc aagggtgcaat tactatgaac
10141 cttttctaag tgttaacatc ccagcacctc aatcccaggt gagtgatgcg agggttgtgg
10201 gagggattcc actcactgat gagaaagcct agggagatgg aggttgagtt ctggtctacc
10261 catcagcaag gcatcccagg ccagctcttg tggatcacc ttcacccctg tcagccacag
10321 cctgtgtctc tgcctcccat tacaataaag ctgcttcatg atctctctcc acttctctc
10381 ctccgtgtgt ctccctccacc catgccagtc aggtttcttc ctgaacacac cactaggctg
10441 ctctcattgt ggtcaccaat gacctcccc tgtgctccac ccaacctcac aggacttccc
10501 agcagcagca gatggccgct cagcctctgt gacatgcttc ctcccagcc cctttcctgg
10561 tccccctcc tccctgtcca caggcctctc ttgtctccct ctccctccct actaattccc
10621 agtgatccca cggagcagcc cagagccctt gtttccagtc cacctcacct ttgaagccca
10681 cagggtgtcac tcactgcctc ttccctgccc ctggctgttt aaagggtgcc ccactcttcc
10741 catggacaag ctgaatcctt gactcccact gcccctcccc aacagagacc cctcctcct
10801 ctaccacac ctccactcag ctgctgctgc cacacctcc cttgggctca ggagaaggcc
10861 ttgtgggcag cctggactct ctctttctc acaacctct tggctccacc ttcagaatcc
10921 acttagaatc cactactttg cacctcttcc ctggctccta cctagccaa agccttctgt
10981 ctgtcacctt gattatgatc ttaagactta ctttcagatc tctgccactt tcttactgag
11041 gctgagtcgc acagtaccga gagtgatcac gtaacatgtg cgttgaatca tgtgatgcct
11101 ctgcacaaaa ccctgcaggg acttcccctg ttccctacca tggcagccga agttcttgcc
11161 agggccccag ggctctctct agctgcacct cctgacctac actttcttcc actgggcccc
11221 agggccattg cacttgcttt ttcttctgcc tgggaattcct ttctttggat atttacaagg
11281 ttcatgtgct cgttgtgctt gttttatgaa ggtctttgcc tccaaaacct caaatgccac
11341 tgcctccaaa aggccttccc tgattgggat ctgaaatggc acccctgtc ccacctctgt
11401 ttttaccctt gttccatttt atttacctca atgcactttg actacctgac atcattatat
11461 tctctctata tatctgttta tctgcttggt atctgtagtg ggagctcagt gtcacacag
11521 gttttgcctg ttttgcctc aggtgtgctc tagctcatag gacagtgtct ggcatatagt
11581 agatgctcag taaatagttg ccagttgagt gaatgtagaa ccatacatca ccacaatgct
11641 gtactaatga aagcagggtc tcagagatgc tgaagcccag ccttattctt aaaggcttat
11701 tgatagcccc cagccgcctc tgtagtcctg aaagtctgc catgatctct ctgctcccca
11761 cctcaacct gttctcttcc ttacagcaca ggggatgaac acagcccggg gaactggggt
11821 cacctggacc agctggctgc cctgcactgg gtccaggaca acattgccag ctttggaggg
11881 aaccaggct ctgtgacct ctttgagggg tcagtgggag gagaaagtgt ctctgttctt
11941 gtgagtattc ctggcaccgg ccccaacccc aggcttgatg tgatctctt tgatctctt
12001 ggacacctgt caatccagct atgcacactt ctgattacag tctcaggact ggccctactc
12061 actgcccatt cagggggtgg aaagctaaag accaagctct ccttggccca caggaagctc
12121 agagctcagc ttagtgctcg ggggccatat gaatctctag ttgcagcctg caacggtggc
12181 attctcctct tctagcttga ttgagctctc tctctctctc tcttctctc tctccttccc
12241 actcatttta acataaggac tcacttccct tctcttggga agtagggata ggcataaatt
12301 atgagtaagg gcaggcgaag acaagagggt gacagaggtc attgttttag ttaaaccaga
12361 actatgagtg gaggggacat ggacactgag caggctggga attcctctgg ggtggtctca
12421 tggctggctc atgcccaga aggaggcaac agtttctgtg actgtgggac cgggtcggct
12481 aggggaggcg ggtgagagga aggagggatg gagctgggga tggggaggga tggggcaggg
12541 gttgtgggtc atagacacaa cccggggtgt gaggggtcct cctggggttg gggattgggt
12601 gcgggagaca ctgggggatc tgggatgaaa acccagatga gaggtgccgg gagctgtagg
12661 aagacttctg cctgcctcga ggtgggcaga gggtcagccc actactggat tectcaatcc
12721 tgtgttggtt ttatagtggg gtagatctag cctggaagag caagtgagtc actgacccca
12781 ctctgagca tgaactctcc tcccctccac tctgctgtca ggttttgtct ccattggcca
12841 agaacctctt ccaccgggcc atttctgaga gtggcgtggc cctcacttct gttctggtga
12901 agaaaggtga tgtcaagccc ttggctgagg taggtctcgg gctggtacgt ctccgctgg
12961 acaccccccac ctcttggct ctatgctcct gaatcctcag ggatctctct tgtggtcgg
13021 tgtagctaatt gttctcctag aatcactgag gcaccaatgg ctgagcagga agggcgagga
13081 gacaccttga tcagcgtccc agtttcacag ccaggcaaac cgacacaggg cttggaaggg
13141 atttgccaag ggcagcaggt gatcagggca gagctgggac tccagctcat ggccctagca
13201 gccagtacag tgccctgtct gtgaccacac tccacctatg tggcagggcc tgggtgccata
13261 ttgggcagtg atggtgtctt gtgtctctca gggctgagtg tctgtggacc cacttgtggg
13321 ctgtgggcct gaagcagttc cacaatgagc gcctgataac cagggttggg tcttgagaa
13381 ttcactcatt gattcatttg ttcacacac aaactaggt gactaagtga aggcaaaaac
13441 aagaaatggg cagacgtcat cctttggctc aaagccagat gtccgtgtgg aggggacata

```

Figure 4 continued

```

13501 gacactgcat ggccctatgt ggctctgcat tctagtcaga cacctaacac ctccccaaagc
13561 ttctgctata atgtagggaa tagatgaata gctacagaat cacacaacta gaaagtgtca
13621 cctatgacaa gagcagtga ggtgaggtac ttgctgccac agcagaatct aaagaaagca
13681 ttgagtcctg gggctggaag aggttaccag ggaaggcttg ctggaaaagt gactaatgag
13741 tcaggagcaa ggaacgtcca gggagtggga gcagcacatg cgcgtgctgt ggcaggagga
13801 cgcattccaa tcgggaggga gagacagaga cagagccgat ggggccagag caggcagaac
13861 aggcggagcg cggcgagtac tgacagaggg gacgttggga ggggccaccc tgcacaggac
13921 cctggcaagg attttgtcat catctggaga gtggttgaaa gccaaaggaa gaggtgatcg
13981 ataggaatcc agacctagge tgaggatcgc ccaactggagc cagtggcatg gaggatttcg
14041 gtagctttga aagcttgttt ggggaaagca tccaaattta aagggtcgtt acataggagg
14101 agagaaaatg gggatgccaa gaatttttag aatttttgag aattttttaa gaattcattg
14161 gttataagca acagttgccc attgaccaga cttaagtcaa gaaggagcat tagcctggtg
14221 tgggtggctca tgccgtgaat ccctgcaatt tgggagaccg aatgagaagg attgcttgag
14281 cccaggagtt tgagaccagc cagggcaaca gagtgatecc ctgtctctac aaaaaaaaaa
14341 aaaattacgc tggtcatggt ggtgcaagcc tgtgggccc tctacatgga gtctgaagag
14401 ggaggatcac ccaagcccag gaggctcagg ctgcagtgag ctgtgtttgc atgacggcac
14461 tccagcctgg gcaacaaagc aagaccttgt ctcaaatct tgtttataaa cggagaatta
14521 atttttcaag gaaatagagg tctgggatga ggtgggctt tgtgaaaaaa acacgaaaca
14581 gaaattacaa aatgatggga atcctagact ctcttattct ttttcctct tctctcctct
14641 gtctctctct ctgtctttct ctctctcgcc ctctacttct cctgtttat ttattcaaga
14701 aacaccctgg gtattgattc tgtcccagat cctacacgag gctcttgggg catgtgcact
14761 cgccctggct cccccaccat ggagcctgcc tgcctcctct ctgtgccctt atttctggct
14821 gcactttgct tccagtctgc atgagcacc tatcagcccc tgagcttaag gactctcctt
14881 tggttccaga aagtagtcag gctgtctttt ctggttgtaa aaatgagcgt ttttacatta
14941 gcagctctcc accctgattc caggctcactt aggagagagg ctactgtct cgtctgtgct
15001 tcagtgtccc tggattctta gaaaagaagg aaaccagagt gtgcttggca gactgacccc
15061 aatccccagg acaggcagaa ggcagcctct cctcaggag ggtttttttt attttaatat
15121 tttttacatt ttttaaagtt ggaaagaggt tcttggtgca ttaattggtg ttgctctgtg
15181 gcttcactta ggacaaaggt gaacagacca tcaactcatg ccaaatctgg atctgtgcct
15241 gcatggccct tgagctaagg atggtttttt aaattttttg atgattgaaa ggaatcaaaa
15301 ggagaataat atttcatgaa ccataaaatt atatgaaatt caaatttcag tgtccataaa
15361 ggaagttttc tttgaacaca ttgacactca ttcatttacc tcatgtctat ggctgtgttt
15421 catgtcaaca gcagtgttga agttacagca gagaccacgt ggccacaca gccctccaata
15481 tttactattt ggctttgat gggcacattg ccgacctgtg ccctaaagag tctgtcttga
15541 ttgctgcttg ggactctagt acacaaatgt gcagggtgga actcctggcc atgaaacccc
15601 atctatgtct tgcaaaggat gacaccctgg tttttttcag aactcacctt taaatcacat
15661 tcgctacctt ctgaagcacc tctgtgtgaa tcatctgtc taggcctcct acaattctgc
15721 acatgtttat tgtgtgggag acatggcact ctcccacatg tgttaactca tgaggtaggt
15781 gcagtcatta ttcccatcca tgagtggaga aaccaaggcc caaagaaatc aggtggactg
15841 cctgcagcct aatggctggg aagtggaaaca accaggattt gaaagcagac aaggaaattc
15901 aggattccat gctctctatc ccagctccac cttgcctttc atgacacatt cagctggaga
15961 tgattatgtc cgtttcaatg gagataaggt agcagagatg agagattaca taattgaccc
16021 atgttacaat tgagattagt aacagaggca acgatctgtg ggacctggaa cccacaccca
16081 tgggtctaca gtatcttctg ctgccccctg ccaactggtac aagttgggct tgggatagaa
16141 tgccactttc catgttgatg gagggaaggg acttcaactct tgaactctgt tgctgtgat
16201 ctttgacgca aattgctatc actgctgggt gtaaaaccac cactcagct gtgatggttc
16261 actgctgctg acagaagacg gaagaggagc tcttgagagc gacattgaaa atggttaggtt
16321 gcctgttccc gtagcccaaa cactgtaaac ttggtcccag acttcttcat ttcagcggctc
16381 ctcttgccct gggacagtta cctgggacaa tttctcaagt ctgggagtc tcagtatctg
16441 aatggggaat ctaatttgtc ctttttaatc gtaaaatggc acaaattgtat aaaaagaaag
16501 tcaaaaaata acatgataaa gaattgacca aatttactat ttgatgaaaa tttaatattt
16561 tgccattctt gttttcagat gtatttttga gaaactaaac attacatatt cccaggggac
16621 accatcatgc tgaatttggg ggatggagat attaagctca aagattttca gaaagatgtc
16681 acaatttatc ttggttgact tagaaactgt ctgtattaga cctgggtggg gtccagttac
16741 atatattttc tgagactctg actcatttat tgttctagga tagtcatttg tccactcaat
16801 cattaatcca tctactatga tcttcttctc catctgtatg ttcactaatt catcgctttc
16861 acttatctat ttttatcaat ctaatatcaa cctgttcgta actttttatg tatctatttt

```

Figure 4 continued

```

16921 caatccatcc accattcatg gatcatccag ccacettata tctcaactcc atcaccatt
16981 cctccaaaat caacaatcca gttatcgct gtctgctagt ttccaccat ctattcatgt
17041 atccattcaa ttcaactgta ctccatgtat tgaccaactc catccatccc tccattgatc
17101 catccatcca tccatgctaa atatgtagg gtgggtgtta gaggtagcaa aacagacatg
17161 aagtggacat agtccctgct ctcaaggaac tatccaaaga gaaatacatt catatacttc
17221 gcagggttgag tatgggtggca gcacagagg gaagcattca ttgtagtaat cagggtatgct
17281 tcacagagaa ggcggaggag ccagatttga gaccagacag tgggaattcag gagttcatgc
17341 ccttaaccct gactccacct tatctttcct gagaaattca gctttcagat cattgtgccc
17401 attttaacag aggtaacaaa gacagataaa ttttgaatt gccccatgtc acagttctaa
17461 ttagtcagag gcaacactgt gggacacaga gccacacct gcgggtctag agtatcttct
17521 gccttccctg ctccctgcca tctcacata ttaggacctg gggaccttta ggattgatgg
17581 atggctgtgg cagtctctca tcacaggga gccacgagg acaaacaccc agatgacaca
17641 gcacccaggg ccgttgggaa ctattccctt tgaggagaa gagtgtgaat tccaagctca
17701 ggagtgcct ggactctctc tttggacctg aggtactcc tgggtcccag ggcggcctc
17761 taccactgga ctctgcttgt gtttccatgg gttgagtcca atgtggtctt ggggttagg
17821 tcttggttca gtctctaagt gaccaaagt ttcaaggaa taaaacacac ctgttttcta
17881 atgccaggaa ggaggcaaat attccagcaa ttctgcatgt ggcacagag gaccagctt
17941 aaaaggggag agttgagctt ttgggaaac ctatccctaa gatcctggaa catccttttg
18001 agtttttctg agatcttgtg gaagtgtct catgattacg cattgcctcc agtatgcaca
18061 catgcagaca cacagacaaa cacacataga cacacacaga gaaacacata caggcacaca
18121 cacacacaca cacagacaca cacagacacc cacacacaca aagagacaca cacacacact
18181 cttggctagt gggacttcaa tcttaatgg agggacatgc actggttacg tccggagcac
18241 aggtcagatc ttggaactac aatgcaacac tacagtact gcaatcttgg tgaacacaca
18301 ccaaagagaa gcatggggtg gggtagtct cagttagctt gcatgagtca gggctttctg
18361 ataggagatg aggaaactcc cttaagtacc ttgaacagaa aaagaaatgt gttgcaatag
18421 catcctcatg taaactgaaa atctcaggag ttgatggctt caggcatggc tggatccaga
18481 tgatcatagg atattattga gaatctacca ttctctatcc ttcagctttg cacatttctg
18541 tgttggtttc actcttaggt agatgtgtct cctctgaagg ttaaagacaa cagcactctg
18601 gacttatgtt ctatctgctt aaaaacctg tttctaaata attcctgcaa aagttttggg
18661 gtaaaactatt ggattgacat ggggtgatgc ccacccctga actaatattc atgttcaggg
18721 agatggagca cccaggtca catatagata tatgaattca tggagttagt ggagaagaag
18781 gaggtctctg aaggagaatt ggtgagttct tcccagggtg agaggaggatg gagaggatg
18841 gggggccagg cagcctcaac cagcgggggt cctctccaga gctccatcga ccctggctca
18901 gagtccccgg cctgtcttcc ccattgctgat gccaccctcc cctgctatg ccttcaccct
18961 gtgcttaggc ctacagagc cagagtctg cccacttct ttcagccag ccagcttggc
19021 accaggaggg agaacctgac acccctgttg cccactcac ccagctcagt gttctcctgg
19081 gaagcctctc acccacatcc tctgcttttg tcttcacaga aattcttacc tctggactta
19141 caggagacc tcaaagaggt aaggacctt tgtttctgga ttacgggttt tgagtcttag
19201 cacttttaag ttccaatgaa ctgtgagtga aagaatcccc tctcggtgg ttatagtaac
19261 tctgtgtgt ttgttgccga gggccagaga ggggtagtga ctacctggg taacacaacc
19321 aggaagacta gtggctggcc tggaaacct tttcctgact ccagtcacag tgctcccagg
19381 catcacctct gtgtgccctg ggctctgccc actcccctgc ttttttacct tttctactcc
19441 caaaaatggc actatggaag aaggctgaac cagaatcttc cactctcagg aggcagagat
19501 aatagcgatg attagccatg gagggaggaa gctaaaatct cagtccaagg acactcgctc
19561 cctcccagca cacaggaaac tccaacagta tctcctgat ctctcacc ccacccccac
19621 ctcccaatgg gttgacagtt tctggtgaca tcacctctga cgaatcttac aatcctgtcc
19681 tctctgctgc ctccctggag gccgcagcac tctcctaaat ggtcatgggc ggagtacatg
19741 agattcattc agcaaagata ataaacactt cctatgtgcc aggacttgtt taggagctgg
19801 ggatataaca gtgaaaaaag aaagactccc tgccctcttg ggaccttct ttacagtga
19861 aggacagtgc aataaaaaag caaatcaagt atgtattcca tttgatgggtg tgcagcatta
19921 ggggatagag tttgtgggtg gtttcaattt cagtagcagg atcagggaat ggttcaatga
19981 gaaggtggca tttgagccaa gctctggagg aggcaggaa ccagctgtca ggaaacctgg
20041 agaagcctat tccaggcaga gggaatagtc actgcaaaga cctgacaga aggggccagt
20101 ctggctggag aggaccgagt gtggggacag actggcttga ggctaaagag gtaattggca
20161 gtgggagaga gatcaattaa agtacttggg tttggattca gagcaagatg ggaaaccttt
20221 ggaggttttg aacagaggag tcacatgttc ttagatctca caaaggctct tgctccgggt
20281 gttcaaatag actgtaggaa aggggcagag tggatgcagt tgaccagctg ggcagccact

```

Figure 4 continued

```

20341 gcattgccgt aatccaggca agggcaaggg ctcttggtg cttagacagg gctgtggcag
20401 tggggatggg aggagtagtt ggagtctgga tatctttgaa ggaatagctg acaggatttg
20461 ctgatggaca ggatgcagcg ggtgagaaag agagaggagt ggaggacgcc ctgagttttc
20521 tgacacaagc aaggttatgc ccacacaggg cagccacttg caaactctga gtggacaaca
20581 gcaatcactt gtcattgttg ctgatggcag tcttcctgca ttgctgctgc ccaactgggg
20641 ctctcctgtc tcttgccctca acttttatcc tcttctactt cacctccctt ccccatthcc
20701 acgtctcatg gtgcagccag actgtccttc cttaacagag ttagagcacg tcaactccac
20761 ttcttaaaag cctgaccttc cttatgagac aaatccctt cctccatgg agttagtcct
20821 tcaaaacaca cctgcgtgca cactgtgtgc cagactgtgc agaaggtagg ctctgcacct
20881 gtccctccct ctgaccttg cccacagca tttctccaag ccagaactc ttggttgttt
20941 ctggaattta cctgcacag tgacacctat gtccctttgc tcatgcggtt cccactgttt
21001 agaaattctc actcctttcc tcacctgtca ataccattca ctgccagag cccaccgcag
21061 gcgggcggca cctcctccac gtagtctgcc ctgcacttcc cttccagagg ggtgctcct
21121 ctttccggcc cctcacggca gccctgctg aactgcacag cctctcaagg caccagact
21181 tgcaccccaa tctgttctg tgcgtcagcc atgttggtt aggatctgat cttgtccagc
21241 taggtctggt ccacttttca gaacatctcc cccaaggcag tccgaactag cttgtcattc
21301 attggctttt gtaggcagac caaaggctcc ggaagccata ggggtctcaa ggctaggaat
21361 tttccagttc catctgatat ccaggaggga aacacagccc atggggcgtg aaactgggaa
21421 gatggcacag aggactggtg tctatgagag ggacccacag caaccttgct aacaagtaat
21481 acctaacagt tattcagcct ttccagacac cagggtgctgt gctaagtatt ttacatgtgt
21541 tcagctatth aatcctcaca ataactcacc tcttagatga gaaaactgag gcccacacag
21601 gtgaaactcc tagccagtaa gtacagagc caggcttcag atgcagttaa tctggtttca
21661 aagtccatgc tctatcttag caactacacc attgtgacta aatatgaaat ataactgtac
21721 ccagaaccag ctgccctgat ggcaacgcgt gagtcagact ctctctaate tgtgaccaca
21781 taaaaattat tcattaaagg taaaacctaa aattaagaca tggatcaata tactgtgagt
21841 taatcactga ttcttttact cataatthtc tctctaatag ggaatctcgt cctagtctag
21901 gtccttgtag tgatgagggt tcagtacttc ctcaaagcca cccatggatc aataacagct
21961 cttgtthtaa tacagataga tagcataate tcttctcat taatcatgga ttttgcgtht
22021 ataaattcac ctgcttgcta aaatgtcctt ataaaccca aatcgatact ggtggcactt
22081 tcatggtcac acaggcaggc acagagcagg ggaagctgga tttgcatgat gggcatgttc
22141 catctgagat ctggtgagge aacctctgtc ttcttgthtc atctctgatg tttacaagtg
22201 tctctctcgc agtctattht gtgcctgtgt ttctgcattt ttgtgcttht tgtagtgtg
22261 tcaactgttht gagcgaaccc aagagttagc ctgctgtctg gcgttctga gctaaacagg
22321 ctgtgctgtg ccttacgggg aagggtgctg tgtgagacaa gcttggtatg ggcaggagct
22381 gcagtgtctg tggcctgca tttggtgttg aaaaatccac ataatagcac attcagaaaa
22441 aggagaagga aattggctga ttcatacatg aggtgtgat agaaagggt aaagtaacac
22501 ctattgtgtg ggatggagcc gtggaagtct ctttcaacag aaaccacac ataaaagaag
22561 gttgatgaaa atgttggtg cagaggthtg caaaaacct aacctgatt tcccccttga
22621 gtgatgaccc tgtattctct aattcagtac tcatgggggc tttatcagaa gcaactgcca
22681 ggactaccga aaatcaactg tggacacaca gaggaaaatg catgcacttc aaattgaaat
22741 atacagacat gtaatthtac atagataata catgtatthc aatatgagtg aagatthctt
22801 cctctcctca tcatctgacc agtgtcagcc tctgcctgaa gttctacccc tgagtccacag
22861 aggcctaccc ccttaagtht ggacctgag tatgtgcagc cgtggcccat ggccatgccg
22921 gtctgtggtg ctggtctcac cctcagacag gcagagthtg ggacatgggt gtgactaggg
22981 agggaatcac aggcactgca cttctctctc cccagagtc accactatth ggccaccgtg
23041 attgatgggg tgggtgctgt gaaaacacct gaagagcttc aagctgaaag gaagthccac
23101 actgtccctt acatggctcg aattaacaag caggagthtg gctggatgct tccaatggtg
23161 agaagggtga agcctcttg agggactcac ccaccacct cagcctcct gtctctagtc
23221 ccagaagtgc ttccctctcc gagccacact ctgagtcctg ggcagctgtt ccttcagca
23281 ggagthtacc tttccatgga aatctgtctc aggagggcac agthgtcacc gggatatcag
23341 tgccctggga tccccctct ccagactctc tgattatctt cctgttgaag aatcctgaag
23401 tgtctgtgct gggaggacca gtagagaaaa taactgtgca gatgggaaaa ctgagggggc
23461 ctgcatgggg gaaggagtht aggagthtg agaggcagtc gggggcttcc cggttctctg
23521 cctcggtgtg ggtgtgcgta gttccctga aaatcactca tgccttctat cacctcattg
23581 aaaatatcca aagaaagtgg cttcttcttg agthaccatt gacttctagg gaacagtaga
23641 catcagggtc tactaagaga gtggagggtc agagggaagat gaggatggaa aagctaccta
23701 atcggtgtct atgtcaata cctggttaaca aaataatcta tacagcaaac tccagagacc

```

Figure 4 continued

```

23761 caaattaccc atgatcaaac ctgcacagtt acccccaaac ctaaaataaa agttggaaac
23821 agataaatgg aaaaagtaat tgagttctag aagtaaagaa gcgttctcct tccagttgaa
23881 gataaacaaa actgaccctt gttaaagcag taggacaggt ttattcagta atatgattgc
23941 aataggggaa atattccggc gtggactgga ctacatccc caaaacaaaa gcctggtgat
24001 gtttcataag ccagggggcg caagggaaac aactgtgggt ctgtggggag aggcttggtc
24061 catgtgacta gaccacctgg atctcttcat cctggctcac ctggggcaga aacaaactcc
24121 tcttatcttt aggacaggaa gccatgctgt agactgcagg aagaatccac tgaagtccag
24181 ctctattctt tccacaggga ctgggaagat caagagttag ctcttgaat gtttgcactg
24241 caaagagaag ggtctcacgt cccccaggaa acagggttgg gtgatagatt cacatcccaa
24301 acgacagaga aagggtttat tattgcaggc tttctaaagt aactgctcta tttgggtcca
24361 ggggacatct gcttttctct aggttttcgc tggagcaaga gtaaattatc ctggcagagt
24421 tgagcctcca caggcaggca tgtaaaggat gctgcggcca ttctagggac acagccttaa
24481 gctgctagaa gtgaggctag agtttgattg agtctgtaag tgcagagggt ctggaagagt
24541 tgttatgtgc gtagagatgt gcagttcacc cctcagaagg atttcagagg ttataccaaa
24601 gaccaaaagg atgaaggcag gatctaaggc tttctgtgtc atctgtgcca ggtgggggta
24661 cttgggaggt ctgtgtcctc tccccagag ccctctattc tacactggcc tggcagaaac
24721 tttccctctg tgtgtcactg accactgtag ggatctcaac cacacacaca tacacacaca
24781 cacacacaca ctgcaaaaca ctgtcacagc gtctccaatt ttttttttat aatctttgta
24841 atgcttgaat ttctggatgt aagcccgttt gattcattaa ttttgccaat ttcagtaaaa
24901 acccatgccc ctgaaagccc aacaaaatth gaggaccttc agactgtatt gatttggttg
24961 ggtggtgaag ttgtttcttc tcattaatth ccaatgactc ataaacgctt aacttttttt
25021 tttaaaacag ttgatcagct atccactctc caaagggaaa ctggaccaga agacagccat
25081 gtcactctta ttgaagtcct atcccctctc tgtaagaatc taggaatcac ggaattggc
25141 tgagacctag agagggacaa ggacttgccc aaatcataga gaaattaaat ggcagaatga
25201 agactggggc ctgagggttt ccccatctt tccaccttcc ccacttccact tcccacctta
25261 gcggggaggt gcacagggtt tgtgagtttc accactgagg taggccttcc tgggtgggctg
25321 gagaagctac atcactcacc aggggggtgt ggtcacttta ttgatctatt ttagcacatt
25381 cctaaggaac tgattccaga agccatttag aagtacttag gaggaacaga tgaccctgtc
25441 aaaaagaaag acctgttctt ggacttaatg ggggatgtga tcttgggtgt cccatctgtg
25501 attctggccc agaaccacag aggtaaagtc cagaggtcga acaggggagg gacatgaacc
25561 ccactgctt ctgtatctga cccacttacc tcccacata gacagacatg gaaacagctg
25621 agggtcaggc ctcaaaagct ctaccatata ggcattgggt gaggtgctgt cttatttttg
25681 tttatgccag cagaagactg tgagaaaaag aatccagagg caagtggttt tttttaaggt
25741 gatgatccca gataacagca gtaggagat ggggaagtga gagaggaaag agatggaatc
25801 caatccatga tgtgttgcca agggagtatc ctctgtggac aactggagct gggaaactca
25861 tggaaacacg gagaatacag cactcatggg catcccatct gagtggtaag ggagccgggt
25921 atttatccac cagcttccac tggtcattgt ttgaagggtc ttgcaaggag ttgtttattc
25981 cctatgactt tgtcctgctg cacacaaggg cagagtagtc tcttgtgaac agacaaaggc
26041 ctgaggtcgg gagctgcaga tgctggaggt ggaagtcagg ctggcatgcc caggaagagg
26101 ggatataggt gagacacaga aagcatctgc tgcaagggtc coatgcctgg agttatcatg
26161 ggacataggg ctgctgtagg acaaatattg aaagatgagt cttggaaagt ttttaaggtc
26221 caagctgggt gcaaacaagt ttgttaagta ttggagctgg aggaacagtg ggagccctgc
26281 atgggatcca gagggttgtc taggaccaag aagacaaaaa ggacaagcca gggcaggtgt
26341 gcgggaggtg gtctggaaaa tggcaatgtc tgagcctggc ttttgtctat ctcagtgtgt
26401 gcaggaacgt gcaagcacag ggccaggaca gggctctatgc aggaccttct accactgacc
26461 caaggtctgt tgggtggggc gtgaggttag ggatgcttgg gctcacctct agagagaagt
26521 atcttttctt ggggtgccaac tagaggaggc tcacaggact ctttcttcca gctctcattc
26581 gcctagtctg caagccagga aatgtcctct ctctaaccat gctggaaggg tttttaaat
26641 tagaacttca taagcatgat aaaaactagc taacactgat ctagcaggtt ccagacact
26701 cttttgcaca ggaaggggca ggtgtcata accctgattt cttcacaggg tcttgataaa
26761 accatttaac caagaccact ggagtttaat ctgttcttcc ttaaaacttg ccttgcccaa
26821 tgctctacgt ctgatctaca tatagtgttc cgtctccat cttctgctgg tttgagcaat
26881 actgaacctc tgttcttcat ctgatttccc atgactttta gttcagggtt tttttgttg
26941 ttgggttgtt tgtttgtttg tttgtttgtt tttaccagga ctacacagat tttatcaact
27001 ttcacttatt gttgtctcac tctttttgat gaaaataaca tttctgaata taaaattaat
27061 atgttaaatt atagaaagt tcaaaaatat ggaaacacaa aaagaagaaa ctaggatga
27121 ctcataatth cactacctgg aggcaagtgc tataaattht ttgtgttttc atctaaagat

```

Figure 4 continued

```

27181 tttccatgtc ctatatagtt taaaaatcaa ttttatatatt tgccttttgtt cgtagaacac
27241 tatgtttgtga ggatttttctc aatgtttgata caaactcatc aaagcatgcc ttattccatg
27301 aaatggctgt gttaaataga ctttaatat cctatcctat tagacattta gcctggctgc
27361 tcacattctg atgctacaat aacattgagc tcagcctaata gttcataagt ctctaaccac
27421 atgtctaata attttcttag gattatttttt agaagtgaat aactaggcag gaccacagaa
27481 acgtttttaga actctgagca aatattgcaa aattgcattc cattgcataa gaaagtggat
27541 ttctttctttt tttttttttt tggggggggg cggatggagt ctgcgtctgt caccagggct
27601 gagagcagtg gcatggtctc ggctcactac cagctctgcc tcctgggttc acaccattct
27661 cctgcctcag cctcctgagt agctgcctgc caccacgctc ggtaattttt ttgtattttt
27721 agtacagacg gggttttctcc gtgttacctc ggatggtctc gatctctgga cctcgtgatc
27781 cgcccatctc agcctcccaa agtgctggga ttacaggcgt gaaccaccgc acccagccaa
27841 taatgtggat tctttttctc atttgggatg ctaagggttaa aaattaacat aacgctgcca
27901 gtttcataaa ctttccccac caataagaag cacatcagct agcataacac tgccagcttc
27961 ataagccttc ccaccaaca agaacacat cagcctgttt taatctgcat gtccctccatg
28021 actagccctc tcattgtctc tgcttctctt attatgtcta attgtcatta cccaaatctg
28081 tgcgtacttt tcctctgaag acttggtgtt tgattgtaag aattgtataa atatcaatgg
28141 ctttagccaa agagttttac aaagtacgtt ttgggcagca tgcctcctat aataagttct
28201 atcattcaca aaagctcttt cttctatggt tggtctgag tttctttgga aaatgttcga
28261 ggggcaaagg caaggtcaaa cctcctttt agcaagtttt gtttgacctt gaagcagcca
28321 tttaacaggt ggattacaga gccacagaag gatggcatct tcccagggtgc gcttcctgct
28381 gagccgaggg tgcatgagc gaggtgaca aagatgagtg aatagattat tcttcattct
28441 cacattgaaa tgacatgtgc ggggtgggca tgagggttagg gatgcgtgag ctcccatgtg
28501 taaatccttt tgtgaacgaa tttttaatga tgttaagaaa tgatggctgg aggcggtggc
28561 tcttgccaac actctcagca ttttgggagg ctgaggtggg aggattatga gagcctagga
28621 tttcaagatc agcctgggca acatagtgag accctgtctc taaaaaaca tttaaaaggg
28681 aagtgattac agtgactac aaaataccat atataatata atccccagtt aaatacataa
28741 aagagagata tgctaaaata aaaaagaagc attctctgag tggtagagtt atgagtata
28801 ctttaaattt gattgtttaa attgattttt aaattatata ggatatggaa aaattttaga
28861 aggaaatata aaaaatgcat ggcccttgcc tcaggtagtg gaattatgag tgatatctag
28921 ttttgcttta tgcgtttcca ttttttgtt tcttctaatt attcctaatt gtttttattt
28981 tccaaatttc ttactgtacc caaaggggag gacattgtaa attttctcca ccagcctggc
29041 attgccctga ctctttccgt cttagtgtca cttctcagtg ggtatcatga ccagtgcata
29101 gtgctgaggg acagggccag gcggcccacg aggtagccct gaaggggcag cgaatctgtg
29161 ctccaatgag ccaatcattg ggaaactggc atcttcctgg ggctgggaca ggagagcttt
29221 gttggtggga gaggagggca gagccgaagg agagagggca ggggagaggc aaggtcaccc
29281 ttgctcagcg ccttcttttc tctgccaggg ctttactgcc ttaacaccct aacactccca
29341 ggccaggcag gcagcaagac acatggggac agtgagctat aaagcagtga gtgccgcagg
29401 ttcttgagac acaagaggca gattgttccc tctctgcata tgtttgtatt taaatagcct
29461 catgttcaga caagcaactt ggatttgctt cagtcttctc aggatgggtg agaaaaacac
29521 tggaaatgct caaaacgagg cagaacagaa gacacctgtc atgggtggaga tggcatccaa
29581 gtgaatgaac gcgagtctgt tatcaatgaa ttatatccgg tgaaccaga tgtagagaga
29641 ttgtcttgtg tgagaatgca ccagcaggca aaggggttta cgaggcagag tcagggtccc
29701 tgcaccggca gctgttttaa atgaattgac ccagggtttc cacattcatt cattaacat
29761 ttgttcattg aacaaatatt tggtctaat tccgtttcca cctcagacag ctatgcagcc
29821 cggagcagat cactgagctt tctgagact ctgtttctct actgtgctga cagccagagg
29881 tttgctgcag agctggagac cttggtagcc tccagggtgtg tgtctggcaa catgaactgt
29941 gcatgaatcc ttgtcttctc agatgctgga gcaccacct acatgtatga gtttcagtac
30001 cgtccaatgc tctcatcaga catgaaaccc aagacgggtg taggagacca cggggatgag
30061 ctcttctccg tctttggggc cccattttta aaaggtaatg ctccttctct tctgtgagct
30121 aggagttaga gacttgggtt ccagtccttg tgtggtggcc ttcagcaatt ttgtctctc
30181 ctggtcttag tttcctgatt ctgtataaca ggggtgaaat tcaactgggt gtccgccggc
30241 ttctctccac tctaacacat gtgaggaggg atgtgtattc atgtgccctc tgcacagacg
30301 ctcagaatcc tgctgtgtga catcaagaca ggtggtgggt ggggtctcct gcaggattca
30361 ctatgacagg gttgaatgtc actgggttgt tccccggctt cctcccgtct taacacacct
30421 gagggaggat gtgtatcatg taccctctgc cacagacact cagaatcctg ggctgtgtca
30481 tcaacaagat aggcggtggg acacatgcgg ggcctgtctc taggttgaga agccaccaat
30541 caatttcctt ctcccaggag cgctgagaca gtgagtgagg gtgagattag aattggtgtt

```

Figure 4 continued

```

30601 aaccctaagt caggtctaata gtgtgacctgt ttgtctttct gtctgtctct aacaaaaact
30661 cataaacctt tacatttagt acattttaata attgttgatt tagaattcac aggatgtatg
30721 aatttggaac aggagaagag aatatTTTTT atatggTTTa tagtcagcaa ggtggccatc
30781 ccacaggctg ggaaggTgcc tctggccaag cccagaaacg ggctcattgg aggaggggtt
30841 gggtaggagc tttatgtgaa cggattggct aaacatacat gttcaacagg ttacaggggg
30901 agcaatggat attcatgaag acagtcctga cacatgtgta ttaacaaac acgtatgtaa
30961 catggcccat gttcacctgg tggTggagac ctaatatTTa aatgtattac aattagggcc
31021 tgtaagtcca aaggTcttct caggacacga agctcagcaa gtgaggagct tctgtacacc
31081 ggccaggTcc agtccatggc tggTgatgtt cttatctgga gaaagttact gaaatcagtc
31141 tcttatgcaa tcaaagcagt agTTaagtct ggcaagtCag ggttctgttt gtctgcgtat
31201 cccagctgca gccgttgTTa ttgttttctt ttgtcttagc tccaggccag tgctggTTTa
31261 gctgctagag aaagagaagc acctcgtggc agtgagaaca aagtggattc tttttttttt
31321 ttattattat actTTaagtt ttagggtaca tgtgcacatt gtgcaggTTa gttacatatg
31381 tatacatgtg ccgtgctggT gcgtgcacc cactaactcg tcatctagca ttaggtatat
31441 ctcccagTgc tatccctccc cctccccc accccaccac agtcccaga gtgtgatatt
31501 ccccttctctg tgtccatgtg atctcattgt tcaattccca cctatgagtg agaatatgcg
31561 gtgtttgggt tttgttctt gcgatagttt actgagaatg atggtttcca atttcatcca
31621 tgtccctaca aaggacatga actcatcatt ttttatggct gcatagtatt ccacggtgta
31681 tatgtgccac attttcttaa tccagtctat cattgttgga catttggtt ggttccaagt
31741 ctttgctatt gtgaataatg ccgcaataaa catacgtgtg catgtgtctt tatagcagca
31801 tgatttatag tcatttggtt atataccag taatgggatg gctgggtcaa atggtatttc
31861 tagttctaga tccctgagga atcgccacac tgacttccac aatggttgaa ctagtttaca
31921 gtcccaccaa cagtgtaaaa gtgttctat ttctccacat cctctccagc acctgttggt
31981 tcctgacttt ttagagatgc aggcctgaac ccttgcccga cgtggcctta ggtctcggtt
32041 ttaatttggt gtcttattgc cacaaagagt gtgttctgtc agaatgatga ccttcatttt
32101 attgctgatg ctggtccggt ggtgtctaaa tcacaaaagg gagggagtat tatgaggcgt
32161 gtctgacctc ctgtccgttc ccaggcagga acagagttga aggtttttcg aggtccct
32221 tagcccagag agggTctgtt cagtCagttg ggggtgtcag gattttattt ttagtttaca
32281 ttatacaata gtttaaaaaa tcttttcacg gtgtcttcta agtcttaact ctCagcagtt
32341 acagatcaga tgggactcac cccatggctg ctCagaacgc accagcgcct ctgggcatct
32401 cactgtgcac gcttaggcgc cttgcggctc tgttgttttt cagaatgtga attgtgggtg
32461 tgtgctgggg agggagaaaa gacccaagag agagcaacc tggaaagtgg ggttgcgtga
32521 gaggaacccc tgggtgggg catttctgtg gaccctgagc ggggaggagg caggcgaggg
32581 tcacaggaga caaaggcagc cagccaatgg aatgacgat gccattgct gagccctagg
32641 tcagatatgc tgaaggttaa agtacacaca aatccaccac gatcttctaa aagtgcagac
32701 ctCagctcag tgggtgtggg gcagtgtgcca agactatgtt tccagcagac tgcagtgtga
32761 tgttgacgct gcagtccacg gaccacactt tgagtaacaa aggtttttt tcttcttccc
32821 cacagagggt gcctcagaag aggagatcag acttagcaag atgggtgatga aattctgggc
32881 caactttgct cgcaatgggt gaggtcgtg gcaaagacag agcacggctg gtgagggtgg
32941 ggggcggggc atgcctattg ggaaggggca gcttctaagg ttctagtgat caaacttctg
33001 accctgtgac catagcactc tgataatgag agctctctgc aaacggagag gccgcccctg
33061 gagatagtga gctctccatc tctggaggta tacaagcctc ttgacagaga taacttgggc
33121 atcctcacac atctctgaag attgttggga acacacagca gctttggggc aattctaatt
33181 gattctgttt ccagaaacc caatggggaa gggctgccc actggccaga gtacaaccag
33241 aaggaagggt atctgcagat tggTgccaac acccaggcgg ccagaagct gaaggacaaa
33301 gaagtagctt tctggacca cctctttgce aagaaggcag tggagaagcc accccagaca
33361 gaacacatag agctgtgaat gaagatccag cggccttgga gagcctggag gagcaaagac
33421 tggggtcttt tgcgaaagg attgcagggt cagaaggcat cttaccatgg ctggggaatt
33481 gtctggtggt ggggggcagg ggacagagg catgaaggag caagttttgt atttgtgacc
33541 tcagctttgg gaataaagga tcttttgaag gcc

```

Figure 4 continued

LOCUS	CES1A1_promoter	1393 bp ds-DNA	linear	14-APR-		
2011						
1	gcttcatcca	cgttgtagca	tgttcttctt	atggctggat	gatatgcctc	tgcacgcata
61	caccacactt	catttacaca	ttcatcaggt	catggacatg	cgggttgtgc	ccactttttg
121	actaggatga	atgacactgc	tctttgtgta	caagcttttg	tgtggacatg	ctttcagttc
181	tcttcggtag	acacctagga	ggggaattgc	tgggtcatat	ggaaactctg	tgttgaactt
241	tttgaggaac	tgcctacacc	gcacaacagt	ttataaccta	ttattttctc	ctttcacaaac
301	gcacccctgtc	ttcattgtta	ataaatagct	gtgcacatca	ctattttgag	tgaatacatc
361	gtatttcctt	agccagcggg	accacagtgt	gtttagttaa	ttccctattg	ttggacattt
421	aataatccca	tgaaagaggg	tttgagtgc	ttccttgtgt	ccccagcagc	ttgtaaatga
481	cagttttcca	gggtgtgcca	gagcactctg	tatcccccta	atttggtgat	ttcacattgc
541	cttgacatca	cccctactgc	tccctccacc	catctgtctt	gtctacaatg	gtgagccctt
601	tgagaagctg	aaccatgtct	tgaattgggc	tttgtatcca	ttatacatgg	ccaggtgcca
661	agcatttttg	taaatgttca	caaaatggat	gggtacttgg	atggatggat	gataacagcc
721	agggactatg	gggggacgga	gttcattttt	acattctgtc	caaaaatccc	tgctttggaa
781	agtagtgagt	tctctgacac	cgatggtgtg	tgacaggagg	ctggaccagt	tggcaggata
841	ttgtgaagct	aattcaggcc	tcagaaaggg	gactcgatga	aatttaagat	cgttccaag
901	cttgagagcc	tggaaagcta	tgaaaacaca	agccctggga	gctgagatat	gtcctaactt
961	accagctga	gctgtgaggt	gtgagtggct	ctaacatttt	ccagttgttt	ctgaggacct
1021	cagatcaaag	cttccctttg	cctaaaagca	tctgcctgct	gggtgtgggc	cttggggggc
1081	gtcacagtgc	actgagggtta	gagtcctgca	aggggtgaacc	cttatgtaac	aagtagttgg
1141	gcaagtttac	agctctctgt	aatctgacag	tagagtccag	actgggttga	tgaaagaggg
1201	taaactgtgg	gtgggcgtgg	cctgaggccc	cactagaagc	ccaggagat	ctgaggaaag
1261	ggagggtttt	tctgatctct	cccaattaga	ggattaggca	attggcagcg	cagggcggta
1321	actctgggcg	gggctgggct	ccagggtctg	acagcacagt	ccctctgaac	tgcacagaga
1381	cctcgcaggg	ccc				

Figure 4 continued

LOCUS	CES1A2_1A3_promoter	1388 bp ds-DNA	linear	14-APR-		
2011						
1	ttcttatggc	tgatgatgat	gcctctgcac	gaatacacca	cactttatatt	acacattcat
61	caggtcatgg	acatgtgggt	tgtgcccact	ttttgactag	gatgaatgac	actgctcttt
121	gtgtacaagt	tttttatgga	catgctttca	gttctgttcg	gtagacacct	aggaggggaa
181	ttgctgggtc	atatggaaac	tctatgttga	atTTTTggag	gaactgccta	caccgcacaa
241	cagtttataa	cctgttattt	tttcctttca	caacgcattc	tgtcttcatt	gttaataagt
301	agctgtgcac	atcactattt	tgagtgaata	catcatattt	ccttagccag	cggatcacca
361	gtgtgtttag	tgaattccct	attgttggac	atttaataat	cccatgaaag	aggtttgagt
421	gacttccttg	tgtccccagc	agcttgtaaa	tgacagtttt	ccaggtgtgc	ccagagcact
481	ctgtatcccc	ttaatttggt	gatttcacat	tgccctgaca	tcacacctac	tgctcctcca
541	ccctatatgt	gttgtctaca	atgggtgaacc	ccttgagaag	ctgaacctatg	tcttgaattg
601	acctttgtat	ccattataca	tggccagggtg	cctggcattt	tggtaaatgt	tcacagaatg
661	gatgggtaca	tgaatggatg	gatgattaaa	gccaggagct	attgagagat	ggaatcattt
721	ttacattctg	tccaaaaatt	cctgctttgg	aaggtagtga	tttacctgag	attggaagtg
781	tgtgagaagg	cctggctagt	tggcaggata	ctgtgaagct	aattcaggca	tcagaaagtg
841	gactggataa	aatttaggat	cgcttccaag	cctgaaatcc	tggaaacgcta	tgaaacacaa
901	gccctgggag	ctgagatatg	tcctaactta	cccagctgag	ctgtgagggtg	tgagtggctc
961	taacattttc	cagttgtttc	tgaggacctc	agatcaaagc	ttccctttgc	ctaaaagcat
1021	ctgcctgctg	gtgcgggcct	ttggggggccc	tcagagagca	ctaagggttag	agtcttgcaa
1081	gggtgacacc	gttatgccac	aagcagttgg	gcgagtttac	agctctctgt	aatctgagag
1141	tagagtccag	attggtttga	tgaagagggtg	taaactgtga	gtgggcgtgg	cctgaggccc
1201	cactggaagc	ccaggagatg	ctgggggaaag	ggagggtctt	tctgatctct	cccaattaga
1261	ggattaggca	attggcagtg	cagggcggtg	actggggggc	agggtggcgc	cagggctgga
1321	cagcacagtc	cctctgagct	gcacggagac	ctcgcaggcc	cccggaaactg	tcgcccttcc
1381	aggatgtg					