



(86) Date de dépôt PCT/PCT Filing Date: 2003/12/22  
(87) Date publication PCT/PCT Publication Date: 2004/07/15  
(45) Date de délivrance/Issue Date: 2018/02/27  
(85) Entrée phase nationale/National Entry: 2005/06/17  
(86) N° demande PCT/PCT Application No.: US 2003/040978  
(87) N° publication PCT/PCT Publication No.: 2004/058052  
(30) Priorités/Priorities: 2002/12/20 (US60/434,778);  
2003/03/10 (US60/453,135); 2003/04/30 (US60/466,412);  
2003/09/23 (US60/504,955)

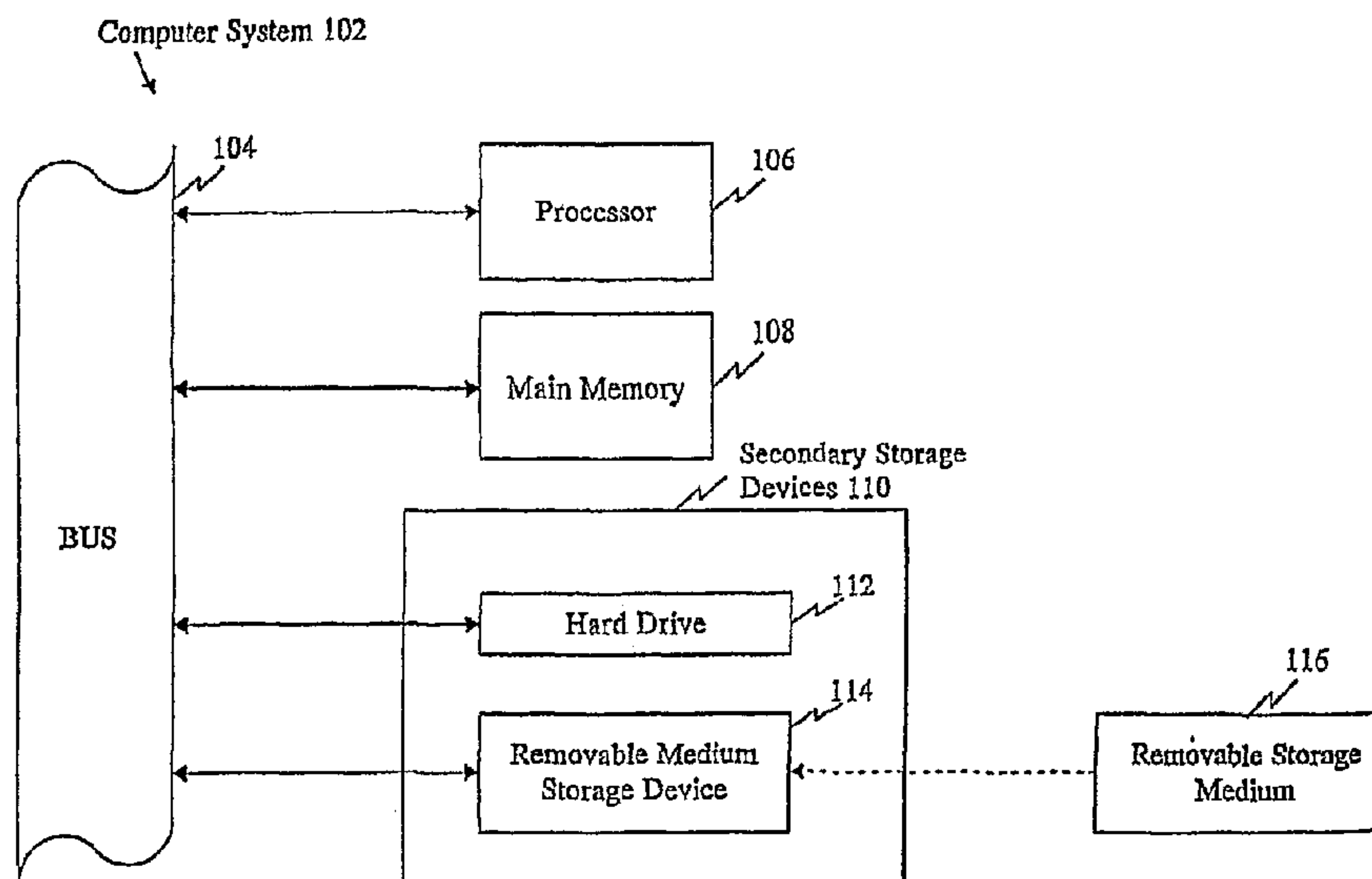
(51) Cl.Int./Int.Cl. *C12N 15/12* (2006.01),  
*C07H 21/00* (2006.01), *C07H 21/04* (2006.01),  
*C07K 14/47* (2006.01), *C07K 16/18* (2006.01),  
*C40B 30/04* (2006.01), *C40B 40/06* (2006.01),  
*G01N 33/48* (2006.01)

(72) Inventeurs/Inventors:  
CARGILL, MICHELE, US;  
DEVLIN, JAMES J., US;  
IAKOUBOVA, OLGA, US

(73) Propriétaire/Owner:  
CELERA CORPORATION, US

(74) Agent: FETHERSTONHAUGH & CO.

(54) Titre : POLYMORPHISMES GENETIQUES ASSOCIES A L'INFARCTUS DU MYOCARDE, TECHNIQUES DE  
DETECTION ET UTILISATIONS DE CEUX-CI  
(54) Title: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION  
AND USES THEREOF



(57) **Abrégé/Abstract:**

The present invention is based on the discovery of genetic polymorphisms that are associated with myocardial infarction. In particular, the present invention relates to nucleic acid molecules containing the polymorphisms, variant proteins encoded by such nucleic acid molecules, reagents for detecting the polymorphic nucleic acid molecules and proteins, and methods of using the nucleic acid and proteins as well as methods of using reagents for their detection.

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

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
15 July 2004 (15.07.2004)

PCT

(10) International Publication Number  
**WO 2004/058052 A2**

(51) International Patent Classification<sup>7</sup>: **A61B**

(21) International Application Number:  
PCT/US2003/040978

(22) International Filing Date:  
22 December 2003 (22.12.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/434,778 20 December 2002 (20.12.2002) US  
60/453,135 10 March 2003 (10.03.2003) US  
60/466,412 30 April 2003 (30.04.2003) US  
60/504,955 23 September 2003 (23.09.2003) US

(71) Applicant (for all designated States except US): **APPLERA CORPORATION** [US/US]; LEE, Victor, K., Assistant Secretary, c/o Celera Genomics, 45 West Gude Drive C-2 4#21, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CARGILL, Michele** [US/US]; c/o Celera Genomics, 45 West Gude Drive

C2-4#21, Rockville, MD 20850 (US). **DEVLIN, James, J.** [US/US]; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **IAKOUBOVA, Olga** [US/US]; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

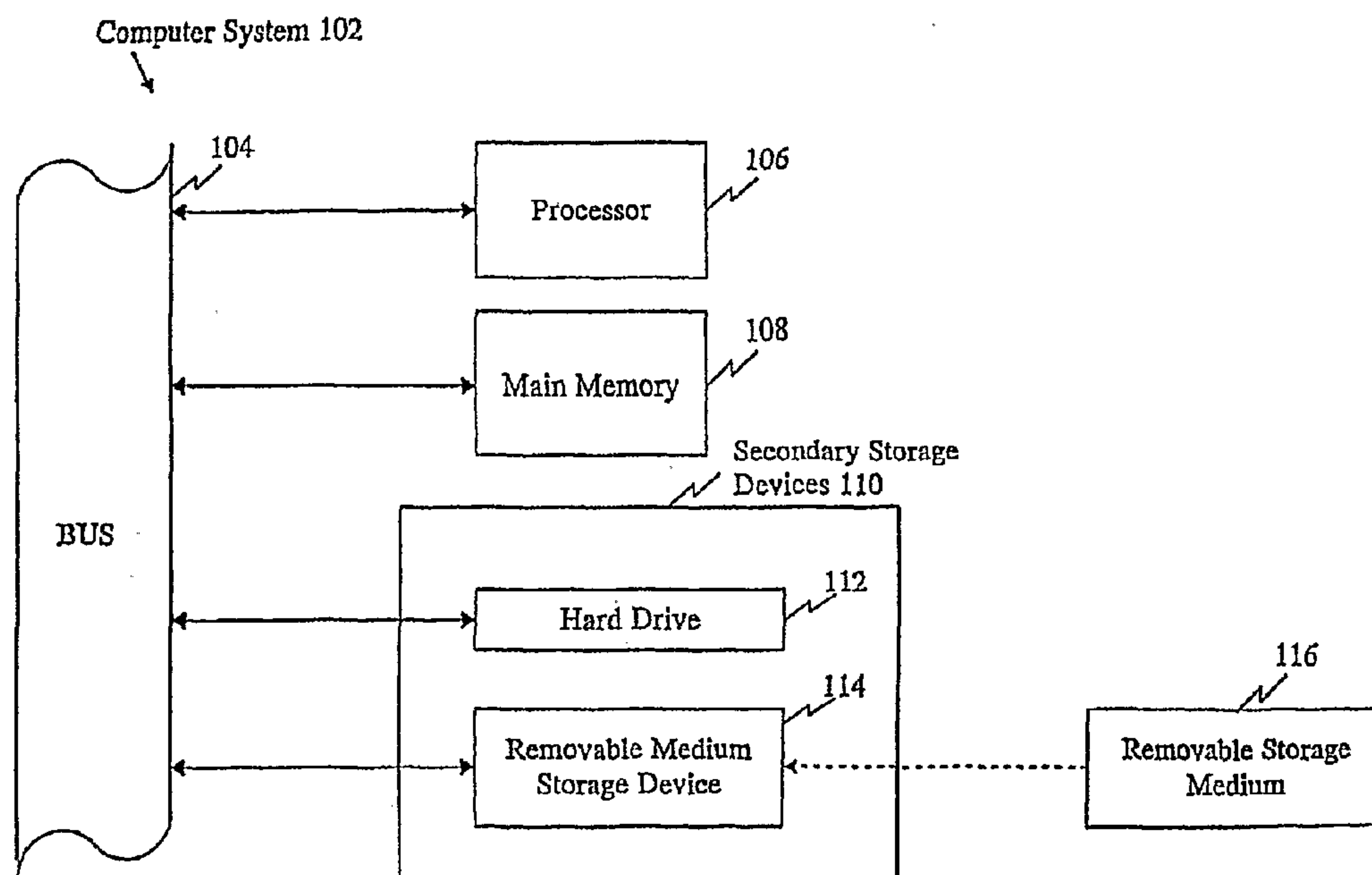
(74) Common Representative: **APPLERA CORPORATION**; LEE, Victor, K., Assistant Secretary, c/o Celera Genomics, 45 West Gude Drive C-2 4#21, Rockville, MD 20850 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,

[Continued on next page]

(54) Title: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF



(57) Abstract: The present invention is based on the discovery of genetic polymorphisms that are associated with myocardial infarction. In particular, the present invention relates to nucleic acid molecules containing the polymorphisms, variant proteins encoded by such nucleic acid molecules, reagents for detecting the polymorphic nucleic acid molecules and proteins, and methods of using the nucleic acid and proteins as well as methods of using reagents for their detection.

WO 2004/058052 A2

# WO 2004/058052 A2



ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# **DEMANDES OU BREVETS VOLUMINEUX**

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS  
COMPREND PLUS D'UN TOME.**

**CECI EST LE TOME \_\_1\_\_ DE \_\_3\_\_**

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

---

# **JUMBO APPLICATIONS / PATENTS**

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE  
THAN ONE VOLUME.**

**THIS IS VOLUME \_\_1\_\_ OF \_\_3\_\_**

NOTE: For additional volumes please contact the Canadian Patent Office.

## **GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF**

### **FIELD OF THE INVENTION**

5       The present invention is in the field of myocardial infarction diagnosis and therapy. In particular, the present invention relates to specific single nucleotide polymorphisms (SNPs) in the human genome, and their association with myocardial infarction (including recurrent myocardial infarction) and related pathologies. Based on differences in allele frequencies in the myocardial infarction patient  
10       population relative to normal individuals, the naturally-occurring SNPs disclosed herein can be used as targets for the design of diagnostic reagents and the development of therapeutic agents, as well as for disease association and linkage analysis. In particular, the SNPs of the present invention are useful for identifying an individual who is at an increased or decreased risk of developing myocardial  
15       infarction and for early detection of the disease, for providing clinically important information for the prevention and/or treatment of myocardial infarction, and for screening and selecting therapeutic agents. The SNPs disclosed herein are also useful for human identification applications. Methods, assays, kits, and reagents for detecting the presence of these polymorphisms and their encoded products are  
20       provided.

### **BACKGROUND OF THE INVENTION**

#### Myocardial Infarction (including Recurrent Myocardial Infarction)

25       Myocardial infarction (MI) is the most common cause of mortality in developed countries. It is a multifactorial disease that involves atherogenesis, thrombus formation and propagation. Thrombosis can result in complete or partial occlusion of coronary arteries. The luminal narrowing or blockage of coronary arteries reduces oxygen and nutrient supply to the cardiac muscle (cardiac ischemia), leading to myocardial necrosis and/or stunning. MI, unstable angina, or sudden  
30       ischemic death are clinical manifestations of cardiac muscle damage. All three endpoints are part of the Acute Coronary Syndrome since the underlying mechanisms of acute complications of atherosclerosis are considered to be the same.

Atherogenesis, the first step of pathogenesis of MI, is a complex interaction between blood elements, mechanical forces, disturbed blood flow, and vessel wall

abnormality. On the cellular level, these include endothelial dysfunction, monocytes/macrophages activation by modified lipoproteins, monocytes/macrophages migration into the neointima and subsequent migration and proliferation of vascular smooth muscle cells (VSMC) from the media that results in plaque accumulation.

5 In recent years, an unstable (vulnerable) plaque was recognized as an underlying cause of arterial thrombotic events and MI. A vulnerable plaque is a plaque, often not stenotic, that has a high likelihood of becoming disrupted or eroded, thus forming a thrombogenic focus. Two vulnerable plaque morphologies have been described. A first type of vulnerable plaque morphology is a rupture of the protective  
10 fibrous cap. It can occur in plaques that have distinct morphological features such as large and soft lipid pool with distinct necrotic core and thinning of the fibrous cap in the region of the plaque shoulders. Fibrous caps have considerable metabolic activity. The imbalance between matrix synthesis and matrix degradation thought to be regulated by inflammatory mediators combined with VSMC apoptosis are the key  
15 underlying mechanisms of plaque rupture. A second type of vulnerable plaque morphology, known as "plaque erosion", can also lead to a fatal coronary thrombotic event. Plaque erosion is morphologically different from plaque rupture. Eroded plaques do not have fractures in the plaque fibrous cap, only superficial erosion of the intima. The loss of endothelial cells can expose the thrombogenic subendothelial  
20 matrix that precipitates thrombus formation. This process could be regulated by inflammatory mediators. The propagation of the acute thrombi for both plaque rupture and plaque erosion events depends on the balance between coagulation and thrombolysis. MI due to a vulnerable plaque is a complex phenomenon that includes: plaque vulnerability, blood vulnerability (hypercoagulation, hypothrombolysis), and  
25 heart vulnerability (sensitivity of the heart to ischemia or propensity for arrhythmia).

Recurrent myocardial infarction can generally be viewed as a severe form of MI progression caused by multiple vulnerable plaques that are able to undergo pre-rupture or a pre-erosive state, coupled with extreme blood coagulability.

30 The incidence of MI is still high despite currently available preventive measures and therapeutic intervention. More than 1,500,000 people in the US suffer acute MI each year (many without seeking help due to unrecognized MI), and one third of these people die. The lifetime risk of coronary artery disease events at age 40 years is 42.4% for men (one in two) and 24.9% for women (one in four) (Lloyd-Jones DM; *Lancet*, 1999 353: 89-92).

the current diagnosis of MI is based on the levels of troponin I or T that indicate the cardiac muscle progressive necrosis, impaired electrocardiogram (ECG), and detection of abnormal ventricular wall motion or angiographic data (the presence of acute thrombi). However, due to the asymptomatic nature of 25% of acute MIs (absence of atypical chest pain, low ECG sensitivity), a significant portion of MIs are not diagnosed and therefore not treated appropriately (e.g., prevention of recurrent MIs).

Despite a very high prevalence and lifetime risk of MI, there are no good prognostic markers that can identify an individual with a high risk of vulnerable plaques and justify preventive treatments. MI risk assessment and prognosis is currently done using classic risk factors or the recently introduced Framingham Risk Index. Both of these assessments put a significant weight on LDL levels to justify preventive treatment. However, it is well established that half of all MIs occur in individuals without overt hyperlipidemia. Hence, there is a need for additional risk factors for predicting predisposition to MI.

Other emerging risk factors are inflammatory biomarkers such as C-reactive protein (CRP), ICAM-1, SAA, TNF  $\alpha$ , homocysteine, impaired fasting glucose, new lipid markers (ox LDL, Lp-a, MAD-LDL, etc.) and pro-thrombotic factors (fibrinogen, PAI-1). Despite showing some promise, these markers have significant limitations such as low specificity and low positive predictive value, and the need for multiple reference intervals to be used for different groups of people (e.g., males-females, smokers-non smokers, hormone replacement therapy users, different age groups). These limitations diminish the utility of such markers as independent prognostic markers for MI screening.

Genetics plays an important role in MI risk. Families with a positive family history of MI account for 14% of the general population, 72% of premature MIs, and 48% of all MIs (Williams R R, *Am J Cardiology*, 2001; 87:129). In addition, replicated linkage studies have revealed evidence of multiple regions of the genome that are associated with MI and relevant to MI genetic traits, including regions on chromosomes 14, 2, 3 and 7 (Broeckel U, *Nature Genetics*, 2002; 30: 210; Harrap S, *Arterioscler Thromb Vasc Biol*, 2002; 22: 874-878, Shearman A, *Human Molecular Genetics*, 2000, 9; 9,1315-1320), implying that genetic risk factors influence the onset, manifestation, and progression of MI. Recent association studies have

identified allelic variants that are associated with acute complications of coronary heart disease, including allelic variants of the ApoE, ApoA5, Lpa, APOCIII, and Klotho genes.

Genetic markers such as single nucleotide polymorphisms are preferable to other types of biomarkers. Genetic markers that are prognostic for MI can be genotyped early in life and could predict individual response to various risk factors. The combination of serum protein levels and genetic predisposition revealed by genetic analysis of susceptibility genes can provide an integrated assessment of the interaction between genotypes and environmental factors, resulting in synergistically increased prognostic value of diagnostic tests.

Thus, there is an urgent need for novel genetic markers that are predictive of predisposition to MI, particularly for individuals who are unrecognized as having a predisposition to MI. Such genetic markers may enable prognosis of MI in much larger populations compared with the populations which can currently be evaluated by using existing risk factors and biomarkers. The availability of a genetic test may allow, for example, appropriate preventive treatments for acute coronary events to be provided for susceptible individuals (such preventive treatments may include, for example, statin treatments and statin dose escalation, as well as changes to modifiable risk factors), lowering of the thresholds for ECG and angiography testing, and allow adequate monitoring of informative biomarkers.

Moreover, the discovery of genetic markers associated with MI will provide novel targets for therapeutic intervention or preventive treatments of MI, and enable the development of new therapeutic agents for treating MI and other cardiovascular disorders.

25

#### SNPs

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution, generating variant forms of progenitor genetic sequences (Gusella, *Ann. Rev. Biochem.* 55, 831-854 (1986)). A variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. Additionally, the effects of a

variant form may be both beneficial and detrimental, depending on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. In many cases, both progenitor and variant forms survive and co-exist in a species population. The coexistence of multiple forms of a genetic sequence gives rise to genetic polymorphisms, including SNPs.

Approximately 90% of all polymorphisms in the human genome are SNPs. SNPs are single base positions in DNA at which different alleles, or alternative nucleotides, exist in a population. The SNP position (interchangeably referred to herein as SNP, SNP site, or SNP locus) is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations). An individual may be homozygous or heterozygous for an allele at each SNP position. A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP is an amino acid coding sequence.

A SNP may arise from a substitution of one nucleotide for another at the polymorphic site. Substitutions can be transitions or transversions. A transition is the replacement of one purine nucleotide by another purine nucleotide, or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine, or vice versa. A SNP may also be a single base insertion or deletion variant referred to as an "indel" (Weber *et al.*, "Human diallelic insertion/deletion polymorphisms", *Am J Hum Genet* 2002 Oct;71(4):854-62).

A synonymous codon change, or silent mutation/SNP (terms such as "SNP", "polymorphism", "mutation", "mutant", "variation", and "variant" are used herein interchangeably), is one that does not result in a change of amino acid due to the degeneracy of the genetic code. A substitution that changes a codon coding for one amino acid to a codon coding for a different amino acid (i.e., a non-synonymous codon change) is referred to as a missense mutation. A nonsense mutation results in a type of non-synonymous codon change in which a stop codon is formed, thereby leading to premature termination of a polypeptide chain and a truncated protein. A read-through mutation is another type of non-synonymous codon change that causes the destruction of a stop codon, thereby resulting in an extended polypeptide product.

while SNPs can be bi-, tri-, or tetra- allelic, the vast majority of the SNPs are bi-allelic, and are thus often referred to as “bi-allelic markers”, or “di-allelic markers”.

As used herein, references to SNPs and SNP genotypes include individual SNPs and/or haplotypes, which are groups of SNPs that are generally inherited  
5 together. Haplotypes can have stronger correlations with diseases or other phenotypic effects compared with individual SNPs, and therefore may provide increased diagnostic accuracy in some cases (Stephens *et al. Science* 293, 489-493, 20 July 2001).

Causative SNPs are those SNPs that produce alterations in gene expression or  
10 in the expression, structure, and/or function of a gene product, and therefore are most predictive of a possible clinical phenotype. One such class includes SNPs falling within regions of genes encoding a polypeptide product, i.e. cSNPs. These SNPs may result in an alteration of the amino acid sequence of the polypeptide product (i.e., non-synonymous codon changes) and give rise to the expression of a defective or other  
15 variant protein. Furthermore, in the case of nonsense mutations, a SNP may lead to premature termination of a polypeptide product. Such variant products can result in a pathological condition, e.g., genetic disease. Examples of genes in which a SNP within a coding sequence causes a genetic disease include sickle cell anemia and cystic fibrosis.

Causative SNPs do not necessarily have to occur in coding regions; causative  
20 SNPs can occur in, for example, any genetic region that can ultimately affect the expression, structure, and/or activity of the protein encoded by a nucleic acid. Such genetic regions include, for example, those involved in transcription, such as SNPs in transcription factor binding domains, SNPs in promoter regions, in areas involved in  
25 transcript processing, such as SNPs at intron-exon boundaries that may cause defective splicing, or SNPs in mRNA processing signal sequences such as polyadenylation signal regions. Some SNPs that are not causative SNPs nevertheless are in close association with, and therefore segregate with, a disease-causing sequence. In this situation, the presence of a SNP correlates with the presence of, or  
30 predisposition to, or an increased risk in developing the disease. These SNPs, although not causative, are nonetheless also useful for diagnostics, disease predisposition screening, and other uses.

An association study of a SNP and a specific disorder involves determining the presence or frequency of the SNP allele in biological samples from individuals

with the disorder of interest, such as myocardial infarction, and comparing the information to that of controls (i.e., individuals who do not have the disorder; controls may be also referred to as "healthy" or "normal" individuals) who are preferably of similar age and race. The appropriate selection of patients and controls is important  
5 to the success of SNP association studies. Therefore, a pool of individuals with well-characterized phenotypes is extremely desirable.

A SNP may be screened in diseased tissue samples or any biological sample obtained from a diseased individual, and compared to control samples, and selected for its increased (or decreased) occurrence in a specific pathological condition, such  
10 as pathologies related to myocardial infarction. Once a statistically significant association is established between one or more SNP(s) and a pathological condition (or other phenotype) of interest, then the region around the SNP can optionally be thoroughly screened to identify the causative genetic locus/sequence(s) (e.g., causative SNP/mutation, gene, regulatory region, etc.) that influences the pathological  
15 condition or phenotype. Association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families (linkage studies).

Clinical trials have shown that patient response to treatment with pharmaceuticals is often heterogeneous. There is a continuing need to improve  
20 pharmaceutical agent design and therapy. In that regard, SNPs can be used to identify patients most suited to therapy with particular pharmaceutical agents (this is often termed "pharmacogenomics"). Similarly, SNPs can be used to exclude patients from certain treatment due to the patient's increased likelihood of developing toxic side effects or their likelihood of not responding to the treatment. Pharmacogenomics can  
25 also be used in pharmaceutical research to assist the drug development and selection process. (Linder et al. (1997), *Clinical Chemistry*, 43, 254; Marshall (1997), *Nature Biotechnology*, 15, 1249; International Patent Application WO 97/40462, Spectra Biomedical; and Schafer et al. (1998), *Nature Biotechnology*, 16, 3).

## 30 SUMMARY OF THE INVENTION

The present invention relates to the identification of novel SNPs, unique combinations of such SNPs, and haplotypes of SNPs that are associated with myocardial infarction (including recurrent myocardial infarction) and related pathologies. The polymorphisms disclosed herein are directly useful as targets for the

design of diagnostic reagents and the development of therapeutic agents for use in the diagnosis and treatment of myocardial infarction and related pathologies.

Based on the identification of SNPs associated with myocardial infarction, the present invention also provides methods of detecting these variants as well as the design and preparation of detection reagents needed to accomplish this task. The invention specifically provides novel SNPs in genetic sequences involved in myocardial infarction, variant proteins encoded by nucleic acid molecules containing such SNPs, antibodies to the encoded variant proteins, computer-based and data storage systems containing the novel SNP information, methods of detecting these SNPs in a test sample, methods of identifying individuals who have an altered (i.e., increased or decreased) risk of developing myocardial infarction based on the presence of a SNP disclosed herein or its encoded product, methods of identifying individuals who are more or less likely to respond to a treatment, methods of screening for compounds useful in the treatment of a disorder associated with a variant gene/protein, compounds identified by these methods, methods of treating disorders mediated by a variant gene/protein, and methods of using the novel SNPs of the present invention for human identification.

In Tables 1-2, the present invention provides gene information, transcript sequences (SEQ ID NOS:1-26), encoded amino acid sequences (SEQ ID NOS:27-46), genomic sequences (SEQ ID NOS:67-73), transcript-based context sequences (SEQ ID NOS:47-66) and genomic-based context sequences (SEQ ID NOS:74-80) that contain the SNPs of the present invention, and extensive SNP information that includes observed alleles, allele frequencies, populations/ethnic groups in which alleles have been observed, information about the type of SNP and corresponding functional effect, and, for cSNPs, information about the encoded polypeptide product. The transcript sequences (SEQ ID NOS:1-26), amino acid sequences (SEQ ID NOS:27-46), genomic sequences (SEQ ID NOS:67-73), transcript-based SNP context sequences (SEQ ID NOS: 47-66), and genomic-based SNP context sequences (SEQ ID NOS:74-80) are also provided in the Sequence Listing.

In a specific embodiment of the present invention, naturally-occurring SNPs in the human genome are provided. These SNPs are associated with myocardial infarction such that they can have a variety of uses in the diagnosis and/or treatment of myocardial infarction. One aspect of the present invention relates to an isolated

nucleic acid molecule comprising a nucleotide sequence in which at least one nucleotide is a SNP disclosed in Tables 3 and/or 4. In an alternative embodiment, a nucleic acid of the invention is an amplified polynucleotide, which is produced by amplification of a SNP-containing nucleic acid template. In another embodiment, the invention provides for a variant protein which is encoded by a nucleic acid molecule containing a SNP disclosed herein.

In yet another embodiment of the invention, a reagent for detecting a SNP in the context of its naturally-occurring flanking nucleotide sequences (which can be, e.g., either DNA or mRNA) is provided. In particular, such a reagent may be in the form of, for example, a hybridization probe or an amplification primer that is useful in the specific detection of a SNP of interest. In an alternative embodiment, a protein detection reagent is used to detect a variant protein which is encoded by a nucleic acid molecule containing a SNP disclosed herein. A preferred embodiment of a protein detection reagent is an antibody or an antigen-reactive antibody fragment.

Also provided in the invention are kits comprising SNP detection reagents, and methods for detecting the SNPs disclosed herein by employing detection reagents. In a specific embodiment, the present invention provides for a method of identifying an individual having an increased or decreased risk of developing myocardial infarction by detecting the presence or absence of a SNP allele disclosed herein. In another embodiment, a method for diagnosis of myocardial infarction by detecting the presence or absence of a SNP allele disclosed herein is provided.

The nucleic acid molecules of the invention can be inserted in an expression vector, such as to produce a variant protein in a host cell. Thus, the present invention also provides for a vector comprising a SNP-containing nucleic acid molecule, genetically-engineered host cells containing the vector, and methods for expressing a recombinant variant protein using such host cells. In another specific embodiment, the host cells, SNP-containing nucleic acid molecules, and/or variant proteins can be used as targets in a method for screening and identifying therapeutic agents or pharmaceutical compounds useful in the treatment of myocardial infarction.

An aspect of this invention is a method for treating myocardial infarction in a human subject wherein said human subject harbors a gene, transcript, and/or encoded protein identified in Tables 1-2, which method comprises administering to said human subject a therapeutically or prophylactically effective amount of one or more agents

counteracting the effects of the disease, such as by inhibiting (or stimulating) the activity of the gene, transcript, and/or encoded protein identified in Tables 1-2.

Another aspect of this invention is a method for identifying an agent useful in therapeutically or prophylactically treating myocardial infarction in a human subject wherein said human subject harbors a gene, transcript, and/or encoded protein identified in Tables 1-2, which method comprises contacting the gene, transcript, or encoded protein with a candidate agent under conditions suitable to allow formation of a binding complex between the gene, transcript, or encoded protein and the candidate agent and detecting the formation of the binding complex, wherein the presence of the complex identifies said agent.

Another aspect of this invention is a method for treating myocardial infarction in a human subject, which method comprises:

(i) determining that said human subject harbors a gene, transcript, and/or encoded protein identified in Tables 1-2, and

(ii) administering to said subject a therapeutically or prophylactically effective amount of one or more agents counteracting the effects of the disease.

Various embodiments of the claimed invention relate to a method for determining a human's risk for myocardial infarction (MI), the method comprising testing nucleic acid from said human for the presence or absence of a polymorphism in gene lipoprotein(a) *LPA* at position 101 of SEQ ID NO:78 or its complement, wherein the presence of G at position 101 of SEQ ID NO:78 or C at position 101 of its complement indicates that said human has an increased risk for MI.

Various embodiments of the claimed invention relate to an allele-specific polynucleotide for use in a method as described herein, wherein said polynucleotide specifically hybridizes to said polymorphism in which said G or said C is present.

Various embodiments of the claimed invention relate to an allele-specific polynucleotide for use in a method as described herein, wherein said polynucleotide comprises a segment of SEQ ID NO: 78 or its complement at least 16 nucleotides in length that includes said position 101.

Various embodiments of the claimed invention relate to an allele-specific polynucleotide for use in a method as described herein, wherein the allele-specific polynucleotide is a primer that comprises SEQ ID NO:90 or SEQ ID NO:91.

Various embodiments of the claimed invention relate to a kit for use in a method as described herein, wherein said kit comprises at least one allele-specific polynucleotide as described herein and at least one further component, wherein the at least one further component is a buffer, deoxynucleotide triphosphates (dNTPs), and amplification primer pair, an enzyme or any  
 5 combination thereof.

Many other uses and advantages of the present invention will be apparent to those skilled in the art upon review of the detailed description of the preferred embodiments herein. Solely for clarity of discussion, the invention is described in the sections below by way of non-limiting  
 10 examples.

### **DESCRIPTION OF TABLE 1 AND TABLE 2**

Table 1 and Table 2 disclose the SNP and associated gene/transcript/protein information of the present invention. For each gene, Table 1 and Table 2 each provide a header containing  
 15 gene/transcript/protein information, followed by a transcript and protein sequence (in Table 1) or genomic sequence (in Table 2), and then SNP information regarding each SNP found in that gene/transcript.

NOTE: SNPs may be included in both Table 1 and Table 2; Table 1 presents the SNPs  
 20 relative to their transcript sequences and encoded protein sequences, whereas Table 2 presents the SNPs relative to their genomic sequences (in some instances Table 2 may also include, after the last gene sequence, genomic sequences of one or more intergenic regions, as well as SNP context sequences and other SNP information for any SNPs that lie within these intergenic regions). SNPs can readily be cross-referenced between Tables based on their hCV (or, in some  
 25 instances, hDV) identification numbers.

The gene/transcript/protein information includes:

30

- a gene number (1 through n, where n = the total number of genes in the Table)
- a Celera hCG and UID internal identification numbers for the gene
- a Celera hCT and UID internal identification numbers for the transcript (Table 1 only)
- a public Genbank accession number (e.g., RefSeq NM number) for the transcript (Table 1 only)
- a Celera hCP and UID internal identification numbers for the protein encoded by the hCT transcript (Table 1 only)
- a public Genbank accession number (e.g., RefSeq NP number) for the protein (Table 1 only)

- an art-known gene symbol
- an art-known gene/protein name
- Celera genomic axis position (indicating start nucleotide position-stop nucleotide position)
- the chromosome number of the chromosome on which the gene is located
- an OMIM (Online Mendelian Inheritance in Man; Johns Hopkins University/NCBI)

public reference number for obtaining further information regarding the medical significance of each gene

- alternative gene/protein name(s) and/or symbol(s) in the OMIM entry

NOTE: Due to the presence of alternative splice forms, multiple transcript/protein entries can be provided for a single gene entry in Table 1; i.e., for a single Gene Number, multiple entries may be provided in series that differ in their transcript/protein information and sequences.

Following the gene/transcript/protein information is a transcript sequence and protein sequence (in Table 1), or a genomic sequence (in Table 2), for each gene, as follows:

- transcript sequence (Table 1 only) (corresponding to SEQ ID NOS:1-26 of the Sequence Listing), with SNPs identified by their IUB codes (transcript sequences can include 5' UTR, protein coding, and 3' UTR regions). (NOTE: If there are differences between the nucleotide sequence of the hCT transcript and the corresponding public transcript sequence identified by the Genbank accession number, the hCT transcript sequence (and encoded protein) is provided, unless the public sequence is a RefSeq transcript sequence identified by an NM number, in which case the RefSeq NM transcript sequence (and encoded protein) is provided. However, whether the hCT transcript or

RefSeq NM transcript is used as the transcript sequence, the disclosed SNPs are represented by their IUB codes within the transcript.)

- the encoded protein sequence (Table 1 only) (corresponding to SEQ ID NOS:27-46 of the Sequence Listing)

5           - the genomic sequence of the gene (Table 2 only), including 6kb on each side of the gene boundaries (i.e., 6kb on the 5' side of the gene plus 6kb on the 3' side of the gene) (corresponding to SEQ ID NOS:67-73 of the Sequence Listing).

After the last gene sequence, Table 2 may include additional genomic sequences of intergenic regions (in such instances, these sequences are identified as "Intergenic region:"

10 followed by a numerical identification number), as well as SNP context sequences and other SNP information for any SNPs that lie within each intergenic region (and such SNPs are identified as "INTERGENIC" for SNP type).

NOTE: The transcript, protein, and transcript-based SNP context sequences are provided in both Table 1 and in the Sequence Listing. The genomic and genomic-based SNP context sequences  
15 are provided in both Table 2 and in the Sequence Listing. SEQ ID NOS are indicated in Table 1 for each transcript sequence (SEQ ID NOS:1-26), protein sequence (SEQ ID NOS:27-46), and transcript-based SNP context sequence (SEQ ID NOS:47-66), and SEQ ID NOS are indicated in Table 2 for each genomic sequence (SEQ ID NOS:67-73), and genomic-based SNP context sequence (SEQ ID NOS:74-80).

20

The SNP information includes:

- context sequence (taken from the transcript sequence in Table 1, and taken from the genomic sequence in Table 2) with the SNP represented by its IUB code, including 100 bp upstream (5') of the SNP position plus 100 bp downstream (3') of the SNP position (the transcript-  
25 based SNP context sequences in Table 1 are provided in the Sequence Listing as SEQ ID NOS: 47-66; the genomic-based SNP context sequences in Table 2 are provided in the Sequence Listing as SEQ ID NOS:74-80).

- Celera hCV internal identification number for the SNP (in some instances, an "hDV" number is given instead of an "hCV" number)

30           - SNP position [position of the SNP within the given transcript sequence (Table 1) or within the given genomic sequence (Table 2)]

- SNP source (may include any combination of one or more of the following five codes, depending on which internal sequencing projects and/or public databases the SNP has been observed in: “Applera” = SNP observed during the re-sequencing of genes and regulatory regions of 39 individuals, “Celera” = SNP observed during shotgun sequencing and assembly of the Celera human genome sequence, “Celera Diagnostics” = SNP observed during re-sequencing of nucleic acid samples from individuals who have myocardial infarction or a related pathology, “dbSNP” = SNP observed in the dbSNP public database, “HGBASE” = SNP observed in the HGBASE public database, “HGMD” = SNP observed in the Human Gene Mutation Database (HGMD) public database) (NOTE: multiple “Applera” source entries for a single SNP indicate that the same SNP was covered by multiple overlapping amplification products and the re-sequencing results (e.g., observed allele counts) from each of these amplification products is being provided)

- Population/allele/allele count information in the format of [population1(allele1,count|allele2,count) population2(allele1,count|allele2,count) total (allele1,total count|allele2,total count)]. The information in this field includes populations/ethnic groups in which particular SNP alleles have been observed (“cau” = Caucasian, “his” = Hispanic, “chn” = Chinese, and “afr” = African-American, “jpn” = Japanese, “ind” = Indian, “mex” = Mexican, “ain” = “American Indian, “cra” = Celera donor, “no\_pop” = no population information available), identified SNP alleles, and observed allele counts (within each population group and total allele counts), where available [“-“ in the allele field represents a deletion allele of an insertion/deletion (“indel”) polymorphism (in which case the corresponding insertion allele, which may be comprised of one or more nucleotides, is indicated in the allele field on the opposite side of the “|”); “-“ in the count field indicates that allele count information is not available].

NOTE: For SNPs of “Applera” SNP source, genes/regulatory regions of 39 individuals (20 Caucasians and 19 African Americans) were re-sequenced and, since

each SNP position is represented by two chromosomes in each individual (with the exception of SNPs on X and Y chromosomes in males, for which each SNP position is represented by a single chromosome), up to 78 chromosomes were genotyped for each SNP position. Thus, the sum of the African-American ("afr") allele counts is up to 38, the sum of the Caucasian allele counts ("cau") is up to 40, and the total sum of all allele counts is up to 78.

(NOTE: semicolons separate population/allele/count information corresponding to each indicated SNP source; i.e., if four SNP sources are indicated, such as "Celera", "dbSNP", "HGBASE", and "HGMD", then population/allele/count information is provided in four groups which are separated by semicolons and listed in the same order as the listing of SNP sources, with each population/allele/count information group corresponding to the respective SNP source based on order; thus, in this example, the first population/allele/count information group would correspond to the first listed SNP source (Celera) and the third population/allele/count information group separated by semicolons would correspond to the third listed SNP source (HGBASE); if population/allele/count information is not available for any particular SNP source, then a pair of semicolons is still inserted as a place-holder in order to maintain correspondence between the list of SNP sources and the corresponding listing of population/allele/count information)

- SNP type (e.g., location within gene/transcript and/or predicted functional effect) ["MIS-SENSE MUTATION" = SNP causes a change in the encoded amino acid (i.e., a non-synonymous coding SNP); "SILENT MUTATION" = SNP does not cause a change in the encoded amino acid (i.e., a synonymous coding SNP); "STOP CODON MUTATION" = SNP is located in a stop codon; "NONSENSE MUTATION" = SNP creates or destroys a stop codon; "UTR 5" = SNP is located in a 5' UTR of a transcript; "UTR 3" = SNP is located in a 3' UTR of a transcript; "PUTATIVE UTR 5" = SNP is located in a putative 5' UTR; "PUTATIVE UTR 3" = SNP is located in a putative 3' UTR; "DONOR SPLICE SITE" = SNP is located in a donor splice site (5' intron boundary); "ACCEPTOR SPLICE SITE" = SNP is located in an acceptor splice site (3' intron boundary); "CODING REGION" = SNP is located in a protein-coding region of the transcript; "EXON" = SNP is located in an exon; "INTRON" = SNP is located in an intron; "hmCS" = SNP is located in a human-mouse conserved segment; "TFBS" = SNP is located in a transcription factor binding

site; “UNKNOWN” = SNP type is not defined; “INTERGENIC” = SNP is intergenic, i.e., outside of any gene boundary]

- Protein coding information (Table 1 only), where relevant, in the format of [protein SEQ ID NO:#, amino acid position, (amino acid-1, codon1) (amino acid-2, codon2)]. The information in this field includes SEQ ID NO of the encoded protein sequence, position of the amino acid residue within the protein identified by the SEQ ID NO that is encoded by the codon containing the SNP, amino acids (represented by one-letter amino acid codes) that are encoded by the alternative SNP alleles (in the case of stop codons, “X” is used for the one-letter amino acid code), and alternative codons containing the alternative SNP nucleotides which encode the amino acid residues (thus, for example, for missense mutation-type SNPs, at least two different amino acids and at least two different codons are generally indicated; for silent mutation-type SNPs, one amino acid and at least two different codons are generally indicated, etc.). In instances where the SNP is located outside of a protein-coding region (e.g., in a UTR region), “None” is indicated following the protein SEQ ID NO.

#### DESCRIPTION OF TABLE 3 AND TABLE 4

Tables 3 and 4 provide a list of a subset of SNPs from Table 1 (in the case of Table 3) or Table 2 (in the case of Table 4) for which the SNP source falls into one of the following three categories: 1) SNPs for which the SNP source is only “Applera” and none other, 2) SNPs for which the SNP source is only “Celera Diagnostics” and none other, and 3) SNPs for which the SNP source is both “Applera” and “Celera Diagnostics” but none other.

These SNPs have not been observed in any of the public databases (dbSNP, HGBASE, and HGMD), and were also not observed during shotgun sequencing and assembly of the Celera human genome sequence (i.e., “Celera” SNP source). Tables 3 and 4 provide the hCV identification number (or hDV identification number for SNPs having “Celera Diagnostics” SNP source) and the SEQ ID NO of the context sequence for each of these SNPs.

#### DESCRIPTION OF TABLE 5

Table 5 provides sequences (SEQ ID NOS:81-101) of primers that have been synthesized and used in the laboratory to carry out allele-specific PCR reactions in order to assay the SNPs disclosed in Tables 6-7 during the course of myocardial infarction association studies.

- 5 Table 5 provides the following:
- the column labeled “hCV” provides an hCV identification number for each SNP site
  - the column labeled “Alleles” designates the two alternative alleles at the SNP site identified by the hCV identification number that are targeted by the allele-specific primers (the allele-specific primers are shown as “Sequence A” and “Sequence B” in each row)
  - 10 - the column labeled “Sequence A (allele-specific primer)” provides an allele-specific primer that is specific for the first allele designated in the “Alleles” column
  - the column labeled “Sequence B (allele-specific primer)” provides an allele-specific primer that is specific for the second allele designated in the “Alleles” column
  - the column labeled “Sequence C (common primer)” provides a common primer that is
  - 15 used in conjunction with each of the allele-specific primers (the “Sequence A” primer and the “Sequence B” primer) and which hybridizes at a site away from the SNP position.

All primer sequences are given in the 5’ to 3’ direction.

- Each of the alleles designated in the “Alleles” column matches the 3’ nucleotide of the allele-specific primer that is specific for that allele. Thus, the first allele designated in the
- 20 “Alleles” column matches the 3’ nucleotide of the “Sequence A” primer, and the second allele designated in the “Alleles” column matches the 3’ nucleotide of the “Sequence B” primer.

#### **DESCRIPTION OF TABLE 6 AND TABLE 7**

- Tables 6 and 7 provide results of statistical analyses for SNPs disclosed in
- 25 Tables 1-5 (SNPs can be cross-referenced between Tables based on their hCV identification numbers). Table 6 provides statistical results for association of SNPs with myocardial infarction, and Table 7 provides statistical results for association of SNPs with recurrent myocardial infarction (RMI). The statistical results shown in Tables 6-7 provide support for the association of these SNPs with MI (Table 6) and/or

30

RMI (Table 7). For example, the statistical results provided in Tables 6-7 show that the association of these SNPs with MI and/or RMI is supported by p-values  $< 0.05$  in at least one of three genotypic association tests and/or an allelic association test. Moreover, in general, the SNPs identified in Tables 6-7 are SNPs for which their association with MI and/or RMI has been replicated by virtue of being significant in at least two independently collected sample sets, which further verifies the association of these SNPs with MI and/or RMI. Furthermore, results of stratification-based analyses are also provided; stratified analysis can, for example, enable increased prediction of MI and/or RMI risk via interaction between conventional risk factors (stratum) and SNPs.

NOTE: SNPs can be cross-referenced between Tables 1-7 based on the hCV identification number of each SNP. However, 1 of the SNPs that is included in Tables 1-7 possesses two different hCV identification numbers, as follows:

- hCV7499900 is equivalent to hCV25620145

TABLE 6 (SNP association with Myocardial Infarction)

Description of column headings for Table 6:

| Table 6 column heading | Definition |
|------------------------|------------|
|------------------------|------------|

|                     |                                                                                                         |
|---------------------|---------------------------------------------------------------------------------------------------------|
| Marker              | Internal hCV identification number for the SNP that is tested                                           |
| Study               | Sample set used in the analysis                                                                         |
| Stratification      | Indicates if the analysis of the dataset was done on a substratum (stratifications are described below) |
| Strata              | Indicates what substratum was used in the analysis (strata are described below)                         |
| Status              | Identifies the inclusion/exclusion criteria for cases and controls (described below)                    |
| Allele1             | Nucleotide (allele) of the tested SNP for which statistics are being reported                           |
| Case Allele1 frq    | Allele frequency of Allele1 in cases                                                                    |
| Control Allele1 frq | Allele frequency of Allele1 in controls                                                                 |
| Allelic p-value     | Result of the Fisher exact test for allelic association                                                 |
| Dom p-value         | Result of the asymptotic chi square test for dominant genotypic association                             |
| Rec p-value         | Result of the asymptotic chi square test for recessive genotypic association                            |
| OR                  | Allelic odds ratio                                                                                      |
| OR 95%CI L          | Lower limit of 95% confidence interval of the allelic odds ratio                                        |
| OR 95%CI U          | Upper limit of 95% confidence interval of the allelic odds ratio                                        |

Definition of entries in the "Stratification" and "Strata" columns (Table 6) for stratification-based analyses:

| Stratification | Strata | Definition      |
|----------------|--------|-----------------|
| no             | ALL    | All individuals |

|          |      |                                                        |
|----------|------|--------------------------------------------------------|
| BMI_GE27 | L /H | Individuals with body mass index lower/higher than 27  |
| HTN      | Y/N  | Individuals with/without (Y/N) history of hypertension |
| SEX      | M/F  | Gender male/female                                     |
| AGE_LT60 | Y/O  | Individuals younger/older than 60 years                |
| SMOKE    | Y/N  | Individuals that have/do not have history of smoking   |

Definition of entries in the "Status" column (Table 6):

| Status            | Definition                                                                                                          |
|-------------------|---------------------------------------------------------------------------------------------------------------------|
| LT60MI_60TO75noMI | MI cases younger than 60 compared to controls between the ages of 60 to 75                                          |
| LT60MI_GT75noMI   | MI cases younger than 60 compared to controls older than 75                                                         |
| MI_GT75noMI       | MI cases compared to controls older than 75                                                                         |
| MI_LT75noMI       | MI cases compared to controls younger than 75                                                                       |
| MI_noASD          | Cases with history of MI, controls with no history of atherosclerotic disease                                       |
| MI_noMI           | Cases with history of MI, controls with no history of MI                                                            |
| MI_YOUNGOLD_noASD | Cases with history of MI under the age of 60 controls with no history of atherosclerotic disease over the age of 60 |
| MI_YOUNGOLD_noMI  | Cases with history of MI under the age of 60 controls with no history of MI over the age of 60                      |
| YoungMI_GT75noASD | Cases with history of MI under the age of 60 controls with no history of atherosclerotic disease over the age of 75 |

TABLE 7 (SNP association with Recurrent Myocardial Infarction)

5 Description of column headings for Table 7:

| Table 7 column heading | Definition                                                                                                                                                                                                   |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Gene                   | Locus Link HUGO approved gene symbol                                                                                                                                                                         |
| Marker                 | Internal hCV identification number for the tested SNP                                                                                                                                                        |
| Sample Set             | Sample Set used in the analysis (CARE, Pre-CARE or WGS_S0012)                                                                                                                                                |
| p-value                | Result of the asymptotic chi square test for allelic association, dominant genotypic association, recessive genotypic association, or the allelic, dominant, or recessive p-value of the stratified analysis |
| OR                     | odds ratio                                                                                                                                                                                                   |
| 95%CI                  | 95% confidence interval of the given odds ratio                                                                                                                                                              |
| Case_Freq              | Allele frequency of minor allele in cases                                                                                                                                                                    |
| Control_Freq           | Allele frequency of minor allele in controls                                                                                                                                                                 |
| Allele1                | Nucleotide (allele) of the tested SNP for which statistics are being reported                                                                                                                                |
| Mode                   | The mode of inheritance                                                                                                                                                                                      |
| Strata                 | Indicates if the analysis of the dataset was based on a substratum such as gender, age, BMI, Hypertension, Fasting Glucose levels, etc. (strata are described below)                                         |

Definition of entries in the "Strata" column (Table 7) for stratification-based analyses:

| Stratum           | Definition                                                 |
|-------------------|------------------------------------------------------------|
| BMI_TERTILE_1     | Individuals in the lowest tertile of body mass index       |
| BMI_TERTILE_2     | Individuals in the middle tertile of body mass index       |
| BMI_TERTILE_3     | Individuals in the highest tertile of body mass index      |
| PLACEBO           | Patients who were in placebo arm of the CARE trail         |
| PRAVASTATIN       | Patients who were in Pravastatin arm of the CARE trail     |
| MALE              | Only males                                                 |
| FEMALE            | Only females                                               |
| HYPERTEN_1        | Individuals with history of Hypertension                   |
| HYPERTEN_0        | Individuals without history of Hypertension                |
| GLUCOSE_TERTILE_1 | Individuals in a lowest tertile of Fasting Glucose levels  |
| GLUCOSE_TERTILE_2 | Individuals in a middle tertile of Fasting Glucose levels  |
| GLUCOSE_TERTILE_3 | Individuals in a highest tertile of Fasting Glucose levels |

|               |                                                       |
|---------------|-------------------------------------------------------|
| AGE_TERTILE_1 | Individuals in a lowest tertile of age (premature MI) |
| AGE_TERTILE_2 | Individuals in a middle tertile of age                |
| AGE_TERTILE_3 | Individuals in a highest tertile age                  |
| EVERSMOKED_0  | Individuals wh0 never smoked                          |
| EVERSMOKED_1  | Former smokers                                        |
| EVERSMOKED_2  | Current smokers                                       |
| FMHX_CHD_1    | Individuals with family history of CHD                |
| FMHX_CHD_0    | Individuals without family history of CHD             |

### DESCRIPTION OF THE FIGURE

Figure 1 provides a diagrammatic representation of a computer-based  
5 discovery system containing the SNP information of the present invention in  
computer readable form.

### DETAILED DESCRIPTION OF THE INVENTION

The present invention provides SNPs associated with myocardial infarction  
10 (including recurrent myocardial infarction), nucleic acid molecules containing  
SNPs, methods and reagents for the detection of the SNPs disclosed herein, uses of  
these SNPs for the development of detection reagents, and assays or kits that utilize  
such reagents. The myocardial infarction-associated SNPs disclosed herein are useful  
for diagnosing, screening for, and evaluating predisposition to myocardial infarction  
15 and related pathologies in humans. Furthermore, such SNPs and their encoded  
products are useful targets for the development of therapeutic agents.

A large number of SNPs have been identified from re-sequencing DNA from  
39 individuals, and they are indicated as "Applera" SNP source in Tables 1-2. Their  
allele frequencies observed in each of the Caucasian and African-American ethnic  
20 groups are provided. Additional SNPs included herein were previously identified  
during shotgun sequencing and assembly of the human genome, and they are  
indicated as "Celera" SNP source in Tables 1-2. Furthermore, the information  
provided in Table 1-2, particularly the allele frequency information obtained from 39  
individuals and the identification of the precise position of each SNP within each

gene/transcript, allows haplotypes (i.e., groups of SNPs that are co-inherited) to be readily inferred. The present invention encompasses SNP haplotypes, as well as individual SNPs.

Thus, the present invention provides individual SNPs associated with myocardial infarction, as well as combinations of SNPs and haplotypes in genetic regions associated with myocardial  
 5 infarction, polymorphic/variant transcript sequences (SEQ ID NOS:1-26) and genomic sequences (SEQ ID NOS:67-73) containing SNPs, encoded amino acid sequences (SEQ ID NOS: 27-46), and both transcript-based SNP context sequences (SEQ ID NOS: 47-66) and genomic-based SNP context sequences (SEQ ID NOS:74-80) (transcript sequences, protein sequences, and transcript-based SNP context sequences are provided in Table 1 and the Sequence Listing; genomic sequences  
 10 and genomic-based SNP context sequences are provided in Table 2 and the Sequence Listing), methods of detecting these polymorphisms in a test sample, methods of determining the risk of an individual of having or developing myocardial infarction, methods of screening for compounds useful for treating disorders associated with a variant gene/protein such as myocardial infarction, compounds identified by these screening methods, methods of using the disclosed SNPs to select a  
 15 treatment strategy, methods of treating a disorder associated with a variant gene/protein (i.e., therapeutic methods), and methods of using the SNPs of the present invention for human identification.

The present invention provides novel SNPs associated with myocardial infarction, as well as SNPs that were previously known in the art, but were not previously known to be associated with  
 20 myocardial infarction. Accordingly, the present invention provides novel compositions and methods based on the novel SNPs disclosed herein, and also provides novel methods of using the known, but previously unassociated, SNPs in methods relating to myocardial infarction (e.g., for diagnosing myocardial infarction, etc.). In Tables 1-2, known SNPs are identified based on the public database in which they have been observed, which is indicated as one or more of the  
 25 following SNP types: “dbSNP” = SNP observed in dbSNP, “HGBASE” = SNP observed in HGBASE, and “HGMD” = SNP observed in the Human Gene Mutation Database (HGMD). Novel SNPs for which the SNP source is only “Applera” and none other, i.e., those that have not been observed in any public databases and which were also not observed during shotgun sequencing and assembly of the

Celera human genome sequence (i.e., "Celera" SNP source), are indicated in Tables 3-4.

Particular SNP alleles of the present invention can be associated with either an increased risk of having or developing myocardial infarction, or a decreased risk of having or developing myocardial infarction. SNP alleles that are associated with a decreased risk of having or developing myocardial infarction may be referred to as "protective" alleles, and SNP alleles that are associated with an increased risk of having or developing myocardial infarction may be referred to as "susceptibility" alleles or "risk factors". Thus, whereas certain SNPs (or their encoded products) can be assayed to determine whether an individual possesses a SNP allele that is indicative of an increased risk of having or developing myocardial infarction (i.e., a susceptibility allele), other SNPs (or their encoded products) can be assayed to determine whether an individual possesses a SNP allele that is indicative of a decreased risk of having or developing myocardial infarction (i.e., a protective allele). Similarly, particular SNP alleles of the present invention can be associated with either an increased or decreased likelihood of responding to a particular treatment or therapeutic compound, or an increased or decreased likelihood of experiencing toxic effects from a particular treatment or therapeutic compound. The term "altered" may be used herein to encompass either of these two possibilities (e.g., an increased or a decreased risk/likelihood).

Those skilled in the art will readily recognize that nucleic acid molecules may be double-stranded molecules and that reference to a particular site on one strand refers, as well, to the corresponding site on a complementary strand. In defining a SNP position, SNP allele, or nucleotide sequence, reference to an adenine, a thymine (uridine), a cytosine, or a guanine at a particular site on one strand of a nucleic acid molecule also defines the thymine (uridine), adenine, guanine, or cytosine (respectively) at the corresponding site on a complementary strand of the nucleic acid molecule. Thus, reference may be made to either strand in order to refer to a particular SNP position, SNP allele, or nucleotide sequence. Probes and primers, may be designed to hybridize to either strand and SNP genotyping methods disclosed herein

may generally target either strand. Throughout the specification, in identifying a SNP position, reference is generally made to the protein-encoding strand, only for the purpose of convenience.

References to variant peptides, polypeptides, or proteins of the present invention include peptides, polypeptides, proteins, or fragments thereof, that contain at least one amino acid residue that differs from the corresponding amino acid sequence of the art-known peptide/polypeptide/protein (the art-known protein may be interchangeably referred to as the “wild-type”, “reference”, or “normal” protein). Such variant peptides/polypeptides/proteins can result from a codon change caused by a nonsynonymous nucleotide substitution at a protein-coding SNP position (i.e., a missense mutation) disclosed by the present invention. Variant peptides/polypeptides/proteins of the present invention can also result from a nonsense mutation, i.e. a SNP that creates a premature stop codon, a SNP that generates a read-through mutation by abolishing a stop codon, or due to any SNP disclosed by the present invention that otherwise alters the structure, function/activity, or expression of a protein, such as a SNP in a regulatory region (e.g. a promoter or enhancer) or a SNP that leads to alternative or defective splicing, such as a SNP in an intron or a SNP at an exon/intron boundary. As used herein, the terms “polypeptide”, “peptide”, and “protein” are used interchangeably.

## **ISOLATED NUCLEIC ACID MOLECULES**

### **AND SNP DETECTION REAGENTS & KITS**

Tables 1 and 2 provide a variety of information about each SNP of the present invention that is associated with myocardial infarction, including the transcript sequences (SEQ ID NOS:1-26), genomic sequences (SEQ ID NOS: 67-73), and protein sequences (SEQ ID NOS:27-46) of the encoded gene products (with the SNPs indicated by IUB codes in the nucleic acid sequences). In addition, Tables 1 and 2 include SNP context sequences, which generally include 100 nucleotide upstream (5') plus 100 nucleotides downstream (3') of each SNP position (SEQ ID NOS:47-66 correspond to transcript-based SNP context sequences disclosed in Table 1, and SEQ ID NOS:74-80 correspond to genomic-based context sequences disclosed in Table 2), the

alternative nucleotides (alleles) at each SNP position, and additional information about the variant where relevant, such as SNP type (coding, missense, splice site, UTR, etc.), human populations in which the SNP was observed, observed allele frequencies, information about the encoded  
5 protein, etc.

### **Isolated Nucleic Acid Molecules**

The present invention provides isolated nucleic acid molecules that contain one or more SNPs disclosed Table 1 and/or Table 2. Preferred isolated nucleic acid  
10 molecules contain one or more SNPs identified in Table 3 and/or Table 4. Isolated nucleic acid molecules containing one or more SNPs disclosed in at least one of Tables 1-4 may be interchangeably referred to throughout the present text as "SNP-containing nucleic acid molecules". Isolated nucleic acid molecules may optionally encode a full-length variant protein or fragment thereof. The isolated nucleic acid  
15 molecules of the present invention also include probes and primers (which are described in greater detail below in the section entitled "SNP Detection Reagents"), which may be used for assaying the disclosed SNPs, and isolated full-length genes, transcripts, cDNA molecules, and fragments thereof, which may be used for such purposes as expressing an encoded protein.

20 As used herein, an "isolated nucleic acid molecule" generally is one that contains a SNP of the present invention or one that hybridizes to such molecule such as a nucleic acid with a complementary sequence, and is separated from most other nucleic acids present in the natural source of the nucleic acid molecule. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule containing a SNP of the present  
25 invention, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. A nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered "isolated". Nucleic acid molecules present in non-human transgenic animals, which do not naturally occur in the animal, are also  
30 considered "isolated". For example, recombinant DNA molecules contained in a vector are considered "isolated". Further examples of "isolated" DNA molecules include recombinant DNA molecules maintained in heterologous host cells, and purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include

in vivo or in vitro RNA transcripts of the isolated SNP-containing DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Generally, an isolated SNP-containing nucleic acid molecule comprises one or more SNP positions disclosed by the present invention with flanking nucleotide sequences on either side of the SNP positions. A flanking sequence can include nucleotide residues that are naturally associated with the SNP site and/or heterologous nucleotide sequences. Preferably the flanking sequence is up to about 500, 300, 100, 60, 50, 30, 25, 20, 15, 10, 8, or 4 nucleotides (or any other length in-between) on either side of a SNP position, or as long as the full-length gene or entire protein-coding sequence (or any portion thereof such as an exon), especially if the SNP-containing nucleic acid molecule is to be used to produce a protein or protein fragment.

For full-length genes and entire protein-coding sequences, a SNP flanking sequence can be, for example, up to about 5KB, 4KB, 3KB, 2KB, 1KB on either side of the SNP. Furthermore, in such instances, the isolated nucleic acid molecule comprises exonic sequences (including protein-coding and/or non-coding exonic sequences), but may also include intronic sequences. Thus, any protein coding sequence may be either contiguous or separated by introns. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences and is of appropriate length such that it can be subjected to the specific manipulations or uses described herein such as recombinant protein expression, preparation of probes and primers for assaying the SNP position, and other uses specific to the SNP-containing nucleic acid sequences.

An isolated SNP-containing nucleic acid molecule can comprise, for example, a full-length gene or transcript, such as a gene isolated from genomic DNA (e.g., by cloning or PCR amplification), a cDNA molecule, or an mRNA transcript molecule. Polymorphic transcript sequences are provided in Table 1 and in the Sequence Listing (SEQ ID NOS: 1-26), and polymorphic genomic sequences are provided in Table 2 and in the Sequence Listing (SEQ ID NOS:67-73). Furthermore, fragments of such full-length genes and transcripts that contain one or more SNPs disclosed herein are also encompassed by the present invention, and such fragments may be used, for example, to express any part of a protein, such as a particular functional domain or an antigenic epitope.

Thus, the present invention also encompasses fragments of the nucleic acid sequences provided in Tables 1-2 (transcript sequences are provided in Table 1

as SEQ ID NOS:1-26, genomic sequences are provided in Table 2 as SEQ ID NOS:67-73, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:47-66, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:74-80) and their complements. A fragment typically comprises a contiguous nucleotide sequence at least about 8 or more nucleotides, more preferably at least about 12 or more nucleotides, and even more preferably at least about 16 or more nucleotides. Further, a fragment could comprise at least about 18, 20, 22, 25, 30, 40, 50, 60, 100, 250 or 500 (or any other number in-between) nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope-bearing regions of a variant peptide or regions of a variant peptide that differ from the normal/wild-type protein, or can be useful as a polynucleotide probe or primer. Such fragments can be isolated using the nucleotide sequences provided in Table 1 and/or Table 2 for the synthesis of a polynucleotide probe. A labeled probe can then be used, for example, to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in amplification reactions, such as for purposes of assaying one or more SNPs sites or for cloning specific regions of a gene.

An isolated nucleic acid molecule of the present invention further encompasses a SNP-containing polynucleotide that is the product of any one of a variety of nucleic acid amplification methods, which are used to increase the copy numbers of a polynucleotide of interest in a nucleic acid sample. Such amplification methods are well known in the art, and they include but are not limited to, polymerase chain reaction (PCR) (U.S. Patent Nos. 4,683,195; and 4,683,202; *PCR Technology: Principles and Applications for DNA Amplification*, ed. H.A. Erlich, Freeman Press, NY, NY, 1992), ligase chain reaction (LCR) (Wu and Wallace, *Genomics* 4:560, 1989; Landegren *et al.*, *Science* 241:1077, 1988), strand displacement amplification (SDA) (U.S. Patent Nos. 5,270,184; and 5,422,252), transcription-mediated amplification (TMA) (U.S. Patent No. 5,399,491), linked linear amplification (LLA) (U.S. Patent No. 6,027,923), and the like, and isothermal amplification methods such as nucleic acid sequence based amplification (NASBA), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Natl. Acad. Sci. USA* 87: 1874, 1990). Based on such methodologies, a person skilled in the art can readily design primers in any suitable

regions 5' and 3' to a SNP disclosed herein. Such primers may be used to amplify DNA of any length so long that it contains the SNP of interest in its sequence.

As used herein, an "amplified polynucleotide" of the invention is a SNP-containing nucleic acid molecule whose amount has been increased at least two fold  
5 by any nucleic acid amplification method performed *in vitro* as compared to its starting amount in a test sample. In other preferred embodiments, an amplified polynucleotide is the result of at least ten fold, fifty fold, one hundred fold, one thousand fold, or even ten thousand fold increase as compared to its starting amount in a test sample. In a typical PCR amplification, a polynucleotide of interest is often  
10 amplified at least fifty thousand fold in amount over the unamplified genomic DNA, but the precise amount of amplification needed for an assay depends on the sensitivity of the subsequent detection method used.

Generally, an amplified polynucleotide is at least about 16 nucleotides in length. More typically, an amplified polynucleotide is at least about 20 nucleotides in  
15 length. In a preferred embodiment of the invention, an amplified polynucleotide is at least about 30 nucleotides in length. In a more preferred embodiment of the invention, an amplified polynucleotide is at least about 32, 40, 45, 50, or 60 nucleotides in length. In yet another preferred embodiment of the invention, an amplified polynucleotide is at least about 100, 200, or 300 nucleotides in length. While the total  
20 length of an amplified polynucleotide of the invention can be as long as an exon, an intron or the entire gene where the SNP of interest resides, an amplified product is typically no greater than about 1,000 nucleotides in length (although certain amplification methods may generate amplified products greater than 1000 nucleotides in length). More preferably, an amplified polynucleotide is not greater than about 600  
25 nucleotides in length. It is understood that irrespective of the length of an amplified polynucleotide, a SNP of interest may be located anywhere along its sequence.

In a specific embodiment of the invention, the amplified product is at least about 201 nucleotides in length, comprises one of the transcript-based context sequences or the genomic-based context sequences shown in Tables 1-2. Such a  
30 product may have additional sequences on its 5' end or 3' end or both. In another embodiment, the amplified product is about 101 nucleotides in length, and it contains a SNP disclosed herein. Preferably, the SNP is located at the middle of the amplified product (e.g., at position 101 in an amplified product that is 201 nucleotides in length, or at position 51 in an amplified product that is 101 nucleotides in length), or within 1,

2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, or 20 nucleotides from the middle of the amplified product (however, as indicated above, the SNP of interest may be located anywhere along the length of the amplified product).

The present invention provides isolated nucleic acid molecules that comprise, consist of, or consist essentially of one or more polynucleotide sequences that contain one or more SNPs disclosed herein, complements thereof, and SNP-containing fragments thereof.

Accordingly, the present invention provides nucleic acid molecules that consist of any of the nucleotide sequences shown in Table 1 and/or Table 2 (transcript sequences are provided in Table 1 as SEQ ID NOS:1-26, genomic sequences are provided in Table 2 as SEQ ID NOS:67-73, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:47-66, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:74-80), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:27-46). A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of any of the nucleotide sequences shown in Table 1 and/or Table 2 (transcript sequences are provided in Table 1 as SEQ ID NOS:1-26, genomic sequences are provided in Table 2 as SEQ ID NOS:67-73, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:47-66, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:74-80), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:27-46). A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleotide residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise any of the nucleotide sequences shown in Table 1 and/or Table 2 or a SNP-containing fragment thereof (transcript sequences are provided in Table 1 as SEQ ID NOS:1-26, genomic sequences are provided in Table 2 as SEQ ID NOS:67-73, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:47-66, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:74-80), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:27-46). A nucleic acid molecule comprises a nucleotide

sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleotide residues, such as residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have one to a few additional nucleotides or can comprise many more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made and isolated is provided below, and such techniques are well known to those of ordinary skill in the art (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY).

The isolated nucleic acid molecules can encode mature proteins plus additional amino or carboxyl-terminal amino acids or both, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life, or facilitate manipulation of a protein for assay or production. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

Thus, the isolated nucleic acid molecules include, but are not limited to, nucleic acid molecules having a sequence encoding a peptide alone, a sequence encoding a mature peptide and additional coding sequences such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), a sequence encoding a mature peptide with or without additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but untranslated sequences that play a role in, for example, transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding, and/or stability of mRNA. In addition, the nucleic acid molecules may be fused to heterologous marker sequences encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA, which may be obtained, for example, by molecular cloning or produced by chemical synthetic techniques or by a combination thereof (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY). Furthermore, isolated nucleic acid molecules, particularly SNP detection reagents such as probes and primers, can also be partially or completely in the form of one or more types of nucleic acid analogs, such as

peptide nucleic acid (PNA) (U.S. Patent Nos. 5,539,082; 5,527,675; 5,623,049; 5,714,331). The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the complementary non-coding strand (anti-sense strand). DNA, RNA, or PNA segments  
5 can be assembled, for example, from fragments of the human genome (in the case of DNA or RNA) or single nucleotides, short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic nucleic acid molecule. Nucleic acid molecules can be readily synthesized using the sequences provided herein as a reference; oligonucleotide and PNA oligomer synthesis techniques are well known in the art  
10 (see, e.g., Corey, "Peptide nucleic acids: expanding the scope of nucleic acid recognition", *Trends Biotechnol.* 1997 Jun;15(6):224-9, and Hyrup et al., "Peptide nucleic acids (PNA): synthesis, properties and potential applications", *Bioorg Med Chem.* 1996 Jan;4(1):5-23). Furthermore, large-scale automated oligonucleotide/PNA synthesis (including synthesis on an array or bead surface or other solid support) can  
15 readily be accomplished using commercially available nucleic acid synthesizers, such as the Applied Biosystems (Foster City, CA) 3900 High-Throughput DNA Synthesizer or Expedite 8909 Nucleic Acid Synthesis System, and the sequence information provided herein.

The present invention encompasses nucleic acid analogs that  
20 contain modified, synthetic, or non-naturally occurring nucleotides or structural elements or other alternative/modified nucleic acid chemistries known in the art. Such nucleic acid analogs are useful, for example, as detection reagents (e.g., primers/probes) for detecting one or more SNPs identified in Table 1 and/or Table 2. Furthermore, kits/systems (such as  
25 beads, arrays, etc.) that include these analogs are also encompassed by the present invention. For example, PNA oligomers that are based on the polymorphic sequences of the present invention are specifically contemplated. PNA oligomers are analogs of DNA in which the phosphate backbone is replaced with a peptide-like backbone (Lagriffoul *et al.*,  
30 *Bioorganic & Medicinal Chemistry Letters*, 4: 1081-1082 (1994), Petersen *et al.*, *Bioorganic & Medicinal Chemistry Letters*, 6: 793-796 (1996), Kumar *et al.*, *Organic Letters* 3(9): 1269-1272 (2001), WO96/04000). PNA hybridizes to complementary RNA or DNA with higher affinity and

specificity than conventional oligonucleotides and oligonucleotide analogs. The properties of PNA enable novel molecular biology and biochemistry applications unachievable with traditional oligonucleotides and peptides.

Additional examples of nucleic acid modifications that improve the  
5 binding properties and/or stability of a nucleic acid include the use of base analogs such as inosine, intercalators (U.S. Patent No. 4,835,263) and the minor groove binders (U.S. Patent No. 5,801,115). Thus, references herein to nucleic acid molecules, SNP-containing nucleic acid molecules, SNP  
detection reagents (e.g., probes and primers),  
10 oligonucleotides/polynucleotides include PNA oligomers and other nucleic acid analogs. Other examples of nucleic acid analogs and alternative/modified nucleic acid chemistries known in the art are described in *Current Protocols in Nucleic Acid Chemistry*, John Wiley & Sons, N.Y. (2002).

15 The present invention further provides nucleic acid molecules that encode fragments of the variant polypeptides disclosed herein as well as nucleic acid molecules that encode obvious variants of such variant polypeptides. Such nucleic acid molecules may be naturally occurring, such as paralogs (different locus) and orthologs (different organism), or may be constructed by recombinant DNA methods  
20 or by chemical synthesis. Non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, the variants can contain nucleotide substitutions, deletions, inversions and insertions (in addition to the SNPs disclosed in Tables 1-2). Variation can occur in either or both the coding and non-coding regions. The variations can  
25 produce conservative and/or non-conservative amino acid substitutions.

Further variants of the nucleic acid molecules disclosed in Tables 1-2, such as naturally occurring allelic variants (as well as orthologs and paralogs) and synthetic variants produced by mutagenesis techniques, can be identified and/or produced using methods well known in the art. Such further variants can comprise a nucleotide  
30 sequence that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with a nucleic acid sequence disclosed in Table 1 and/or Table 2 (or a fragment thereof) and that includes a novel SNP allele disclosed in Table 1 and/or Table 2. Further, variants can comprise a nucleotide

sequence that encodes a polypeptide that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with a polypeptide sequence disclosed in Table 1 (or a fragment thereof) and that includes a novel SNP allele disclosed in Table 1 and/or Table 2. Thus, the present invention

5 specifically contemplates isolated nucleic acid molecule that have a certain degree of sequence variation compared with the sequences shown in Tables 1-2, but that contain a novel SNP allele disclosed herein. In other words, as long as an isolated nucleic acid molecule contains a novel SNP allele disclosed herein, other portions of the nucleic acid molecule that flank the novel SNP allele can vary to some degree from the

10 specific transcript, genomic, and context sequences shown in Tables 1-2, and can encode a polypeptide that varies to some degree from the specific polypeptide sequences shown in Table 1.

To determine the percent identity of two amino acid sequences or two nucleotide sequences of two molecules that share sequence homology, the sequences

15 are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino

20 acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein, amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology").

25 The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between

30 two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular*

*Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch algorithm (*J. Mol. Biol.* (48):444-453 (1970))

5 which has been incorporated into the GAP program in the GCG software package, using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6.

In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software  
10 package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN  
15 program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4.

The nucleotide and amino acid sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be  
20 performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50,  
25 wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. In  
30 addition to BLAST, examples of other search and sequence comparison programs used in the art include, but are not limited to, FASTA (Pearson, *Methods Mol. Biol.* 25, 365-389 (1994)) and KERR (Dufresne *et al.*, *Nat Biotechnol* 2002 Dec;20(12):1269-71). For further information regarding bioinformatics techniques, see *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., N.Y.

The present invention further provides non-coding fragments of the nucleic acid molecules disclosed in Table 1 and/or Table 2. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, intronic sequences, 5' untranslated regions (UTRs), 3' untranslated regions, gene modulating sequences and gene termination sequences. Such fragments are useful, for example, in controlling heterologous gene expression and in developing screens to identify gene-modulating agents.

### **SNP Detection Reagents**

In a specific aspect of the present invention, the SNPs disclosed in Table 1 and/or Table 2, and their associated transcript sequences (provided in Table 1 as SEQ ID NOS:1-26), genomic sequences (provided in Table 2 as SEQ ID NOS:67-73), and context sequences (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:47-66; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:74-80), can be used for the design of SNP detection reagents. As used herein, a "SNP detection reagent" is a reagent that specifically detects a specific target SNP position disclosed herein, and that is preferably specific for a particular nucleotide (allele) of the target SNP position (i.e., the detection reagent preferably can differentiate between different alternative nucleotides at a target SNP position, thereby allowing the identity of the nucleotide present at the target SNP position to be determined). Typically, such detection reagent hybridizes to a target SNP-containing nucleic acid molecule by complementary base-pairing in a sequence specific manner, and discriminates the target variant sequence from other nucleic acid sequences such as an art-known form in a test sample. An example of a detection reagent is a probe that hybridizes to a target nucleic acid containing one or more of the SNPs provided in Table 1 and/or Table 2. In a preferred embodiment, such a probe can differentiate between nucleic acids having a particular nucleotide (allele) at a target SNP position from other nucleic acids that have a different nucleotide at the same target SNP position. In addition, a detection reagent may hybridize to a specific region 5' and/or 3' to a SNP position, particularly a region corresponding to the context sequences provided in Table 1 and/or Table 2 (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:47-66; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:74-80). Another example of a detection reagent is a primer which acts as an initiation point of nucleotide extension along a complementary strand of a target

polynucleotide. The SNP sequence information provided herein is also useful for designing primers, e.g. allele-specific primers, to amplify (e.g., using PCR) any SNP of the present invention.

In one preferred embodiment of the invention, a SNP detection reagent is an isolated or synthetic DNA or RNA polynucleotide probe or primer or PNA oligomer, or a combination of  
 5 DNA, RNA and/or PNA, that hybridizes to a segment of a target nucleic acid molecule containing a SNP identified in Table 1 and/or Table 2. A detection reagent in the form of a polynucleotide may optionally contain modified base analogs, intercalators or minor groove binders. Multiple detection reagents such as probes may be, for example, affixed to a solid support (e.g., arrays or beads) or supplied in solution (e.g., probe/primer sets for enzymatic reactions such as PCR, RT-PCR,  
 10 TaqMan assays, or primer-extension reactions) to form a SNP detection kit.

A probe or primer typically is a substantially purified oligonucleotide or PNA oligomer. Such oligonucleotide typically comprises a region of complementary nucleotide sequence that hybridizes under stringent conditions to at least about 8, 10, 12, 16, 18, 20, 22, 25, 30, 40, 50, 60, 100 (or any other number in-between) or more consecutive nucleotides in a target nucleic acid molecule.  
 15 Depending on the particular assay, the consecutive nucleotides can either include the target SNP position, or be a specific region in close enough proximity 5' and/or 3' to the SNP position to carry out the desired assay.

Other preferred primer and probe sequences can readily be determined using the transcript sequences (SEQ ID NOS:1-26), genomic sequences (SEQ ID NOS:67-73), and SNP context  
 20 sequences (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:47-66; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:74-80) disclosed in the Sequence Listing and in Tables 1-2. It will be apparent to one of skill in the art that such primers and probes are directly useful as reagents for genotyping the SNPs of the present invention, and can be incorporated into any kit/system format.

25 In order to produce a probe or primer specific for a target SNP-containing sequence, the gene/transcript and/or context sequence surrounding the SNP of interest is typically examined using a computer algorithm which starts at the 5' or at the 3' end

of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene/SNP context sequence, have a GC content within a range suitable for hybridization, lack predicted secondary structure that may interfere with hybridization, and/or possess other desired characteristics or that lack other undesired characteristics.

A primer or probe of the present invention is typically at least about 8 nucleotides in length. In one embodiment of the invention, a primer or a probe is at least about 10 nucleotides in length. In a preferred embodiment, a primer or a probe is at least about 12 nucleotides in length. In a more preferred embodiment, a primer or probe is at least about 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 nucleotides in length. While the maximal length of a probe can be as long as the target sequence to be detected, depending on the type of assay in which it is employed, it is typically less than about 50, 60, 65, or 70 nucleotides in length. In the case of a primer, it is typically less than about 30 nucleotides in length. In a specific preferred embodiment of the invention, a primer or a probe is within the length of about 18 and about 28 nucleotides. However, in other embodiments, such as nucleic acid arrays and other embodiments in which probes are affixed to a substrate, the probes can be longer, such as on the order of 30-70, 75, 80, 90, 100, or more nucleotides in length (see the section below entitled "SNP Detection Kits and Systems").

For analyzing SNPs, it may be appropriate to use oligonucleotides specific for alternative SNP alleles. Such oligonucleotides which detect single nucleotide variations in target sequences may be referred to by such terms as "allele-specific oligonucleotides", "allele-specific probes", or "allele-specific primers". The design and use of allele-specific probes for analyzing polymorphisms is described in, e.g., *Mutation Detection A Practical Approach*, ed. Cotton *et al.* Oxford University Press, 1998; Saiki *et al.*, *Nature* 324, 163-166 (1986); Dattagupta, EP235,726; and Saiki, WO 89/11548.

While the design of each allele-specific primer or probe depends on variables such as the precise composition of the nucleotide sequences flanking a SNP position in a target nucleic acid molecule, and the length of the primer or probe, another factor in the use of primers and probes is the stringency of the condition under which the hybridization between the probe or primer and the target sequence is performed. Higher stringency conditions utilize buffers with lower ionic strength and/or a higher reaction temperature, and tend to require a more perfect match between probe/primer

and a target sequence in order to form a stable duplex. If the stringency is too high, however, hybridization may not occur at all. In contrast, lower stringency conditions utilize buffers with higher ionic strength and/or a lower reaction temperature, and permit the formation of stable duplexes with more mismatched bases between a probe/primer and a target sequence. By way of example and not limitation, exemplary conditions for high stringency hybridization conditions using an allele-specific probe are as follows: Prehybridization with a solution containing 5X standard saline phosphate EDTA (SSPE), 0.5% NaDodSO<sub>4</sub> (SDS) at 55°C, and incubating probe with target nucleic acid molecules in the same solution at the same temperature, followed by washing with a solution containing 2X SSPE, and 0.1% SDS at 55°C or room temperature.

Moderate stringency hybridization conditions may be used for allele-specific primer extension reactions with a solution containing, e.g., about 50mM KCl at about 46°C. Alternatively, the reaction may be carried out at an elevated temperature such as 60°C. In another embodiment, a moderately stringent hybridization condition suitable for oligonucleotide ligation assay (OLA) reactions wherein two probes are ligated if they are completely complementary to the target sequence may utilize a solution of about 100mM KCl at a temperature of 46°C.

In a hybridization-based assay, allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms (e.g., alternative SNP alleles/nucleotides) in the respective DNA segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant detectable difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles or significantly more strongly to one allele. While a probe may be designed to hybridize to a target sequence that contains a SNP site such that the SNP site aligns anywhere along the sequence of the probe, the probe is preferably designed to hybridize to a segment of the target sequence such that the SNP site aligns with a central position of the probe (e.g., a position within the probe that is at least three nucleotides from either end of the probe). This design of probe generally achieves good discrimination in hybridization between different allelic forms.

In another embodiment, a probe or primer may be designed to hybridize to a segment of target DNA such that the SNP aligns with either the 5' most end or the 3' most end of the probe or primer. In a specific preferred embodiment which is particularly suitable for use in a oligonucleotide ligation assay (U.S. Patent No.

5 4,988,617), the 3' most nucleotide of the probe aligns with the SNP position in the target sequence.

Oligonucleotide probes and primers may be prepared by methods well known in the art. Chemical synthetic methods include, but are limited to, the phosphotriester method described by Narang *et al.*, 1979, *Methods in Enzymology* 68:90; the  
10 phosphodiester method described by Brown *et al.*, 1979, *Methods in Enzymology* 68:109, the diethylphosphoamidate method described by Beaucage *et al.*, 1981, *Tetrahedron Letters* 22:1859; and the solid support method described in U.S. Patent No. 4,458,066.

Allele-specific probes are often used in pairs (or, less commonly, in sets of 3  
15 or 4, such as if a SNP position is known to have 3 or 4 alleles, respectively, or to assay both strands of a nucleic acid molecule for a target SNP allele), and such pairs may be identical except for a one nucleotide mismatch that represents the allelic variants at the SNP position. Commonly, one member of a pair perfectly matches a reference form of a target sequence that has a more common SNP allele (i.e., the allele that is  
20 more frequent in the target population) and the other member of the pair perfectly matches a form of the target sequence that has a less common SNP allele (i.e., the allele that is rarer in the target population). In the case of an array, multiple pairs of probes can be immobilized on the same support for simultaneous analysis of multiple different polymorphisms.

25 In one type of PCR-based assay, an allele-specific primer hybridizes to a region on a target nucleic acid molecule that overlaps a SNP position and only primes amplification of an allelic form to which the primer exhibits perfect complementarity (Gibbs, 1989, *Nucleic Acid Res.* 17 2427-2448). Typically, the primer's 3'-most nucleotide is aligned with and complementary to the SNP position of the target  
30 nucleic acid molecule. This primer is used in conjunction with a second primer that hybridizes at a distal site. Amplification proceeds from the two primers, producing a detectable product that indicates which allelic form is present in the test sample. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect

complementarity to a distal site. The single-base mismatch prevents amplification or substantially reduces amplification efficiency, so that either no detectable product is formed or it is formed in lower amounts or at a slower pace. The method generally works most effectively when the mismatch is at the 3'-most position of the oligonucleotide (i.e., the 3'-most position of the  
 5 oligonucleotide aligns with the target SNP position) because this position is most destabilizing to elongation from the primer (see, e.g., WO 93/22456). This PCR-based assay can be utilized as part of the TaqMan assay, described below.

In a specific embodiment of the invention, a primer of the invention contains a sequence substantially complementary to a segment of a target SNP-containing nucleic acid molecule except that  
 10 the primer has a mismatched nucleotide in one of the three nucleotide positions at the 3'-most end of the primer, such that the mismatched nucleotide does not base pair with a particular allele at the SNP site. In a preferred embodiment, the mismatched nucleotide in the primer is the second from the last nucleotide at the 3'-most position of the primer. In a more preferred embodiment, the mismatched nucleotide in the primer is the last nucleotide at the 3'-most position of the primer.

In another embodiment of the invention, a SNP detection reagent of the invention is labeled with a fluorogenic reporter dye that emits a detectable signal. While the preferred reporter dye is a fluorescent dye, any reporter dye that can be attached to a detection reagent such as an oligonucleotide probe or primer is suitable for use in the invention. Such dyes include, but are not limited to, Acridine, AMCA, BODIPY, Cascade Blue, Cy2, Cy3, Cy5, Cy7, Dabcyl, Edans, Eosin, Erythrosin, Fluorescein,  
 15 6-Fam, Tet, Joe, Hex, Oregon Green, Rhodamine, Rhodol Green, Tamra, Rox, and Texas Red.

In yet another embodiment of the invention, the detection reagent may be further labeled with a quencher dye such as Tamra, especially when the reagent is used as a self-quenching probe such as a TaqMan (U.S. Patent Nos. 5,210,015 and 5,538,848) or Molecular Beacon probe (U.S. Patent Nos. 5,118,801 and 5,312,728), or other stemless or linear beacon probe (Livak *et al.*, 1995, PCR Method  
 25 Appl. 4:357-362; Tyagi *et al.*, 1996, Nature Biotechnology 14: 303-308; Nazarenko *et al.*, 1997, Nucl. Acids Res. 25:2516-2521; U.S. Patent Nos. 5,866,336 and 6,117,635).

The detection reagents of the invention may also contain other labels, including but not limited to, biotin for streptavidin binding, hapten for antibody binding, and

oligonucleotide for binding to another complementary oligonucleotide such as pairs of zipcodes.

The present invention also contemplates reagents that do not contain (or that are complementary to) a SNP nucleotide identified herein but that are used to assay one or more SNPs disclosed herein. For example, primers that flank, but do not hybridize directly to a target SNP position provided herein are useful in primer extension reactions in which the primers hybridize to a region adjacent to the target SNP position (i.e., within one or more nucleotides from the target SNP site). During the primer extension reaction, a primer is typically not able to extend past a target SNP site if a particular nucleotide (allele) is present at that target SNP site, and the primer extension product can readily be detected in order to determine which SNP allele is present at the target SNP site. For example, particular ddNTPs are typically used in the primer extension reaction to terminate primer extension once a ddNTP is incorporated into the extension product (a primer extension product which includes a ddNTP at the 3'-most end of the primer extension product, and in which the ddNTP corresponds to a SNP disclosed herein, is a composition that is encompassed by the present invention). Thus, reagents that bind to a nucleic acid molecule in a region adjacent to a SNP site, even though the bound sequences do not necessarily include the SNP site itself, are also encompassed by the present invention.

#### **SNP Detection Kits and Systems**

A person skilled in the art will recognize that, based on the SNP and associated sequence information disclosed herein, detection reagents can be developed and used to assay any SNP of the present invention individually or in combination, and such detection reagents can be readily incorporated into one of the established kit or system formats which are well known in the art. The terms "kits" and "systems", as used herein in the context of SNP detection reagents, are intended to refer to such things as combinations of multiple SNP detection reagents, or one or more SNP detection reagents in combination with one or more other types of elements or components (e.g., other types of biochemical reagents, containers, packages such as

packaging intended for commercial sale, substrates to which SNP detection reagents are attached, electronic hardware components, etc.). Accordingly, the present invention further provides SNP detection kits and systems, including but not limited to, packaged probe and primer sets (e.g., TaqMan probe/primer sets),

5 arrays/microarrays of nucleic acid molecules, and beads that contain one or more probes, primers, or other detection reagents for detecting one or more SNPs of the present invention. The kits/systems can optionally include various electronic hardware components; for example, arrays ("DNA chips") and microfluidic systems ("lab-on-a-chip" systems) provided by various manufacturers typically comprise hardware  
10 components. Other kits/systems (e.g., probe/primer sets) may not include electronic hardware components, but may be comprised of, for example, one or more SNP detection reagents (along with, optionally, other biochemical reagents) packaged in one or more containers.

In some embodiments, a SNP detection kit typically contains one or more  
15 detection reagents and other components (e.g., a buffer, enzymes such as DNA polymerases or ligases, chain extension nucleotides such as deoxynucleotide triphosphates, and in the case of Sanger-type DNA sequencing reactions, chain terminating nucleotides, positive control sequences, negative control sequences, and the like) necessary to carry out an assay or reaction, such as amplification and/or  
20 detection of a SNP-containing nucleic acid molecule. A kit may further contain means for determining the amount of a target nucleic acid, and means for comparing the amount with a standard, and can comprise instructions for using the kit to detect the SNP-containing nucleic acid molecule of interest. In one embodiment of the present invention, kits are provided which contain the necessary reagents to carry out one or  
25 more assays to detect one or more SNPs disclosed herein. In a preferred embodiment of the present invention, SNP detection kits/systems are in the form of nucleic acid arrays, or compartmentalized kits, including microfluidic/lab-on-a-chip systems.

SNP detection kits/systems may contain, for example, one or more probes, or pairs of probes, that hybridize to a nucleic acid molecule at or near each target SNP  
30 position. Multiple pairs of allele-specific probes may be included in the kit/system to simultaneously assay large numbers of SNPs, at least one of which is a SNP of the present invention. In some kits/systems, the allele-specific probes are immobilized to a substrate such as an array or bead. For example, the same substrate can comprise allele-specific probes for detecting at least 1; 10; 100; 1000; 10,000; 100,000 (or any

other number in-between) or substantially all of the SNPs shown in Table 1 and/or Table 2.

The terms “arrays”, “microarrays”, and “DNA chips” are used herein interchangeably to refer to an array of distinct polynucleotides affixed to a substrate, such as glass, plastic, paper, nylon or other type of membrane, filter, chip, or any other suitable solid support. The

5 polynucleotides can be synthesized directly on the substrate, or synthesized separate from the substrate and then affixed to the substrate. In one embodiment, the microarray is prepared and used according to the methods described in U.S. Patent No. 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; *Nat. Biotech.* 14: 1675-1680) and Schena, M. et al. (1996; *Proc. Natl. Acad. Sci.* 93: 10614-10619). In other embodiments, such arrays are  
10 produced by the methods described by Brown et al., U.S. Patent No. 5,807,522.

Nucleic acid arrays are reviewed in the following references: Zammattéo et al., “New chips for molecular biology and diagnostics”, *Biotechnol Annu Rev.* 2002;8:85-101; Sosnowski et al., “Active microelectronic array system for DNA hybridization, genotyping and pharmacogenomic applications”, *Psychiatr Genet.* 2002 Dec;12(4):181-92; Heller, “DNA microarray technology:  
15 devices, systems, and applications”, *Annu Rev Biomed Eng.* 2002;4:129-53. Epub 2002 Mar 22; Kolchinsky et al., “Analysis of SNPs and other genomic variations using gel-based chips”, *Hum Mutat.* 2002 Apr;19(4):343-60; and McGall et al., “High-density genechip oligonucleotide probe arrays”, *Adv Biochem Eng Biotechnol.* 2002;77:21-42.

Any number of probes, such as allele-specific probes, may be implemented in an array, and  
20 each probe or pair of probes can hybridize to a different SNP position. In the case of polynucleotide probes, they can be synthesized at designated areas (or synthesized separately and then affixed to designated areas) on a substrate using a light-directed chemical process. Each DNA chip can contain, for example, thousands to millions of

individual synthetic polynucleotide probes arranged in a grid-like pattern and miniaturized (e.g., to the size of a dime). Preferably, probes are attached to a solid support in an ordered, addressable array.

5 A microarray can be composed of a large number of unique, single-stranded polynucleotides; usually either synthetic antisense polynucleotides or fragments of cDNAs, fixed to a solid support. Typical polynucleotides are preferably about 6-60 nucleotides in length, more preferably about 15-30 nucleotides in length, and most preferably about 18-25 nucleotides in length. For certain types of microarrays or other detection kits/systems, it may be preferable to use oligonucleotides that are only about 10 7-20 nucleotides in length. In other types of arrays, such as arrays used in conjunction with chemiluminescent detection technology, preferred probe lengths can be, for example, about 15-80 nucleotides in length, preferably about 50-70 nucleotides in length, more preferably about 55-65 nucleotides in length, and most preferably about 60 nucleotides in length. The microarray or detection kit can contain polynucleotides 15 that cover the known 5' or 3' sequence of a gene/transcript or target SNP site, sequential polynucleotides that cover the full-length sequence of a gene/transcript; or unique polynucleotides selected from particular areas along the length of a target gene/transcript sequence, particularly areas corresponding to one or more SNPs disclosed in Table 1 and/or Table 2. Polynucleotides used in the microarray or 20 detection kit can be specific to a SNP or SNPs of interest (e.g., specific to a particular SNP allele at a target SNP site, or specific to particular SNP alleles at multiple different SNP sites), or specific to a polymorphic gene/transcript or genes/transcripts of interest.

Hybridization assays based on polynucleotide arrays rely on the 25 differences in hybridization stability of the probes to perfectly matched and mismatched target sequence variants. For SNP genotyping, it is generally preferable that stringency conditions used in hybridization assays are high enough such that nucleic acid molecules that differ from one another at as little as a single SNP position can be differentiated (e.g., 30 typical SNP hybridization assays are designed so that hybridization will occur only if one particular nucleotide is present at a SNP position, but will not occur if an alternative nucleotide is present at that SNP position). Such high stringency conditions may be preferable when using, for example,

nucleic acid arrays of allele-specific probes for SNP detection. Such high stringency conditions are described in the preceding section, and are well known to those skilled in the art and can be found in, for example, *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.

In other embodiments, the arrays are used in conjunction with chemiluminescent detection technology. The following patents and patent applications, provide additional information pertaining to chemiluminescent detection: U.S. patent applications 10/620332 and 10/620333 describe chemiluminescent approaches for microarray detection; U.S. Patent Nos. 6124478, 6107024, 5994073, 5981768, 5871938, 5843681, 5800999, and 5773628 describe methods and compositions of dioxetane for performing chemiluminescent detection; and U.S. published application US2002/0110828 discloses methods and compositions for microarray controls.

In one embodiment of the invention, a nucleic acid array can comprise an array of probes of about 15-25 nucleotides in length. In further embodiments, a nucleic acid array can comprise any number of probes, in which at least one probe is capable of detecting one or more SNPs disclosed in Table 1 and/or Table 2, and/or at least one probe comprises a fragment of one of the sequences selected from the group consisting of those disclosed in Table 1, Table 2, the Sequence Listing, and sequences complementary thereto, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 12, 15, 16, 18, 20, more preferably 22, 25, 30, 40, 47, 50, 55, 60, 65, 70, 80, 90, 100, or more consecutive nucleotides (or any other number in-between) and containing (or being complementary to) a novel SNP allele disclosed in Table 1 and/or Table 2. In some embodiments, the nucleotide complementary to the SNP site is within 5, 4, 3, 2, or 1 nucleotide from the center of the probe, more preferably at the center of said probe.

A polynucleotide probe can be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding

procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more polynucleotides, or any other number which lends itself to the efficient use of commercially available instrumentation.

Using such arrays or other kits/systems, the present invention provides methods of identifying the SNPs disclosed herein in a test sample. Such methods typically involve incubating a test sample of nucleic acids with an array comprising one or more probes corresponding to at least one SNP position of the present invention, and assaying for binding of a nucleic acid from the test sample with one or more of the probes. Conditions for incubating a SNP detection reagent (or a kit/system that employs one or more such SNP detection reagents) with a test sample vary. Incubation conditions depend on such factors as the format employed in the assay, the detection methods employed, and the type and nature of the detection reagents used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification and array assay formats can readily be adapted to detect the SNPs disclosed herein.

A SNP detection kit/system of the present invention may include components that are used to prepare nucleic acids from a test sample for the subsequent amplification and/or detection of a SNP-containing nucleic acid molecule. Such sample preparation components can be used to produce nucleic acid extracts (including DNA and/or RNA), proteins or membrane extracts from any bodily fluids (such as blood, serum, plasma, urine, saliva, phlegm, gastric juices, semen, tears, sweat, etc.), skin, hair, cells (especially nucleated cells), biopsies, buccal swabs or tissue specimens. The test samples used in the above-described methods will vary based on such factors as the assay format, nature of the detection method, and the specific tissues, cells or extracts used as the test sample to be assayed. Methods of preparing nucleic acids, proteins, and cell extracts are well known in the art and can be readily adapted to obtain a sample that is compatible with the system utilized. Automated sample preparation systems for extracting nucleic acids from a test sample are commercially available, and examples are Qiagen's BioRobot 9600,

Applied Biosystems' PRISM 6700, and Roche Molecular Systems' COBAS AmpliPrep System.

Another form of kit contemplated by the present invention is a compartmentalized kit. A compartmentalized kit includes any kit in which reagents are  
5 contained in separate containers. Such containers include, for example, small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allow one to efficiently transfer reagents from one compartment to another compartment such that the test samples and reagents are not cross-contaminated, or from one container to another vessel not included in the kit,  
10 and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another or to another vessel. Such containers may include, for example, one or more containers which will accept the test sample, one or more containers which contain at least one probe or other SNP detection reagent for detecting one or more SNPs of the present invention, one or more containers which  
15 contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and one or more containers which contain the reagents used to reveal the presence of the bound probe or other SNP detection reagents. The kit can optionally further comprise compartments and/or reagents for, for example, nucleic acid amplification or other enzymatic reactions such as primer extension reactions, hybridization, ligation,  
20 electrophoresis (preferably capillary electrophoresis), mass spectrometry, and/or laser-induced fluorescent detection. The kit may also include instructions for using the kit. Exemplary compartmentalized kits include microfluidic devices known in the art (see, e.g., Weigl et al., "Lab-on-a-chip for drug development", *Adv Drug Deliv Rev.* 2003 Feb 24;55(3):349-77). In such microfluidic devices, the containers may be referred to as, for  
25 example, microfluidic "compartments", "chambers", or "channels".

Microfluidic devices, which may also be referred to as "lab-on-a-chip" systems, biomedical micro-electro-mechanical systems (bioMEMs), or multicomponent integrated systems, are exemplary kits/systems of the present invention for analyzing SNPs. Such systems miniaturize and  
30 compartmentalize processes such as probe/target hybridization, nucleic acid amplification, and capillary electrophoresis reactions in a single functional device. Such microfluidic devices typically utilize detection reagents in at least one aspect of the system, and such detection reagents

may be used to detect one or more SNPs of the present invention. One example of a microfluidic system is disclosed in U.S. Patent No. 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips. Exemplary microfluidic systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples may be controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts. Varying the voltage can be used as a means to control the liquid flow at intersections between the micro-machined channels and to change the liquid flow rate for pumping across different sections of the microchip. See, for example, U.S. Patent Nos. 6,153,073, Dubrow *et al.*, and 6,156,181, Parce *et al.*

For genotyping SNPs, an exemplary microfluidic system may integrate, for example, nucleic acid amplification, primer extension, capillary electrophoresis, and a detection method such as laser induced fluorescence detection. In a first step of an exemplary process for using such an exemplary system, nucleic acid samples are amplified, preferably by PCR. Then, the amplification products are subjected to automated primer extension reactions using ddNTPs (specific fluorescence for each ddNTP) and the appropriate oligonucleotide primers to carry out primer extension reactions which hybridize just upstream of the targeted SNP. Once the extension at the 3' end is completed, the primers are separated from the unincorporated fluorescent ddNTPs by capillary electrophoresis. The separation medium used in capillary electrophoresis can be, for example, polyacrylamide, polyethyleneglycol or dextran. The incorporated ddNTPs in the single nucleotide primer extension products are identified by laser-induced fluorescence detection. Such an exemplary microchip can be used to process, for example, at least 96 to 384 samples, or more, in parallel.

## **USES OF NUCLEIC ACID MOLECULES**

The nucleic acid molecules of the present invention have a variety of uses, especially in the diagnosis and treatment of myocardial infarction. For example, the nucleic acid molecules are useful as hybridization probes, such as for genotyping SNPs in messenger RNA, transcript, cDNA, genomic DNA, amplified DNA or other nucleic

acid molecules, and for isolating full-length cDNA and genomic clones encoding the variant peptides disclosed in Table 1 as well as their orthologs.

5 A probe can hybridize to any nucleotide sequence along the entire length of a nucleic acid molecule provided in Table 1 and/or Table 2. Preferably, a probe of the present invention hybridizes to a region of a target sequence that encompasses a SNP position indicated in Table 1 and/or Table 2. More preferably, a probe hybridizes to a SNP-containing target sequence in a sequence-specific manner such that it distinguishes the target sequence from other nucleotide sequences which vary from the target sequence only by which nucleotide is present at the SNP site. Such a probe is particularly useful  
10 for detecting the presence of a SNP-containing nucleic acid in a test sample, or for determining which nucleotide (allele) is present at a particular SNP site (i.e., genotyping the SNP site).

A nucleic acid hybridization probe may be used for determining the presence, level, form, and/or distribution of nucleic acid expression. The  
15 nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes specific for the SNPs described herein can be used to assess the presence, expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in gene expression relative to normal levels. *In vitro*  
20 techniques for detection of mRNA include, for example, Northern blot hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern blot hybridizations and *in situ* hybridizations (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY).

25 Probes can be used as part of a diagnostic test kit for identifying cells or tissues in which a variant protein is expressed, such as by measuring the level of a variant protein-encoding nucleic acid (e.g., mRNA) in a sample of cells from a subject or determining if a polynucleotide contains a SNP of interest.

Thus, the nucleic acid molecules of the invention can be used as  
30 hybridization probes to detect the SNPs disclosed herein, thereby determining whether an individual with the polymorphisms is at risk for myocardial infarction or has developed early stage myocardial infarction. Detection of a SNP associated with a disease phenotype provides a

diagnostic tool for an active disease and/or genetic predisposition to the disease.

5 The nucleic acid molecules of the invention are also useful as primers to amplify any given region of a nucleic acid molecule, particularly a region containing a SNP identified in Table 1 and/or Table 2.

10 The nucleic acid molecules of the invention are also useful for constructing recombinant vectors (described in greater detail below). Such vectors include expression vectors that express a portion of, or all of, any of the variant peptide sequences provided in Table 1. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced SNPs.

15 The nucleic acid molecules of the invention are also useful for expressing antigenic portions of the variant proteins, particularly antigenic portions that contain a variant amino acid sequence (e.g., an amino acid substitution) caused by a SNP disclosed in Table 1 and/or Table 2.

20 The nucleic acid molecules of the invention are also useful for constructing vectors containing a gene regulatory region of the nucleic acid molecules of the present invention.

The nucleic acid molecules of the invention are also useful for designing ribozymes corresponding to all, or a part, of an mRNA molecule expressed from a SNP-containing nucleic acid molecule described herein.

25 The nucleic acid molecules of the invention are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and variant peptides.

30 The nucleic acid molecules of the invention are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and variant peptides. The production of recombinant cells and transgenic animals having nucleic acid molecules which contain the SNPs disclosed in Table 1 and/or Table 2 allow, for example, effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules of the invention are also useful in assays for drug screening to identify compounds that, for example, modulate nucleic acid expression.

The nucleic acid molecules of the invention are also useful in gene therapy in patients whose cells have aberrant gene expression. Thus, recombinant cells, which include a patient's cells that have been engineered *ex vivo* and returned to the patient, can be introduced into an individual where the recombinant cells produce the desired protein to treat the individual.

### **SNP Genotyping Methods**

The process of determining which specific nucleotide (i.e., allele) is present at each of one or more SNP positions, such as a SNP position in a nucleic acid molecule disclosed in Table 1 and/or Table 2, is referred to as SNP genotyping. The present invention provides methods of SNP genotyping, such as for use in screening for myocardial infarction or related pathologies, or determining predisposition thereto, or determining responsiveness to a form of treatment, or in genome mapping or SNP association analysis, etc.

Nucleic acid samples can be genotyped to determine which allele(s) is/are present at any given genetic region (e.g., SNP position) of interest by methods well known in the art. The neighboring sequence can be used to design SNP detection reagents such as oligonucleotide probes, which may optionally be implemented in a kit format. Exemplary SNP genotyping methods are described in Chen et al., "Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput", *Pharmacogenomics J.* 2003;3(2):77-96; Kwok et al., "Detection of single nucleotide polymorphisms", *Curr Issues Mol Biol.* 2003 Apr;5(2):43-60; Shi, "Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes", *Am J Pharmacogenomics.* 2002;2(3):197-205; and Kwok, "Methods for genotyping single nucleotide polymorphisms", *Annu Rev Genomics Hum Genet* 2001;2:235-58. Exemplary techniques for high-throughput SNP genotyping are described in Marnellos, "High-throughput SNP analysis for genetic association studies", *Curr Opin Drug Discov Devel.* 2003 May;6(3):317-21. Common SNP genotyping methods include, but are not limited to, TaqMan assays, molecular beacon assays, nucleic acid arrays, allele-specific primer extension, allele-specific PCR, arrayed primer

extension, homogeneous primer extension assays, primer extension with detection by mass spectrometry, pyrosequencing, multiplex primer extension sorted on genetic arrays, ligation with rolling circle amplification, homogeneous ligation, OLA (U.S. Patent No. 4,988,167), multiplex ligation reaction sorted on genetic arrays, restriction-fragment  
5 length polymorphism, single base extension-tag assays, and the Invader assay. Such methods may be used in combination with detection mechanisms such as, for example, luminescence or chemiluminescence detection, fluorescence detection, time-resolved fluorescence detection, fluorescence resonance energy transfer, fluorescence polarization, mass spectrometry, and electrical detection.

10 Various methods for detecting polymorphisms include, but are not limited to, methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985); Cotton *et al.*, *PNAS* 85:4397 (1988); and Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), comparison of the electrophoretic mobility of variant and wild type nucleic acid  
15 molecules (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and assaying the movement of polymorphic or wild-type fragments in polyacrylamide gels containing a gradient of denaturant using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.*, *Nature* 313:495 (1985)). Sequence variations at specific locations can also be  
20 assessed by nuclease protection assays such as RNase and S1 protection or chemical cleavage methods.

In a preferred embodiment, SNP genotyping is performed using the TaqMan assay, which is also known as the 5' nuclease assay (U.S. Patent Nos. 5,210,015 and 5,538,848). The TaqMan assay detects the  
25 accumulation of a specific amplified product during PCR. The TaqMan assay utilizes an oligonucleotide probe labeled with a fluorescent reporter dye and a quencher dye. The reporter dye is excited by irradiation at an appropriate wavelength, it transfers energy to the quencher dye in the same probe via a process called fluorescence resonance energy transfer  
30 (FRET). When attached to the probe, the excited reporter dye does not emit a signal. The proximity of the quencher dye to the reporter dye in the intact probe maintains a reduced fluorescence for the reporter. The reporter dye and quencher dye may be at the 5' most and the 3' most ends,

respectively, or vice versa. Alternatively, the reporter dye may be at the 5' or 3' most end while the quencher dye is attached to an internal nucleotide, or vice versa. In yet another embodiment, both the reporter and the quencher may be attached to internal nucleotides at a distance  
5 from each other such that fluorescence of the reporter is reduced.

During PCR, the 5' nuclease activity of DNA polymerase cleaves the probe, thereby separating the reporter dye and the quencher dye and resulting in increased fluorescence of the reporter. Accumulation of PCR product is detected directly by monitoring the increase in fluorescence of  
10 the reporter dye. The DNA polymerase cleaves the probe between the reporter dye and the quencher dye only if the probe hybridizes to the target SNP-containing template which is amplified during PCR, and the probe is designed to hybridize to the target SNP site only if a particular SNP allele is present.

15 Preferred TaqMan primer and probe sequences can readily be determined using the SNP and associated nucleic acid sequence information provided herein. A number of computer programs, such as Primer Express (Applied Biosystems, Foster City, CA), can be used to rapidly obtain optimal primer/probe sets. It will be apparent to one of skill in the art that such primers and probes for detecting the SNPs of the  
20 present invention are useful in diagnostic assays for myocardial infarction and related pathologies, and can be readily incorporated into a kit format. The present invention also includes modifications of the Taqman assay well known in the art such as the use of Molecular Beacon probes (U.S. Patent Nos. 5,118,801 and 5,312,728) and other variant formats (U.S. Patent Nos. 5,866,336 and 6,117,635).

25 Another preferred method for genotyping the SNPs of the present invention is the use of two oligonucleotide probes in an OLA (see, e.g., U.S. Patent No. 4,988,617). In this method, one probe hybridizes to a segment of a target nucleic acid with its 3' most end aligned with the SNP site. A second probe hybridizes to an adjacent segment of the target nucleic acid molecule directly 3' to the first probe. The  
30 two juxtaposed probes hybridize to the target nucleic acid molecule, and are ligated in the presence of a linking agent such as a ligase if there is perfect complementarity between the 3' most nucleotide of the first probe with the SNP site. If there is a mismatch, ligation would not occur. After the reaction, the ligated probes are

separated from the target nucleic acid molecule, and detected as indicators of the presence of a SNP.

The following patents, patent applications, and published international patent applications, provide additional information pertaining to techniques for carrying out various types of OLA: U.S. Patent Nos. 6027889, 6268148, 5494810, 5830711, and 6054564 describe OLA strategies for performing SNP detection; WO 97/31256 and WO 00/56927 describe OLA strategies for performing SNP detection using universal arrays, wherein a zipcode sequence can be introduced into one of the hybridization probes, and the resulting product, or amplified product, hybridized to a universal zip code array; U.S. application US01/17329 (and 09/584,905) describes OLA (or LDR) followed by PCR, wherein zipcodes are incorporated into OLA probes, and amplified PCR products are determined by electrophoretic or universal zipcode array readout; U.S. applications 60/427818, 60/445636, and 60/445494 describe SNPLEX methods and software for multiplexed SNP detection using OLA followed by PCR, wherein zipcodes are incorporated into OLA probes, and amplified PCR products are hybridized with a zipchute reagent, and the identity of the SNP determined from electrophoretic readout of the zipchute. In some embodiments, OLA is carried out prior to PCR (or another method of nucleic acid amplification). In other embodiments, PCR (or another method of nucleic acid amplification) is carried out prior to OLA.

Another method for SNP genotyping is based on mass spectrometry. Mass spectrometry takes advantage of the unique mass of each of the four nucleotides of DNA. SNPs can be unambiguously genotyped by mass spectrometry by measuring the differences in the mass of nucleic acids having alternative SNP alleles. MALDI-TOF (Matrix Assisted Laser Desorption Ionization – Time of Flight) mass spectrometry technology is preferred for extremely precise determinations of molecular mass, such as SNPs. Numerous approaches to SNP analysis have been developed based on mass spectrometry. Preferred mass spectrometry-based methods of SNP genotyping include primer extension assays, which can also be utilized in combination with other approaches, such as traditional gel-based formats and microarrays.

Typically, the primer extension assay involves designing and annealing a primer to a template PCR amplicon upstream (5') from a target SNP position. A mix of dideoxynucleotide triphosphates (ddNTPs) and/or deoxynucleotide triphosphates

(dNTPs) are added to a reaction mixture containing template (e.g., a SNP-containing nucleic acid molecule which has typically been amplified, such as by PCR), primer, and DNA polymerase. Extension of the primer terminates at the first position in the template where a nucleotide complementary to one of the ddNTPs in the mix occurs.

5 The primer can be either immediately adjacent (i.e., the nucleotide at the 3' end of the primer hybridizes to the nucleotide next to the target SNP site) or two or more nucleotides removed from the SNP position. If the primer is several nucleotides removed from the target SNP position, the only limitation is that the template sequence between the 3' end of the primer and the SNP position cannot contain a

10 nucleotide of the same type as the one to be detected, or this will cause premature termination of the extension primer. Alternatively, if all four ddNTPs alone, with no dNTPs, are added to the reaction mixture, the primer will always be extended by only one nucleotide, corresponding to the target SNP position. In this instance, primers are designed to bind one nucleotide upstream from the SNP position (i.e., the nucleotide

15 at the 3' end of the primer hybridizes to the nucleotide that is immediately adjacent to the target SNP site on the 5' side of the target SNP site). Extension by only one nucleotide is preferable, as it minimizes the overall mass of the extended primer, thereby increasing the resolution of mass differences between alternative SNP nucleotides. Furthermore, mass-tagged ddNTPs can be employed in the primer

20 extension reactions in place of unmodified ddNTPs. This increases the mass difference between primers extended with these ddNTPs, thereby providing increased sensitivity and accuracy, and is particularly useful for typing heterozygous base positions. Mass-tagging also alleviates the need for intensive sample-preparation procedures and decreases the necessary resolving power of the mass spectrometer.

25 The extended primers can then be purified and analyzed by MALDI-TOF mass spectrometry to determine the identity of the nucleotide present at the target SNP position. In one method of analysis, the products from the primer extension reaction are combined with light absorbing crystals that form a matrix. The matrix is then hit with an energy source such as a laser to ionize and desorb the nucleic acid

30 molecules into the gas-phase. The ionized molecules are then ejected into a flight tube and accelerated down the tube towards a detector. The time between the ionization event, such as a laser pulse, and collision of the molecule with the detector is the time of flight of that molecule. The time of flight is precisely correlated with the mass-to-charge ratio ( $m/z$ ) of the ionized molecule. Ions with smaller  $m/z$  travel down the

tube faster than ions with larger  $m/z$  and therefore the lighter ions reach the detector before the heavier ions. The time-of-flight is then converted into a corresponding, and highly precise,  $m/z$ . In this manner, SNPs can be identified based on the slight differences in mass, and the corresponding time of flight differences, inherent in

5 nucleic acid molecules having different nucleotides at a single base position. For further information regarding the use of primer extension assays in conjunction with MALDI-TOF mass spectrometry for SNP genotyping, see, e.g., Wise *et al.*, "A standard protocol for single nucleotide primer extension in the human genome using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", *Rapid*

10 *Commun Mass Spectrom.* 2003;17(11):1195-202.

The following references provide further information describing mass spectrometry-based methods for SNP genotyping: Bocker, "SNP and mutation discovery using base-specific cleavage and MALDI-TOF mass spectrometry", *Bioinformatics.* 2003 Jul;19 Suppl 1:I44-I53; Storm *et al.*, "MALDI-TOF mass

15 spectrometry-based SNP genotyping", *Methods Mol Biol.* 2003;212:241-62; Jurinke *et al.*, "The use of MassARRAY technology for high throughput genotyping", *Adv Biochem Eng Biotechnol.* 2002;77:57-74; and Jurinke *et al.*, "Automated genotyping using the DNA MassArray technology", *Methods Mol Biol.* 2002;187:179-92.

SNPs can also be scored by direct DNA sequencing. A variety of automated

20 sequencing procedures can be utilized ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)). The nucleic acid sequences of the present invention enable one of ordinary skill in the art to readily design sequencing

25 primers for such automated sequencing procedures. Commercial instrumentation, such as the Applied Biosystems 377, 3100, 3700, 3730, and 3730xl DNA Analyzers (Foster City, CA), is commonly used in the art for automated sequencing.

Other methods that can be used to genotype the SNPs of the present invention include single-strand conformational polymorphism (SSCP), and denaturing gradient

30 gel electrophoresis (DGGE) (Myers *et al.*, *Nature* 313:495 (1985)). SSCP identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, *Proc. Nat. Acad.* Single-stranded PCR products can be generated by heating or otherwise denaturing double stranded PCR products. Single-stranded nucleic acids may refold or form secondary structures that are

partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products are related to base-sequence differences at SNP positions. DGGE differentiates SNP alleles based on the different sequence-dependent stabilities and melting properties inherent in polymorphic DNA and the corresponding differences in electrophoretic migration patterns in a denaturing gradient gel (Erlich, ed., PCR Technology, *Principles and Applications for DNA Amplification*, W.H. Freeman and Co, New York, 1992, Chapter 7).

Sequence-specific ribozymes (U.S. Patent No. 5,498,531) can also be used to score SNPs based on the development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature. If the SNP affects a restriction enzyme cleavage site, the SNP can be identified by alterations in restriction enzyme digestion patterns, and the corresponding changes in nucleic acid fragment lengths determined by gel electrophoresis.

SNP genotyping can include the steps of, for example, collecting a biological sample from a human subject (e.g., sample of tissues, cells, fluids, secretions, etc.), isolating nucleic acids (e.g., genomic DNA, mRNA or both) from the cells of the sample, contacting the nucleic acids with one or more primers which specifically hybridize to a region of the isolated nucleic acid containing a target SNP under conditions such that hybridization and amplification of the target nucleic acid region occurs, and determining the nucleotide present at the SNP position of interest, or, in some assays, detecting the presence or absence of an amplification product (assays can be designed so that hybridization and/or amplification will only occur if a particular SNP allele is present or absent). In some assays, the size of the amplification product is detected and compared to the length of a control sample; for example, deletions and insertions can be detected by a change in size of the amplified product compared to a normal genotype.

SNP genotyping is useful for numerous practical applications, as described below. Examples of such applications include, but are not limited to, SNP-disease association analysis, disease predisposition screening, disease diagnosis, disease prognosis, disease progression monitoring, determining therapeutic strategies based

on an individual's genotype ("pharmacogenomics"), developing therapeutic agents based on SNP genotypes associated with a disease or likelihood of responding to a drug, stratifying a patient population for clinical trial for a treatment regimen, predicting the likelihood that an individual will experience toxic side effects from a therapeutic agent, and human identification applications such as forensics.

#### **Analysis of Genetic Association Between SNPs and Phenotypic Traits**

SNP genotyping for disease diagnosis, disease predisposition screening, disease prognosis, determining drug responsiveness (pharmacogenomics), drug toxicity screening, and other uses described herein, typically relies on initially establishing a genetic association between one or more specific SNPs and the particular phenotypic traits of interest.

Different study designs may be used for genetic association studies (*Modern Epidemiology*, Lippincott Williams & Wilkins (1998), 609-622). Observational studies are most frequently carried out in which the response of the patients is not interfered with. The first type of observational study identifies a sample of persons in whom the suspected cause of the disease is present and another sample of persons in whom the suspected cause is absent, and then the frequency of development of disease in the two samples is compared. These sampled populations are called cohorts, and the study is a prospective study. The other type of observational study is case-control or a retrospective study. In typical case-control studies, samples are collected from individuals with the phenotype of interest (cases) such as certain manifestations of a disease, and from individuals without the phenotype (controls) in a population (target population) that conclusions are to be drawn from. Then the possible causes of the disease are investigated retrospectively. As the time and costs of collecting samples in case-control studies are considerably less than those for prospective studies, case-control studies are the more commonly used study design in genetic association studies, at least during the exploration and discovery stage.

In both types of observational studies, there may be potential confounding factors that should be taken into consideration. Confounding factors are those that are associated with both the real cause(s) of the disease and the disease itself, and they include demographic information such as age, gender, ethnicity as well as environmental factors. When confounding factors are not matched in cases and controls in a study, and are not controlled properly, spurious association results can

arise. If potential confounding factors are identified, they should be controlled for by analysis methods explained below.

In a genetic association study, the cause of interest to be tested is a certain allele or a SNP or a combination of alleles or a haplotype from several SNPs. Thus, tissue specimens (e.g., whole blood) from the sampled individuals may be collected and genomic DNA genotyped for the SNP(s) of interest. In addition to the phenotypic trait of interest, other information such as demographic (e.g., age, gender, ethnicity, etc.), clinical, and environmental information that may influence the outcome of the trait can be collected to further characterize and define the sample set. In many cases, these factors are known to be associated with diseases and/or SNP allele frequencies. There are likely gene-environment and/or gene-gene interactions as well. Analysis methods to address gene-environment and gene-gene interactions (for example, the effects of the presence of both susceptibility alleles at two different genes can be greater than the effects of the individual alleles at two genes combined) are discussed below.

After all the relevant phenotypic and genotypic information has been obtained, statistical analyses are carried out to determine if there is any significant correlation between the presence of an allele or a genotype with the phenotypic characteristics of an individual. Preferably, data inspection and cleaning are first performed before carrying out statistical tests for genetic association. Epidemiological and clinical data of the samples can be summarized by descriptive statistics with tables and graphs. Data validation is preferably performed to check for data completion, inconsistent entries, and outliers. Chi-squared tests and t-tests (Wilcoxon rank-sum tests if distributions are not normal) may then be used to check for significant differences between cases and controls for discrete and continuous variables, respectively. To ensure genotyping quality, Hardy-Weinberg disequilibrium tests can be performed on cases and controls separately. Significant deviation from Hardy-Weinberg equilibrium (HWE) in both cases and controls for individual markers can be indicative of genotyping errors. If HWE is violated in a majority of markers, it is indicative of population substructure that should be further investigated. Moreover, Hardy-Weinberg disequilibrium in cases only can indicate genetic association of the markers with the disease (*Genetic Data Analysis*, Weir B., Sinauer (1990)).

To test whether an allele of a single SNP is associated with the case or control status of a phenotypic trait, one skilled in the art can compare allele frequencies in

cases and controls. Standard chi-squared tests and Fisher exact tests can be carried out on a 2x2 table (2 SNP alleles x 2 outcomes in the categorical trait of interest). To test whether genotypes of a SNP are associated, chi-squared tests can be carried out on a 3x2 table (3 genotypes x 2 outcomes). Score tests are also carried out for

5 genotypic association to contrast the three genotypic frequencies (major homozygotes, heterozygotes and minor homozygotes) in cases and controls, and to look for trends using 3 different modes of inheritance, namely dominant (with contrast coefficients 2, -1, -1), additive (with contrast coefficients 1, 0, -1) and recessive (with contrast coefficients 1, 1, -2). Odds ratios for minor versus major alleles, and odds ratios for

10 heterozygote and homozygote variants versus the wild type genotypes are calculated with the desired confidence limits, usually 95%.

In order to control for confounders and to test for interaction and effect modifiers, stratified analyses may be performed using stratified factors that are likely to be confounding, including demographic information such as age, ethnicity, and

15 gender, or an interacting element or effect modifier, such as a known major gene (e.g., APOE for Alzheimer's disease or HLA genes for autoimmune diseases), or environmental factors such as smoking in lung cancer. Stratified association tests may be carried out using Cochran-Mantel-Haenszel tests that take into account the ordinal nature of genotypes with 0, 1, and 2 variant alleles. Exact tests by StatXact

20 may also be performed when computationally possible. Another way to adjust for confounding effects and test for interactions is to perform stepwise multiple logistic regression analysis using statistical packages such as SAS or R. Logistic regression is a model-building technique in which the best fitting and most parsimonious model is built to describe the relation between the dichotomous outcome (for instance, getting

25 a certain disease or not) and a set of independent variables (for instance, genotypes of different associated genes, and the associated demographic and environmental factors). The most common model is one in which the logit transformation of the odds ratios is expressed as a linear combination of the variables (main effects) and their cross-product terms (interactions) (Applied Logistic Regression, Hosmer and

30 Lemeshow, Wiley (2000)). To test whether a certain variable or interaction is significantly associated with the outcome, coefficients in the model are first estimated and then tested for statistical significance of their departure from zero.

In addition to performing association tests one marker at a time, haplotype association analysis may also be performed to study a number of markers that are

closely linked together. Haplotype association tests can have better power than genotypic or allelic association tests when the tested markers are not the disease-causing mutations themselves but are in linkage disequilibrium with such mutations. The test will even be more powerful if the disease is indeed caused by a combination of alleles on a haplotype (e.g., APOE is a haplotype formed by 2 SNPs that are very close to each other). In order to perform haplotype association effectively, marker-marker linkage disequilibrium measures, both  $D'$  and  $R^2$ , are typically calculated for the markers within a gene to elucidate the haplotype structure. Recent studies (Daly et al, *Nature Genetics*, 29, 232-235, 2001) in linkage disequilibrium indicate that SNPs within a gene are organized in block pattern, and a high degree of linkage disequilibrium exists within blocks and very little linkage disequilibrium exists between blocks. Haplotype association with the disease status can be performed using such blocks once they have been elucidated.

Haplotype association tests can be carried out in a similar fashion as the allelic and genotypic association tests. Each haplotype in a gene is analogous to an allele in a multi-allelic marker. One skilled in the art can either compare the haplotype frequencies in cases and controls or test genetic association with different pairs of haplotypes. It has been proposed (Schaid et al, *Am. J. Hum. Genet.*, 70, 425-434, 2002) that score tests can be done on haplotypes using the program "haplo.score". In that method, haplotypes are first inferred by EM algorithm and score tests are carried out with a generalized linear model (GLM) framework that allows the adjustment of other factors.

An important decision in the performance of genetic association tests is the determination of the significance level at which significant association can be declared when the p-value of the tests reaches that level. In an exploratory analysis where positive hits will be followed up in subsequent confirmatory testing, an unadjusted p-value  $<0.1$  (a significance level on the lenient side) may be used for generating hypotheses for significant association of a SNP with certain phenotypic characteristics of a disease. It is preferred that a p-value  $<0.05$  (a significance level traditionally used in the art) is achieved in order for a SNP to be considered to have an association with a disease. It is more preferred that a p-value  $<0.01$  (a significance level on the stringent side) is achieved for an association to be declared. When hits are followed up in confirmatory analyses in more samples of the same source or in different samples from different sources, adjustment for multiple testing will be

performed as to avoid excess number of hits while maintaining the experiment-wise error rates at 0.05. While there are different methods to adjust for multiple testing to control for different kinds of error rates, a commonly used but rather conservative method is Bonferroni correction to control the experiment-wise or family-wise error rate (*Multiple comparisons and multiple tests*, Westfall et al, SAS Institute (1999)).  
5 Permutation tests to control for the false discovery rates, FDR, can be more powerful (Benjamini and Hochberg, Journal of the Royal Statistical Society, Series B 57, 1289-1300, 1995, *Resampling-based Multiple Testing*, Westfall and Young, Wiley (1993)). Such methods to control for multiplicity would be preferred when the tests are  
10 dependent and controlling for false discovery rates is sufficient as opposed to controlling for the experiment-wise error rates.

In replication studies using samples from different populations after statistically significant markers have been identified in the exploratory stage, meta-analyses can then be performed by combining evidence of different studies (*Modern*  
15 *Epidemiology*, Lippincott Williams & Wilkins, 1998, 643-673). If available, association results known in the art for the same SNPs can be included in the meta-analyses.

Since both genotyping and disease status classification can involve errors, sensitivity analyses may be performed to see how odds ratios and  
20 p-values would change upon various estimates on genotyping and disease classification error rates.

It has been well known that subpopulation-based sampling bias between cases and controls can lead to spurious results in case-control association studies (Ewens and Spielman, *Am. J. Hum. Genet.* 62, 450-  
25 458, 1995) when prevalence of the disease is associated with different subpopulation groups. Such bias can also lead to a loss of statistical power in genetic association studies. To detect population stratification, Pritchard and Rosenberg (Pritchard et al. *Am. J. Hum. Gen.* 1999, 65:220-228) suggested typing markers that are unlinked to the disease and using  
30 results of association tests on those markers to determine whether there is any population stratification. When stratification is detected, the genomic control (GC) method as proposed by Devlin and Roeder (Devlin et al. *Biometrics* 1999, 55:997-1004) can be used to adjust for the inflation of test

statistics due to population stratification. GC method is robust to changes in population structure levels as well as being applicable to DNA pooling designs (Devlin et al. *Genet. Epidemiol.* 2000, 21:273-284).

While Pritchard's method recommended using 15-20 unlinked microsatellite markers, it suggested using more than 30 biallelic markers to get enough power to detect population stratification. For the GC method, it has been shown (Bacanu et al. *Am. J. Hum. Genet.* 2000, 66:1933-1944) that about 60-70 biallelic markers are sufficient to estimate the inflation factor for the test statistics due to population stratification. Hence, 70 intergenic SNPs can be chosen in unlinked regions as indicated in a genome scan (Kehoe et al. *Hum. Mol. Genet.* 1999, 8:237-245).

Once individual risk factors, genetic or non-genetic, have been found for the predisposition to disease, the next step is to set up a classification/prediction scheme to predict the category (for instance, disease or no-disease) that an individual will be in depending on his genotypes of associated SNPs and other non-genetic risk factors. Logistic regression for discrete trait and linear regression for continuous trait are standard techniques for such tasks (*Applied Regression Analysis*, Draper and Smith, Wiley (1998)). Moreover, other techniques can also be used for setting up classification. Such techniques include, but are not limited to, MART, CART, neural network, and discriminant analyses that are suitable for use in comparing the performance of different methods (*The Elements of Statistical Learning*, Hastie, Tibshirani & Friedman, Springer (2002)).

#### **Disease Diagnosis and Predisposition Screening**

Information on association/correlation between genotypes and disease-related phenotypes can be exploited in several ways. For example, in the case of a highly statistically significant association between one or more SNPs with predisposition to a disease for which treatment is available, detection of such a genotype pattern in an individual may justify immediate administration of treatment, or at least the institution of regular monitoring of the individual. Detection of the susceptibility alleles associated with serious disease in a couple contemplating having children may also be valuable to the couple in their reproductive decisions. In the case of a weaker but still statistically significant association between a SNP and a human disease, immediate therapeutic intervention or monitoring may not be justified after detecting the

susceptibility allele or SNP. Nevertheless, the subject can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little or no cost to the individual but would confer potential benefits in reducing the risk of developing conditions for which that individual may have an increased risk by virtue  
5 of having the susceptibility allele(s).

The SNPs of the invention may contribute to myocardial infarction in an individual in different ways. Some polymorphisms occur within a protein coding sequence and contribute to disease phenotype by affecting protein structure. Other polymorphisms occur in noncoding regions but may exert phenotypic effects  
10 indirectly via influence on, for example, replication, transcription, and/or translation. A single SNP may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by multiple SNPs in different genes.

As used herein, the terms "diagnose", "diagnosis", and "diagnostics" include, but are not limited to any of the following: detection of myocardial infarction that an  
15 individual may presently have or be at risk for, predisposition screening (i.e., determining the increased risk for an individual in developing myocardial infarction in the future, or determining whether an individual has a decreased risk of developing myocardial infarction in the future; in the case of recurrent myocardial infarction (RMI), predisposition screening may typically involve determining the risk that an  
20 individual who has previously had a myocardial infarction will develop another myocardial infarction in the future, or determining whether an individual who has not experienced a myocardial infarction will be at risk for developing recurrent myocardial infarctions in the future), determining a particular type or subclass of myocardial infarction in an individual known to have myocardial infarction,  
25 confirming or reinforcing a previously made diagnosis of myocardial infarction, pharmacogenomic evaluation of an individual to determine which therapeutic strategy that individual is most likely to positively respond to or to predict whether a patient is likely to respond to a particular treatment, predicting whether a patient is likely to experience toxic effects from a particular treatment or therapeutic compound, and  
30 evaluating the future prognosis of an individual having myocardial infarction. Such diagnostic uses are based on the SNPs individually or in a unique combination or SNP haplotypes of the present invention.

Haplotypes are particularly useful in that, for example, fewer SNPs can be genotyped to determine if a particular genomic region harbors a locus that influences

a particular phenotype, such as in linkage disequilibrium-based SNP association analysis.

Linkage disequilibrium (LD) refers to the co-inheritance of alleles (e.g., alternative nucleotides) at two or more different SNP sites at frequencies greater than  
5 would be expected from the separate frequencies of occurrence of each allele in a given population. The expected frequency of co-occurrence of two alleles that are inherited independently is the frequency of the first allele multiplied by the frequency of the second allele. Alleles that co-occur at expected frequencies are said to be in "linkage equilibrium". In contrast, LD refers to any non-random genetic association  
10 between allele(s) at two or more different SNP sites, which is generally due to the physical proximity of the two loci along a chromosome. LD can occur when two or more SNPs sites are in close physical proximity to each other on a given chromosome and therefore alleles at these SNP sites will tend to remain unseparated for multiple generations with the consequence that a particular nucleotide (allele) at one SNP site  
15 will show a non-random association with a particular nucleotide (allele) at a different SNP site located nearby. Hence, genotyping one of the SNP sites will give almost the same information as genotyping the other SNP site that is in LD.

For diagnostic purposes, if a particular SNP site is found to be useful for diagnosing myocardial infarction, then the skilled artisan would recognize that other  
20 SNP sites which are in LD with this SNP site would also be useful for diagnosing the condition. Various degrees of LD can be encountered between two or more SNPs with the result being that some SNPs are more closely associated (i.e., in stronger LD) than others. Furthermore, the physical distance over which LD extends along a chromosome differs between different regions of the genome, and therefore the  
25 degree of physical separation between two or more SNP sites necessary for LD to occur can differ between different regions of the genome.

For diagnostic applications, polymorphisms (e.g., SNPs and/or haplotypes) that are not the actual disease-causing (causative) polymorphisms, but are in LD with such causative polymorphisms, are also useful. In such instances, the genotype of the  
30 polymorphism(s) that is/are in LD with the causative polymorphism is predictive of the genotype of the causative polymorphism and, consequently, predictive of the phenotype (e.g., myocardial infarction) that is influenced by the causative SNP(s). Thus, polymorphic markers that are in LD with causative polymorphisms are useful

as diagnostic markers, and are particularly useful when the actual causative polymorphism(s) is/are unknown.

Linkage disequilibrium in the human genome is reviewed in: Wall et al., "Haplotype blocks and linkage disequilibrium in the human genome", *Nat Rev Genet.* 2003 Aug;4(8):587-97; Garner et al., "On selecting markers for association studies: patterns of linkage disequilibrium between two and three diallelic loci", *Genet Epidemiol.* 2003 Jan;24(1):57-67; Ardlie et al., "Patterns of linkage disequilibrium in the human genome", *Nat Rev Genet.* 2002 Apr;3(4):299-309 (erratum in *Nat Rev Genet* 2002 Jul;3(7):566); and Remm et al., "High-density genotyping and linkage disequilibrium in the human genome using chromosome 22 as a model"; *Curr Opin Chem Biol.* 2002 Feb;6(1):24-30.

The contribution or association of particular SNPs and/or SNP haplotypes with disease phenotypes, such as myocardial infarction, enables the SNPs of the present invention to be used to develop superior diagnostic tests capable of identifying individuals who express a detectable trait, such as myocardial infarction, as the result of a specific genotype, or individuals whose genotype places them at an increased or decreased risk of developing a detectable trait at a subsequent time as compared to individuals who do not have that genotype. As described herein, diagnostics may be based on a single SNP or a group of SNPs. Combined detection of a plurality of SNPs (for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 24, 25, 30, 32, 48, 50, 64, 96, 100, or any other number in-between, or more, of the SNPs provided in Table 1 and/or Table 2) typically increases the probability of an accurate diagnosis. For example, the presence of a single SNP known to correlate with myocardial infarction might indicate a probability of 20% that an individual has or is at risk of developing myocardial infarction, whereas detection of five SNPs, each of which correlates with myocardial infarction, might indicate a probability of 80% that an individual has or is at risk of developing myocardial infarction. To further increase the accuracy of diagnosis or predisposition screening, analysis of the SNPs of the present invention can be combined with that of other polymorphisms or other risk factors of myocardial infarction, such as disease symptoms, pathological

characteristics, family history, diet, environmental factors or lifestyle factors.

It will, of course, be understood by practitioners skilled in the treatment or diagnosis of myocardial infarction that the present invention generally does not intend  
5 to provide an absolute identification of individuals who are at risk (or less at risk) of developing myocardial infarction, and/or pathologies related to myocardial infarction, but rather to indicate a certain increased (or decreased) degree or likelihood of developing the disease based on statistically significant association results. However, this information is extremely valuable as it can be used to, for example, initiate  
10 preventive treatments or to allow an individual carrying one or more significant SNPs or SNP haplotypes to foresee warning signs such as minor clinical symptoms, or to have regularly scheduled physical exams to monitor for appearance of a condition in order to identify and begin treatment of the condition at an early stage. Particularly with diseases that are extremely debilitating or fatal if not treated on time, the  
15 knowledge of a potential predisposition, even if this predisposition is not absolute, would likely contribute in a very significant manner to treatment efficacy.

The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a SNP or a SNP pattern associated with an increased or decreased risk of developing a detectable trait or  
20 whether the individual suffers from a detectable trait as a result of a particular polymorphism/mutation, including, for example, methods which enable the analysis of individual chromosomes for haplotyping, family studies, single sperm DNA analysis, or somatic hybrids. The trait analyzed using the diagnostics of the invention may be any detectable trait that is commonly observed in pathologies and disorders  
25 related to myocardial infarction.

Another aspect of the present invention relates to a method of determining whether an individual is at risk (or less at risk) of developing one or more traits or whether an individual expresses one or more traits as a consequence of possessing a particular trait-causing or trait-influencing allele. These methods generally involve  
30 obtaining a nucleic acid sample from an individual and assaying the nucleic acid sample to determine which nucleotide(s) is/are present at one or more SNP positions, wherein the assayed nucleotide(s) is/are indicative of an increased or decreased risk of

developing the trait or indicative that the individual expresses the trait as a result of possessing a particular trait-causing or trait-influencing allele.

In another embodiment, the SNP detection reagents of the present invention are used to determine whether an individual has one or more SNP allele(s) affecting the level (e.g., the concentration of mRNA or protein in a sample, etc.) or pattern (e.g., the kinetics of expression, rate of decomposition, stability profile,  $K_m$ ,  $V_{max}$ , etc.) of gene expression (collectively, the "gene response" of a cell or bodily fluid). Such a determination can be accomplished by screening for mRNA or protein expression (e.g., by using nucleic acid arrays, RT-PCR, TaqMan assays, or mass spectrometry), identifying genes having altered expression in an individual, genotyping SNPs disclosed in Table 1 and/or Table 2 that could affect the expression of the genes having altered expression (e.g., SNPs that are in and/or around the gene(s) having altered expression, SNPs in regulatory/control regions, SNPs in and/or around other genes that are involved in pathways that could affect the expression of the gene(s) having altered expression, or all SNPs could be genotyped), and correlating SNP genotypes with altered gene expression. In this manner, specific SNP alleles at particular SNP sites can be identified that affect gene expression.

#### **Pharmacogenomics and Therapeutics/Drug Development**

The present invention provides methods for assessing the pharmacogenomics of a subject harboring particular SNP alleles or haplotypes to a particular therapeutic agent or pharmaceutical compound, or to a class of such compounds. Pharmacogenomics deals with the roles which clinically significant hereditary variations (e.g., SNPs) play in the response to drugs due to altered drug disposition and/or abnormal action in affected persons. See, e.g., Roses, *Nature* 405, 857-865 (2000); Gould Rothberg, *Nature Biotechnology* 19, 209-211 (2001); Eichelbaum, *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996); and Linder, *Clin. Chem.* 43(2):254-266 (1997). The clinical outcomes of these variations can result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the SNP genotype of an individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. For example, SNPs in drug metabolizing enzymes can affect the activity of these enzymes, which in turn can affect both the intensity and duration of drug action, as well as drug metabolism and clearance.

The discovery of SNPs in drug metabolizing enzymes, drug transporters, proteins for pharmaceutical agents, and other drug targets has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. SNPs can be expressed in the phenotype of the extensive metabolizer and in the phenotype of the poor metabolizer. Accordingly, SNPs may lead to allelic variants of a protein in which one or more of the protein functions in one population are different from those in another population. SNPs and the encoded variant peptides thus provide targets to ascertain a genetic predisposition that can affect treatment modality. For example, in a ligand-based treatment, SNPs may give rise to amino terminal extracellular domains and/or other ligand-binding regions of a receptor that are more or less active in ligand binding, thereby affecting subsequent protein activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing particular SNP alleles or haplotypes.

As an alternative to genotyping, specific variant proteins containing variant amino acid sequences encoded by alternative SNP alleles could be identified. Thus, pharmacogenomic characterization of an individual permits the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic uses based on the individual's SNP genotype, thereby enhancing and optimizing the effectiveness of the therapy. Furthermore, the production of recombinant cells and transgenic animals containing particular SNPs/haplotypes allow effective clinical design and testing of treatment compounds and dosage regimens. For example, transgenic animals can be produced that differ only in specific SNP alleles in a gene that is orthologous to a human disease susceptibility gene.

Pharmacogenomic uses of the SNPs of the present invention provide several significant advantages for patient care, particularly in treating myocardial infarction. Pharmacogenomic characterization of an individual, based on an individual's SNP genotype, can identify those individuals unlikely to respond to treatment with a particular medication and thereby allows physicians to avoid prescribing the ineffective medication to those individuals. On the other hand, SNP genotyping of an individual may enable physicians to select the appropriate medication and dosage regimen that will be most effective based on an individual's SNP genotype. This information increases a physician's confidence in prescribing medications and motivates patients to comply with their drug regimens. Furthermore, pharmacogenomics may identify patients predisposed

to toxicity and adverse reactions to particular drugs or drug dosages. Adverse drug reactions lead to more than 100,000 avoidable deaths per year in the United States alone and therefore represent a significant cause of hospitalization and death, as well as a significant economic burden on the healthcare system (Pfof et al., *Trends in*  
 5 *Biotechnology*, Aug. 2000.). Thus, pharmacogenomics based on the SNPs disclosed herein has the potential to both save lives and reduce healthcare costs substantially.

Pharmacogenomics in general is discussed further in Rose et al.,  
 “Pharmacogenetic analysis of clinically relevant genetic polymorphisms”, *Methods*  
*Mol Med.* 2003;85:225-37. Pharmacogenomics as it relates to Alzheimer’s disease  
 10 and other neurodegenerative disorders is discussed in Cacabelos, “Pharmacogenomics for the treatment of dementia”, *Ann Med.* 2002;34(5):357-79, Maimone, et al.,  
 “Pharmacogenomics of neurodegenerative diseases”, *Eur J Pharmacol.* 2001 Feb 9;413(1):11-29, and Poirier, “Apolipoprotein E: a pharmacogenetic target for the treatment of Alzheimer’s disease”, *Mol Diagn.* 1999 Dec;4(4):335-41.  
 15 Pharmacogenomics as it relates to cardiovascular disorders is discussed in Siest et al.,  
 “Pharmacogenomics of drugs affecting the cardiovascular system”, *Clin Chem Lab Med.* 2003 Apr;41(4):590-9, Mukherjee et al., “Pharmacogenomics in cardiovascular diseases”, *Prog Cardiovasc Dis.* 2002 May-Jun;44(6):479-98, and Mooser et al.,  
 “Cardiovascular pharmacogenetics in the SNP era”, *J Thromb Haemost.* 2003  
 20 Jul;1(7):1398-402. Pharmacogenomics as it relates to cancer is discussed in McLeod et al., “Cancer pharmacogenomics: SNPs, chips, and the individual patient”, *Cancer Invest.* 2003;21(4):630-40 and Watters et al., “Cancer pharmacogenomics: current and future applications”, *Biochim Biophys Acta.* 2003 Mar 17;1603(2):99-111.

The SNPs of the present invention also can be used to identify novel  
 25 therapeutic targets for myocardial infarction. For example, genes containing the disease-associated variants (“variant genes”) or their products, as well as genes or their products that are directly or indirectly regulated by or interacting with these variant genes or their products, can be targeted for the development of therapeutics that, for example, treat  
 30 the disease or prevent or delay disease onset. The therapeutics may be composed of, for example, small molecules, proteins, protein fragments or peptides, antibodies, nucleic acids, or their derivatives or mimetics which modulate the functions or levels of the target genes or gene products.

The SNP-containing nucleic acid molecules disclosed herein, and their complementary nucleic acid molecules, may be used as antisense constructs to control gene expression in cells, tissues, and organisms. Antisense technology is well established in the art and extensively reviewed in *Antisense Drug Technology: Principles, Strategies, and Applications*, Crooke (ed.), Marcel Dekker, Inc.: New York (2001). An antisense nucleic acid molecule is generally designed to be complementary to a region of mRNA expressed by a gene so that the antisense molecule hybridizes to the mRNA and thereby blocks translation of mRNA into protein. Various classes of antisense oligonucleotides are used in the art, two of which are cleavers and blockers. Cleavers, by binding to target RNAs, activate intracellular nucleases (e.g., RNaseH or RNase L) that cleave the target RNA. Blockers, which also bind to target RNAs, inhibit protein translation through steric hindrance of ribosomes. Exemplary blockers include peptide nucleic acids, morpholinos, locked nucleic acids, and methylphosphonates (see, e.g., Thompson, *Drug Discovery Today*, 7 (17): 912-917 (2002)). Antisense oligonucleotides are directly useful as therapeutic agents, and are also useful for determining and validating gene function (e.g., in gene knock-out or knock-down experiments).

Antisense technology is further reviewed in: Lavery et al., "Antisense and RNAi: powerful tools in drug target discovery and validation", *Curr Opin Drug Discov Devel.* 2003 Jul;6(4):561-9; Stephens et al., "Antisense oligonucleotide therapy in cancer", *Curr Opin Mol Ther.* 2003 Apr;5(2):118-22; Kurreck, "Antisense technologies. Improvement through novel chemical modifications", *Eur J Biochem.* 2003 Apr;270(8):1628-44; Dias et al., "Antisense oligonucleotides: basic concepts and mechanisms", *Mol Cancer Ther.* 2002 Mar;1(5):347-55; Chen, "Clinical development of antisense oligonucleotides as anti-cancer therapeutics", *Methods Mol Med.* 2003;75:621-36; Wang et al., "Antisense anticancer oligonucleotide therapeutics", *Curr Cancer Drug Targets.* 2001 Nov;1(3):177-96; and Bennett, "Efficiency of antisense oligonucleotide drug discovery", *Antisense Nucleic Acid Drug Dev.* 2002 Jun;12(3):215-24.

The SNPs of the present invention are particularly useful for designing antisense reagents that are specific for particular nucleic acid variants. Based on the SNP information disclosed herein, antisense oligonucleotides can be produced that specifically target mRNA molecules that contain one or more particular SNP nucleotides. In this manner, expression of mRNA molecules that contain one or more

undesired polymorphisms (e.g., SNP nucleotides that lead to a defective protein such as an amino acid substitution in a catalytic domain) can be inhibited or completely blocked. Thus, antisense oligonucleotides can be used to specifically bind a particular polymorphic form (e.g., a SNP allele that encodes a defective protein), thereby inhibiting translation of this form, but which do not bind an alternative polymorphic form (e.g., an alternative SNP nucleotide that encodes a protein having normal function).

Antisense molecules can be used to inactivate mRNA in order to inhibit gene expression and production of defective proteins. Accordingly, these molecules can be used to treat a disorder, such as myocardial infarction, characterized by abnormal or undesired gene expression or expression of certain defective proteins. This technique can involve cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible mRNA regions include, for example, protein-coding regions and particularly protein-coding regions corresponding to catalytic activities, substrate/ligand binding, or other functional activities of a protein.

The SNPs of the present invention are also useful for designing RNA interference reagents that specifically target nucleic acid molecules having particular SNP variants. RNA interference (RNAi), also referred to as gene silencing, is based on using double-stranded RNA (dsRNA) molecules to turn genes off. When introduced into a cell, dsRNAs are processed by the cell into short fragments (generally about 21-22bp in length) known as small interfering RNAs (siRNAs) which the cell uses in a sequence-specific manner to recognize and destroy complementary RNAs (Thompson, *Drug Discovery Today*, 7 (17): 912-917 (2002)). Thus, because RNAi molecules, including siRNAs, act in a sequence-specific manner, the SNPs of the present invention can be used to design RNAi reagents that recognize and destroy nucleic acid molecules having specific SNP alleles/nucleotides (such as deleterious alleles that lead to the production of defective proteins), while not affecting nucleic acid molecules having alternative SNP alleles (such as alleles that encode proteins having normal function). As with antisense reagents, RNAi reagents may be directly useful as therapeutic agents (e.g., for turning off defective, disease-causing genes), and are also useful for characterizing and validating gene function (e.g., in gene knock-out or knock-down experiments).

The following references provide a further review of RNAi: Agami, "RNAi and related mechanisms and their potential use for therapy", *Curr Opin Chem Biol.* 2002 Dec;6(6):829-34; Lavery et al., "Antisense and RNAi: powerful tools in drug target discovery and validation", *Curr Opin Drug Discov Devel.* 2003 Jul;6(4):561-9;  
5 Shi, "Mammalian RNAi for the masses", *Trends Genet* 2003 Jan;19(1):9-12, Shuey et al., "RNAi: gene-silencing in therapeutic intervention", *Drug Discovery Today* 2002 Oct;7(20):1040-1046; McManus et al., *Nat Rev Genet* 2002 Oct;3(10):737-47; Xia et al., *Nat Biotechnol* 2002 Oct;20(10):1006-10; Plasterk et al., *Curr Opin Genet Dev* 2000 Oct;10(5):562-7; Bosher et al., *Nat Cell Biol* 2000 Feb;2(2):E31-6; and  
10 Hunter, *Curr Biol* 1999 Jun 17;9(12):R440-2).

A subject suffering from a pathological condition, such as myocardial infarction, ascribed to a SNP may be treated so as to correct the genetic defect (see Kren et al., *Proc. Natl. Acad. Sci. USA* 96:10349-10354 (1999)). Such a subject can be identified by any method that can detect the polymorphism in a biological sample  
15 drawn from the subject. Such a genetic defect may be permanently corrected by administering to such a subject a nucleic acid fragment incorporating a repair sequence that supplies the normal/wild-type nucleotide at the position of the SNP. This site-specific repair sequence can encompass an RNA/DNA oligonucleotide that operates to promote endogenous repair of a subject's genomic DNA. The site-specific  
20 repair sequence is administered in an appropriate vehicle, such as a complex with polyethylenimine, encapsulated in anionic liposomes, a viral vector such as an adenovirus, or other pharmaceutical composition that promotes intracellular uptake of the administered nucleic acid. A genetic defect leading to an inborn pathology may then be overcome, as the chimeric oligonucleotides induce incorporation of the  
25 normal sequence into the subject's genome. Upon incorporation, the normal gene product is expressed, and the replacement is propagated, thereby engendering a permanent repair and therapeutic enhancement of the clinical condition of the subject.

In cases in which a cSNP results in a variant protein that is ascribed to be the cause of, or a contributing factor to, a pathological condition, a method of treating  
30 such a condition can include administering to a subject experiencing the pathology the wild-type/normal cognate of the variant protein. Once administered in an effective dosing regimen, the wild-type cognate provides complementation or remediation of the pathological condition.

The invention further provides a method for identifying a compound or agent that can be used to treat myocardial infarction. The SNPs disclosed herein are useful as targets for the identification and/or development of therapeutic agents. A method for identifying a therapeutic agent or compound typically includes assaying the ability of the agent or compound to modulate the activity and/or expression of a SNP-containing nucleic acid or the encoded product and thus identifying an agent or a compound that can be used to treat a disorder characterized by undesired activity or expression of the SNP-containing nucleic acid or the encoded product. The assays can be performed in cell-based and cell-free systems. Cell-based assays can include cells naturally expressing the nucleic acid molecules of interest or recombinant cells genetically engineered to express certain nucleic acid molecules.

Variant gene expression in a myocardial infarction patient can include, for example, either expression of a SNP-containing nucleic acid sequence (for instance, a gene that contains a SNP can be transcribed into an mRNA transcript molecule containing the SNP, which can in turn be translated into a variant protein) or altered expression of a normal/wild-type nucleic acid sequence due to one or more SNPs (for instance, a regulatory/control region can contain a SNP that affects the level or pattern of expression of a normal transcript).

Assays for variant gene expression can involve direct assays of nucleic acid levels (e.g., mRNA levels), expressed protein levels, or of collateral compounds involved in a signal pathway. Further, the expression of genes that are up- or down-regulated in response to the signal pathway can also be assayed. In this embodiment, the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Modulators of variant gene expression can be identified in a method wherein, for example, a cell is contacted with a candidate compound/agent and the expression of mRNA determined. The level of expression of mRNA in the presence of the candidate compound is compared to the level of expression of mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of variant gene expression based on this comparison and be used to treat a disorder such as myocardial infarction that is characterized by variant gene expression (e.g., either expression of a SNP-containing nucleic acid or altered expression of a normal/wild-type nucleic acid molecule due to one or more SNPs that affect expression of the nucleic acid molecule) due to one or more SNPs of the present invention. When expression of

mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is  
5 identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the SNP or associated nucleic acid domain (e.g., catalytic domain, ligand/substrate-binding domain, regulatory/control region, etc.) or gene, or the encoded mRNA transcript, as a target, using a compound identified through drug screening as a gene modulator to modulate  
10 variant nucleic acid expression. Modulation can include either up-regulation (i.e., activation or agonization) or down-regulation (i.e., suppression or antagonization) of nucleic acid expression.

Expression of mRNA transcripts and encoded proteins, either wild type or variant, may be altered in individuals with a particular SNP allele in a regulatory/control  
15 element, such as a promoter or transcription factor binding domain, that regulates expression. In this situation, methods of treatment and compounds can be identified, as discussed herein, that regulate or overcome the variant regulatory/control element, thereby generating normal, or healthy, expression levels of either the wild type or variant protein.

20 The SNP-containing nucleic acid molecules of the present invention are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of a variant gene, or encoded product, in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as an indicator for the continuing effectiveness of treatment with the compound, particularly with compounds to which a  
25 patient can develop resistance, as well as an indicator for toxicities. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression  
30 falls below a desirable level, administration of the compound could be commensurately decreased.

In another aspect of the present invention, there is provided a pharmaceutical pack comprising a therapeutic agent (e.g., a small molecule drug, antibody, peptide, antisense or RNAi nucleic acid molecule, etc.) and a set of instructions for

administration of the therapeutic agent to humans diagnostically tested for one or more SNPs or SNP haplotypes provided by the present invention.

The SNPs/haplotypes of the present invention are also useful for improving many different aspects of the drug development process. For example, individuals can be selected for clinical trials based on their SNP genotype. Individuals with SNP genotypes that indicate that they are most likely to respond to the drug can be included in the trials and those individuals whose SNP genotypes indicate that they are less likely to or would not respond to the drug, or suffer adverse reactions, can be eliminated from the clinical trials. This not only improves the safety of clinical trials, but also will enhance the chances that the trial will demonstrate statistically significant efficacy. Furthermore, the SNPs of the present invention may explain why certain previously developed drugs performed poorly in clinical trials and may help identify a subset of the population that would benefit from a drug that had previously performed poorly in clinical trials, thereby "rescuing" previously developed drugs, and enabling the drug to be made available to a particular myocardial infarction patient population that can benefit from it.

SNPs have many important uses in drug discovery, screening, and development. A high probability exists that, for any gene/protein selected as a potential drug target, variants of that gene/protein will exist in a patient population. Thus, determining the impact of gene/protein variants on the selection and delivery of a therapeutic agent should be an integral aspect of the drug discovery and development process. (Jazwinska, *A Trends Guide to Genetic Variation and Genomic Medicine*, 2002 Mar; S30-S36).

Knowledge of variants (e.g., SNPs and any corresponding amino acid polymorphisms) of a particular therapeutic target (e.g., a gene, mRNA transcript, or protein) enables parallel screening of the variants in order to identify therapeutic candidates (e.g., small molecule compounds, antibodies, antisense or RNAi nucleic acid compounds, etc.) that demonstrate efficacy across variants (Rothberg, *Nat Biotechnol* 2001 Mar;19(3):209-11). Such therapeutic candidates would be expected to show equal efficacy across a larger segment of the patient population, thereby leading to a larger potential market for the therapeutic candidate.

Furthermore, identifying variants of a potential therapeutic target enables the most common form of the target to be used for selection of therapeutic candidates, thereby helping to ensure that the experimental activity that is observed for the

selected candidates reflects the real activity expected in the largest proportion of a patient population (Jazwinska, *A Trends Guide to Genetic Variation and Genomic Medicine*, 2002 Mar; S30-S36).

5 Additionally, screening therapeutic candidates against all known variants of a target can enable the early identification of potential toxicities and adverse reactions relating to particular variants. For example, variability in drug absorption, distribution, metabolism and excretion (ADME) caused by, for example, SNPs in therapeutic targets or drug metabolizing genes, can be identified, and this information can be utilized during the drug development process to minimize variability in drug  
10 disposition and develop therapeutic agents that are safer across a wider range of a patient population. The SNPs of the present invention, including the variant proteins and encoding polymorphic nucleic acid molecules provided in Tables 1-2, are useful in conjunction with a variety of toxicology methods established in the art, such as those set forth in *Current Protocols in Toxicology*, John Wiley & Sons, Inc., N.Y.

15 Furthermore, therapeutic agents that target any art-known proteins (or nucleic acid molecules, either RNA or DNA) may cross-react with the variant proteins (or polymorphic nucleic acid molecules) disclosed in Table 1, thereby significantly affecting the pharmacokinetic properties of the drug. Consequently, the protein variants and the SNP-containing  
20 nucleic acid molecules disclosed in Tables 1-2 are useful in developing, screening, and evaluating therapeutic agents that target corresponding art-known protein forms (or nucleic acid molecules). Additionally, as discussed above, knowledge of all polymorphic forms of a particular drug target enables the design of therapeutic agents that are effective against  
25 most or all such polymorphic forms of the drug target.

#### **Pharmaceutical Compositions and Administration Thereof**

Any of the myocardial infarction-associated proteins, and encoding nucleic acid molecules, disclosed herein can be used as therapeutic targets (or directly used  
30 themselves as therapeutic compounds) for treating myocardial infarction and related pathologies, and the present disclosure enables therapeutic compounds (e.g., small molecules, antibodies, therapeutic proteins, RNAi and antisense molecules, etc.) to be developed that target (or are comprised of) any of these therapeutic targets.

In general, a therapeutic compound will be administered in a therapeutically effective amount by any of the accepted modes of administration for agents that serve similar utilities. The actual amount of the therapeutic compound of this invention, i.e., the active ingredient, will depend upon numerous factors such as the severity of the disease to be treated, the age and relative health of the subject, the potency of the compound used, the route and form of administration, and other factors.

Therapeutically effective amounts of therapeutic compounds may range from, for example, approximately 0.01-50 mg per kilogram body weight of the recipient per day; preferably about 0.1-20 mg/kg/day. Thus, as an example, for administration to a 70 kg person, the dosage range would most preferably be about 7 mg to 1.4 g per day.

In general, therapeutic compounds will be administered as pharmaceutical compositions by any one of the following routes: oral, systemic (e.g., transdermal, intranasal, or by suppository), or parenteral (e.g., intramuscular, intravenous, or subcutaneous) administration. The preferred manner of administration is oral or parenteral using a convenient daily dosage regimen, which can be adjusted according to the degree of affliction. Oral compositions can take the form of tablets, pills, capsules, semisolids, powders, sustained release formulations, solutions, suspensions, elixirs, aerosols, or any other appropriate compositions.

The choice of formulation depends on various factors such as the mode of drug administration (e.g., for oral administration, formulations in the form of tablets; pills, or capsules are preferred) and the bioavailability of the drug substance. Recently, pharmaceutical formulations have been developed especially for drugs that show poor bioavailability based upon the principle that bioavailability can be increased by increasing the surface area, i.e., decreasing particle size. For example, U.S. Patent No. 4,107,288 describes a pharmaceutical formulation having particles in the size range from 10 to 1,000 nm in which the active material is supported on a cross-linked matrix of macromolecules. U.S. Patent No. 5,145,684 describes the production of a pharmaceutical formulation in which the drug substance is pulverized to nanoparticles (average particle size of 400 nm) in the presence of a surface modifier and then dispersed in a liquid medium to give a pharmaceutical formulation that exhibits remarkably high bioavailability.

Pharmaceutical compositions are comprised of, in general, a therapeutic compound in combination with at least one pharmaceutically acceptable excipient. Acceptable excipients are non-toxic, aid administration, and do not adversely affect

the therapeutic benefit of the therapeutic compound. Such excipients may be any solid, liquid, semi-solid or, in the case of an aerosol composition, gaseous excipient that is generally available to one skilled in the art.

5 Solid pharmaceutical excipients include starch, cellulose, talc, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, magnesium stearate, sodium stearate, glycerol monostearate, sodium chloride, dried skim milk and the like. Liquid and semisolid excipients may be selected from glycerol, propylene glycol, water, ethanol and various oils, including those of petroleum, animal, vegetable or synthetic origin, e.g., peanut oil, soybean oil, mineral oil, sesame oil, etc. Preferred  
10 liquid carriers, particularly for injectable solutions, include water, saline, aqueous dextrose, and glycols.

Compressed gases may be used to disperse a compound of this invention in aerosol form. Inert gases suitable for this purpose are nitrogen, carbon dioxide, etc.

Other suitable pharmaceutical excipients and their formulations are described  
15 in Remington's Pharmaceutical Sciences, edited by E. W. Martin (Mack Publishing Company, 18th ed., 1990).

The amount of the therapeutic compound in a formulation can vary within the full range employed by those skilled in the art. Typically, the formulation will contain, on a weight percent (wt %) basis, from about 0.01-99.99 wt % of the  
20 therapeutic compound based on the total formulation, with the balance being one or more suitable pharmaceutical excipients. Preferably, the compound is present at a level of about 1-80 wt %.

Therapeutic compounds can be administered alone or in combination with other therapeutic compounds or in combination with one or more other active  
25 ingredient(s). For example, an inhibitor or stimulator of a myocardial infarction-associated protein can be administered in combination with another agent that inhibits or stimulates the activity of the same or a different myocardial infarction-associated protein to thereby counteract the affects of myocardial infarction.

For further information regarding pharmacology, see *Current Protocols in*  
30 *Pharmacology*, John Wiley & Sons, Inc., N.Y.

### **Human Identification Applications**

In addition to their diagnostic and therapeutic uses in myocardial infarction and related pathologies, the SNPs provided by the present invention are also useful as

human identification markers for such applications as forensics, paternity testing, and biometrics (see, e.g., Gill, "An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes", *Int J Legal Med.* 2001;114(4-5):204-10). Genetic variations in the nucleic acid sequences between individuals can be used  
5 as genetic markers to identify individuals and to associate a biological sample with an individual. Determination of which nucleotides occupy a set of SNP positions in an individual identifies a set of SNP markers that distinguishes the individual. The more SNP positions that are analyzed, the lower the probability that the set of SNPs in one individual is the same as that in an unrelated individual. Preferably, if multiple sites  
10 are analyzed, the sites are unlinked (i.e., inherited independently). Thus, preferred sets of SNPs can be selected from among the SNPs disclosed herein, which may include SNPs on different chromosomes, SNPs on different chromosome arms, and/or SNPs that are dispersed over substantial distances along the same chromosome arm.

Furthermore, among the SNPs disclosed herein, preferred SNPs for use in  
15 certain forensic/human identification applications include SNPs located at degenerate codon positions (i.e., the third position in certain codons which can be one of two or more alternative nucleotides and still encode the same amino acid), since these SNPs do not affect the encoded protein. SNPs that do not affect the encoded protein are expected to be under less selective pressure and are therefore expected to be more  
20 polymorphic in a population, which is typically an advantage for forensic/human identification applications. However, for certain forensics/human identification applications, such as predicting phenotypic characteristics (e.g., inferring ancestry or inferring one or more physical characteristics of an individual) from a DNA sample, it may be desirable to utilize SNPs that affect the encoded protein.

25 For many of the SNPs disclosed in Tables 1-2 (which are identified as "Applera" SNP source), Tables 1-2 provide SNP allele frequencies obtained by re-sequencing the DNA of chromosomes from 39 individuals (Tables 1-2 also provide allele frequency information for "Celera" source SNPs and, where available, public SNPs from dbEST, HGBASE, and/or HGMD). The allele frequencies provided in  
30 Tables 1-2 enable these SNPs to be readily used for human identification applications. Although any SNP disclosed in Table 1 and/or Table 2 could be used for human identification, the closer that the frequency of the minor allele at a particular SNP site is to 50%, the greater the ability of that SNP to discriminate between different individuals in a population since it becomes increasingly likely that two randomly

selected individuals would have different alleles at that SNP site. Using the SNP allele frequencies provided in Tables 1-2, one of ordinary skill in the art could readily select a subset of SNPs for which the frequency of the minor allele is, for example, at least 1%, 2%, 5%, 10%, 20%, 25%, 30%, 40%, 45%, or 50%, or any other frequency in-between. Thus, since Tables 1-2 provide allele frequencies based on the re-sequencing of the chromosomes from 39 individuals, a subset of SNPs could readily be selected for human identification in which the total allele count of the minor allele at a particular SNP site is, for example, at least 1, 2, 4, 8, 10, 16, 20, 24, 30, 32, 36, 38, 39, 40, or any other number in-between.

Furthermore, Tables 1-2 also provide population group (interchangeably referred to herein as ethnic or racial groups) information coupled with the extensive allele frequency information. For example, the group of 39 individuals whose DNA was re-sequenced was made-up of 20 Caucasians and 19 African-Americans. This population group information enables further refinement of SNP selection for human identification. For example, preferred SNPs for human identification can be selected from Tables 1-2 that have similar allele frequencies in both the Caucasian and African-American populations; thus, for example, SNPs can be selected that have equally high discriminatory power in both populations. Alternatively, SNPs can be selected for which there is a statistically significant difference in allele frequencies between the Caucasian and African-American populations (as an extreme example, a particular allele may be observed only in either the Caucasian or the African-American population group but not observed in the other population group); such SNPs are useful, for example, for predicting the race/ethnicity of an unknown perpetrator from a biological sample such as a hair or blood stain recovered at a crime scene. For a discussion of using SNPs to predict ancestry from a DNA sample, including statistical methods, see Frudakis *et al.*, "A Classifier for the SNP-Based Inference of Ancestry", *Journal of Forensic Sciences* 2003; 48(4):771-782.

SNPs have numerous advantages over other types of polymorphic markers, such as short tandem repeats (STRs). For example, SNPs can be easily scored and are amenable to automation, making SNPs the markers of choice for large-scale forensic databases. SNPs are found in much greater abundance throughout the genome than repeat polymorphisms. Population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci. SNPs are mutationally more stable than repeat polymorphisms. SNPs

are not susceptible to artefacts such as stutter bands that can hinder analysis. Stutter bands are frequently encountered when analyzing repeat polymorphisms, and are particularly troublesome when analyzing samples such as crime scene samples that may contain mixtures of DNA from multiple sources. Another significant advantage  
5 of SNP markers over STR markers is the much shorter length of nucleic acid needed to score a SNP. For example, STR markers are generally several hundred base pairs in length. A SNP, on the other hand, comprises a single nucleotide, and generally a short conserved region on either side of the SNP position for primer and/or probe binding. This makes SNPs more amenable to typing in highly degraded or aged  
10 biological samples that are frequently encountered in forensic casework in which DNA may be fragmented into short pieces.

SNPs also are not subject to microvariant and "off-ladder" alleles frequently encountered when analyzing STR loci. Microvariants are deletions or insertions within a repeat unit that change the size of the amplified DNA product so that the  
15 amplified product does not migrate at the same rate as reference alleles with normal sized repeat units. When separated by size, such as by electrophoresis on a polyacrylamide gel, microvariants do not align with a reference allelic ladder of standard sized repeat units, but rather migrate between the reference alleles. The reference allelic ladder is used for precise sizing of alleles for allele classification;  
20 therefore alleles that do not align with the reference allelic ladder lead to substantial analysis problems. Furthermore, when analyzing multi-allelic repeat polymorphisms, occasionally an allele is found that consists of more or less repeat units than has been previously seen in the population, or more or less repeat alleles than are included in a reference allelic ladder. These alleles will migrate outside the size range of known  
25 alleles in a reference allelic ladder, and therefore are referred to as "off-ladder" alleles. In extreme cases, the allele may contain so few or so many repeats that it migrates well out of the range of the reference allelic ladder. In this situation, the allele may not even be observed, or, with multiplex analysis, it may migrate within or close to the size range for another locus, further confounding analysis.

30 SNP analysis avoids the problems of microvariants and off-ladder alleles encountered in STR analysis. Importantly, microvariants and off-ladder alleles may provide significant problems, and may be completely missed, when using analysis methods such as oligonucleotide hybridization arrays, which utilize oligonucleotide probes specific for certain known alleles. Furthermore, off-ladder alleles and

microvariants encountered with STR analysis, even when correctly typed, may lead to improper statistical analysis, since their frequencies in the population are generally unknown or poorly characterized, and therefore the statistical significance of a matching genotype may be questionable. All these advantages of SNP analysis are  
5 considerable in light of the consequences of most DNA identification cases, which may lead to life imprisonment for an individual, or re-association of remains to the family of a deceased individual.

DNA can be isolated from biological samples such as blood, bone, hair, saliva, or semen, and compared with the DNA from a reference source at particular SNP  
10 positions. Multiple SNP markers can be assayed simultaneously in order to increase the power of discrimination and the statistical significance of a matching genotype. For example, oligonucleotide arrays can be used to genotype a large number of SNPs simultaneously. The SNPs provided by the present invention can be assayed in  
15 combination with other polymorphic genetic markers, such as other SNPs known in the art or STRs, in order to identify an individual or to associate an individual with a particular biological sample.

Furthermore, the SNPs provided by the present invention can be genotyped for inclusion in a database of DNA genotypes, for example, a criminal DNA databank such as the FBI's Combined DNA Index System (CODIS) database. A genotype  
20 obtained from a biological sample of unknown source can then be queried against the database to find a matching genotype, with the SNPs of the present invention providing nucleotide positions at which to compare the known and unknown DNA sequences for identity. Accordingly, the present invention provides a database comprising novel SNPs or SNP alleles of the present invention (e.g., the database can  
25 comprise information indicating which alleles are possessed by individual members of a population at one or more novel SNP sites of the present invention), such as for use in forensics, biometrics, or other human identification applications. Such a database typically comprises a computer-based system in which the SNPs or SNP alleles of the present invention are recorded on a computer readable medium (see the section of the  
30 present specification entitled "Computer-Related Embodiments").

The SNPs of the present invention can also be assayed for use in paternity testing. The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing

investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child, with the SNPs of the present invention providing nucleotide positions at which to compare the putative  
5 father's and child's DNA sequences for identity. If the set of polymorphisms in the child attributable to the father does not match the set of polymorphisms of the putative father, it can be concluded, barring experimental error, that the putative father is not the father of the child. If the set of polymorphisms in the child attributable to the father match the set of polymorphisms of the putative father, a statistical calculation  
10 can be performed to determine the probability of coincidental match, and a conclusion drawn as to the likelihood that the putative father is the true biological father of the child.

In addition to paternity testing, SNPs are also useful for other types of kinship testing, such as for verifying familial relationships for immigration purposes, or for  
15 cases in which an individual alleges to be related to a deceased individual in order to claim an inheritance from the deceased individual, etc. For further information regarding the utility of SNPs for paternity testing and other types of kinship testing, including methods for statistical analysis, see Krawczak, "Informativity assessment for biallelic single nucleotide polymorphisms", *Electrophoresis* 1999 Jun;20(8):1676-  
20 81.

The use of the SNPs of the present invention for human identification further extends to various authentication systems, commonly referred to as biometric systems, which typically convert physical characteristics of humans (or other organisms) into digital data. Biometric systems include various technological devices that measure such  
25 unique anatomical or physiological characteristics as finger, thumb, or palm prints; hand geometry; vein patterning on the back of the hand; blood vessel patterning of the retina and color and texture of the iris; facial characteristics; voice patterns; signature and typing dynamics; and DNA. Such physiological measurements can be used to verify identity and, for example, restrict or allow access based on the identification. Examples  
30 of applications for biometrics include physical area security, computer and network security, aircraft passenger check-in and boarding, financial transactions, medical records access, government benefit distribution, voting, law enforcement, passports, visas and immigration, prisons, various military applications, and for restricting access to

expensive or dangerous items, such as automobiles or guns (see, for example, O'Connor, *Stanford Technology Law Review* and U.S. Patent No. 6,119,096).

Groups of SNPs, particularly the SNPs provided by the present invention, can be typed to uniquely identify an individual for biometric applications such as those described above. Such SNP typing can readily be accomplished using, for example, DNA chips/arrays. Preferably, a minimally invasive means for obtaining a DNA sample is utilized. For example, PCR amplification enables sufficient quantities of DNA for analysis to be obtained from buccal swabs or fingerprints, which contain DNA-containing skin cells and oils that are naturally transferred during contact.

Further information regarding techniques for using SNPs in forensic/human identification applications can be found in, for example, *Current Protocols in Human Genetics*, John Wiley & Sons, N.Y. (2002), 14.1-14.7.

## **VARIANT PROTEINS, ANTIBODIES, VECTORS & HOST CELLS, & USES THEREOF**

### **Variant Proteins Encoded by SNP-Containing Nucleic Acid Molecules**

The present invention provides SNP-containing nucleic acid molecules, many of which encode proteins having variant amino acid sequences as compared to the art-known (i.e., wild-type) proteins. Amino acid sequences encoded by the polymorphic nucleic acid molecules of the present invention are provided as SEQ ID NOS:27-46 in Table 1 and the Sequence Listing. These variants will generally be referred to herein as variant proteins/peptides/polypeptides, or polymorphic proteins/peptides/polypeptides of the present invention. The terms "protein", "peptide", and "polypeptide" are used herein interchangeably.

A variant protein of the present invention may be encoded by, for example, a nonsynonymous nucleotide substitution at any one of the cSNP positions disclosed herein. In addition, variant proteins may also include proteins whose expression, structure, and/or function is altered by a SNP disclosed herein, such as a SNP that creates or destroys a stop codon, a SNP that affects splicing, and a SNP in control/regulatory elements, e.g. promoters, enhancers, or transcription factor binding domains.

As used herein, a protein or peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or chemical precursors or other chemicals. The variant proteins of the present invention can be purified to homogeneity or other lower degrees of purity. The level of purification will be based on the intended use. The  
5 key feature is that the preparation allows for the desired function of the variant protein, even if in the presence of considerable amounts of other components.

As used herein, "substantially free of cellular material" includes preparations of the variant protein having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other  
10 proteins, or less than about 5% other proteins. When the variant protein is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the variant protein in which it is separated from chemical  
15 precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the variant protein having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than  
20 about 5% chemical precursors or other chemicals.

An isolated variant protein may be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant host cells), or synthesized using known protein synthesis methods. For example, a nucleic acid molecule containing SNP(s) encoding the variant protein can be cloned into an  
25 expression vector, the expression vector introduced into a host cell, and the variant protein expressed in the host cell. The variant protein can then be isolated from the cells by any appropriate purification scheme using standard protein purification techniques. Examples of these techniques are described in detail below (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press,  
30 Cold Spring Harbor, NY).

The present invention provides isolated variant proteins that comprise, consist of or consist essentially of amino acid sequences that

contain one or more variant amino acids encoded by one or more codons which contain a SNP of the present invention.

Accordingly, the present invention provides variant proteins that consist of amino acid sequences that contain one or more amino acid polymorphisms (or  
5 truncations or extensions due to creation or destruction of a stop codon, respectively) encoded by the SNPs provided in Table 1 and/or Table 2. A protein consists of an amino acid sequence when the amino acid sequence is the entire amino acid sequence of the protein.

The present invention further provides variant proteins that consist essentially of  
10 amino acid sequences that contain one or more amino acid polymorphisms (or truncations or extensions due to creation or destruction of a stop codon, respectively) encoded by the SNPs provided in Table 1 and/or Table 2. A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues in the final protein.

15 The present invention further provides variant proteins that comprise amino acid sequences that contain one or more amino acid polymorphisms (or truncations or extensions due to creation or destruction of a stop codon, respectively) encoded by the SNPs provided in Table 1 and/or Table 2. A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the  
20 protein. In such a fashion, the protein may contain only the variant amino acid sequence or have additional amino acid residues, such as a contiguous encoded sequence that is naturally associated with it or heterologous amino acid residues. Such a protein can have a few additional amino acid residues or can comprise many more additional amino acids. A brief description of how various types of these proteins can be made and isolated is  
25 provided below.

The variant proteins of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a variant protein operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the variant protein.  
30 "Operatively linked" indicates that the coding sequences for the variant protein and the heterologous protein are ligated in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the variant protein. In another embodiment, the fusion protein is encoded by a fusion polynucleotide that is synthesized by

conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see  
 5 Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A variant protein-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the variant protein.

10 In many uses, the fusion protein does not affect the activity of the variant protein. The fusion protein can include, but is not limited to, enzymatic fusion proteins, for example, beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate their purification following recombinant expression. In certain  
 15 host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Fusion proteins are further described in, for example, Terpe, "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems", *Appl Microbiol Biotechnol*: 2003 Jan;60(5):523-33. Epub 2002 Nov 07; Graddis et al., "Designing proteins that work  
 20 using recombinant technologies", *Curr Pharm Biotechnol*. 2002 Dec;3(4):285-97; and Nilsson et al., "Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins", *Protein Expr Purif*. 1997 Oct;11(1):1-16.

The present invention also relates to further obvious variants of the variant polypeptides of the present invention, such as naturally-occurring mature forms (e.g.,  
 25 allelic variants), non-naturally occurring recombinantly-derived variants, and orthologs and paralogs of such proteins that share sequence homology. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude those known in the prior art before the present invention.

30 Further variants of the variant polypeptides disclosed in Table 1 can comprise an amino acid sequence that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with an amino acid sequence disclosed in Table 1 (or a fragment thereof) and that includes a novel amino acid residue (allele) disclosed in Table 1 (which is encoded by a novel SNP allele).

Thus, the present invention specifically contemplates polypeptides that have a certain degree of sequence variation compared with the polypeptide sequences shown in Table 1, but that contain a novel amino acid residue (allele) encoded by a novel SNP allele disclosed herein. In other words, as long as a polypeptide contains a novel amino acid residue disclosed herein, other portions of the polypeptide that flank the novel amino acid residue can vary to some degree from the polypeptide sequences shown in Table 1.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the amino acid sequences disclosed herein can readily be identified as having complete sequence identity to one of the variant proteins of the present invention as well as being encoded by the same genetic locus as the variant proteins provided herein.

Orthologs of a variant peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of a variant peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from non-human mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs can be encoded by a nucleic acid sequence that hybridizes to a variant peptide-encoding nucleic acid molecule under moderate to stringent conditions depending on the degree of relatedness of the two organisms yielding the homologous proteins.

Variant proteins include, but are not limited to, proteins containing deletions, additions and substitutions in the amino acid sequence caused by the SNPs of the present invention. One class of substitutions is conserved amino acid substitutions in which a given amino acid in a polypeptide is substituted for another amino acid of like characteristics. Typical conservative substitutions are replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in, for example, Bowie *et al.*, *Science* 247:1306-1310 (1990).

Variant proteins can be fully functional or can lack function in one or more activities, e.g. ability to bind another molecule, ability to catalyze a substrate, ability

to mediate signaling, etc. Fully functional variants typically contain only conservative variations or variations in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions  
5 may positively or negatively affect function to some degree. Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, truncations or extensions, or a substitution, insertion, inversion, or deletion of a critical residue or in a critical region.

Amino acids that are essential for function of a protein can be identified by  
10 methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)), particularly using the amino acid sequence and polymorphism information provided in Table 1. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as enzyme activity  
15 or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

Polypeptides can contain amino acids other than the 20 amino acids  
20 commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Accordingly, the variant proteins of the present invention also encompass derivatives or analogs in which a  
25 substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (e.g., polyethylene glycol), or in which additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of  
30 the mature polypeptide or a pro-protein sequence.

Known protein modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of

phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such protein modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993); Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990); and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663:48-62 (1992).

The present invention further provides fragments of the variant proteins in which the fragments contain one or more amino acid sequence variations (e.g., substitutions, or truncations or extensions due to creation or destruction of a stop codon) encoded by one or more SNPs disclosed herein. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that have been disclosed in the prior art before the present invention.

As used herein, a fragment may comprise at least about 4, 8, 10, 12, 14, 16, 18, 20, 25, 30, 50, 100 (or any other number in-between) or more contiguous amino acid residues from a variant protein, wherein at least one amino acid residue is affected by a SNP of the present invention, e.g., a variant amino acid residue encoded by a nonsynonymous nucleotide substitution at a cSNP position provided by the present invention. The variant amino acid encoded by a cSNP may occupy any residue position along the sequence of the fragment. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the variant protein or the ability to perform a function, e.g., act as an immunogen. Particularly important fragments are biologically active fragments. Such fragments will typically comprise a domain or motif of a variant protein of the present invention, e.g., active site, transmembrane domain, or ligand/substrate binding domain. Other fragments include, but are not limited to, domain or motif-containing fragments, soluble peptide fragments, and fragments

containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known to those of skill in the art (e.g., PROSITE analysis) (*Current Protocols in Protein Science*, John Wiley & Sons, N.Y. (2002)).

## 5            Uses of Variant Proteins

The variant proteins of the present invention can be used in a variety of ways, including but not limited to, in assays to determine the biological activity of a variant protein, such as in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another type of immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively  
10    determine levels of the variant protein (or its binding partner) in biological fluids; as a marker for cells or tissues in which it is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); as a target for screening for a therapeutic agent; and as a direct therapeutic agent to be administered into a human subject. Any of the variant proteins disclosed herein may be developed into reagent grade or kit format for commercialization  
15    as research products. Methods for performing the uses listed above are well known to those skilled in the art (see, e.g., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Sambrook and Russell, 2000, and *Methods in Enzymology: Guide to Molecular Cloning Techniques*, Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987).

In a specific embodiment of the invention, the methods of the present invention include  
20    detection of one or more variant proteins disclosed herein. Variant proteins are disclosed in Table 1 and in the Sequence Listing as SEQ ID NOS: 27-46. Detection of such proteins can be accomplished using, for example, antibodies, small molecule compounds, aptamers, ligands/substrates, other proteins or protein fragments, or other protein-binding agents. Preferably, protein detection agents are specific for a variant protein of the present invention and can therefore  
25    discriminate between a variant protein of the present invention and the wild-type protein or another variant form. This can generally be accomplished by, for example, selecting or designing detection agents that bind to the region of a protein that differs between the

variant and wild-type protein, such as a region of a protein that contains one or more amino acid substitutions that is/are encoded by a non-synonymous cSNP of the present invention, or a region of a protein that follows a nonsense mutation-type SNP that creates a stop codon thereby leading to a shorter polypeptide, or a region of a protein that follows a read-through mutation-type SNP that destroys a stop codon thereby leading to a longer polypeptide in which a portion of the polypeptide is present in one version of the polypeptide but not the other.

In another specific aspect of the invention, the variant proteins of the present invention are used as targets for diagnosing myocardial infarction or for determining predisposition to myocardial infarction in a human. Accordingly, the invention provides methods for detecting the presence of, or levels of, one or more variant proteins of the present invention in a cell, tissue, or organism. Such methods typically involve contacting a test sample with an agent (e.g., an antibody, small molecule compound, or peptide) capable of interacting with the variant protein such that specific binding of the agent to the variant protein can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an array, for example, an antibody or aptamer array (arrays for protein detection may also be referred to as "protein chips"). The variant protein of interest can be isolated from a test sample and assayed for the presence of a variant amino acid sequence encoded by one or more SNPs disclosed by the present invention. The SNPs may cause changes to the protein and the corresponding protein function/activity, such as through non-synonymous substitutions in protein coding regions that can lead to amino acid substitutions, deletions, insertions, and/or rearrangements; formation or destruction of stop codons; or alteration of control elements such as promoters. SNPs may also cause inappropriate post-translational modifications.

One preferred agent for detecting a variant protein in a sample is an antibody capable of selectively binding to a variant form of the protein (antibodies are described in greater detail in the next section). Such samples include, for example, tissues, cells, and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

*In vitro* methods for detection of the variant proteins associated with myocardial infarction that are disclosed herein and fragments thereof include, but are not limited to, enzyme linked immunosorbent assays (ELISAs), radioimmunoassays (RIA), Western blots, immunoprecipitations, immunofluorescence, and protein arrays/chips (e.g., arrays

of antibodies or aptamers). For further information regarding immunoassays and related protein detection methods, see *Current Protocols in Immunology*, John Wiley & Sons, N.Y., and Hage, "Immunoassays", *Anal Chem.* 1999 Jun 15;71(12):294R-304R.

Additional analytic methods of detecting amino acid variants include, but are not  
5 limited to, altered electrophoretic mobility, altered tryptic peptide digest, altered protein activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, and direct amino acid sequencing.

Alternatively, variant proteins can be detected *in vivo* in a subject by introducing into the subject a labeled antibody (or other type of detection reagent) specific for a  
10 variant protein. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

Other uses of the variant peptides of the present invention are based on the class or action of the protein. For example, proteins isolated from humans and their  
15 mammalian orthologs serve as targets for identifying agents (e.g., small molecule drugs or antibodies) for use in therapeutic applications, particularly for modulating a biological or pathological response in a cell or tissue that expresses the protein. Pharmaceutical agents can be developed that modulate protein activity.

As an alternative to modulating gene expression, therapeutic compounds can be  
20 developed that modulate protein function. For example, many SNPs disclosed herein affect the amino acid sequence of the encoded protein (e.g., non-synonymous cSNPs and nonsense mutation-type SNPs). Such alterations in the encoded amino acid sequence may affect protein function, particularly if such amino acid sequence variations occur in functional protein domains, such as catalytic domains, ATP-binding domains, or  
25 ligand/substrate binding domains. It is well established in the art that variant proteins having amino acid sequence variations in functional domains can cause or influence pathological conditions. In such instances, compounds (e.g., small molecule drugs or antibodies) can be developed that target the variant protein and modulate (e.g., up- or down-regulate) protein function/activity.

30 The therapeutic methods of the present invention further include methods that target one or more variant proteins of the present invention. Variant proteins can be targeted using, for example, small molecule compounds, antibodies, aptamers, ligands/substrates, other proteins, or

other protein-binding agents. Additionally, the skilled artisan will recognize that the novel protein variants (and polymorphic nucleic acid molecules) disclosed in Table 1 may themselves be directly used as therapeutic agents by acting as competitive inhibitors of corresponding  
5 art-known proteins (or nucleic acid molecules such as mRNA molecules).

The variant proteins of the present invention are particularly useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can utilize cells that naturally express the protein, a biopsy specimen, or cell cultures. In one embodiment, cell-based assays involve recombinant host cells expressing the variant  
10 protein. Cell-free assays can be used to detect the ability of a compound to directly bind to a variant protein or to the corresponding SNP-containing nucleic acid fragment that encodes the variant protein.

A variant protein of the present invention, as well as appropriate fragments thereof, can be used in high-throughput screening assays to test candidate compounds for  
15 the ability to bind and/or modulate the activity of the variant protein. These candidate compounds can be further screened against a protein having normal function (e.g., a wild-type/non-variant protein) to further determine the effect of the compound on the protein activity. Furthermore, these compounds can be tested in animal or invertebrate systems to determine *in vivo* activity/effectiveness. Compounds can be identified that  
20 activate (agonists) or inactivate (antagonists) the variant protein, and different compounds can be identified that cause various degrees of activation or inactivation of the variant protein.

Further, the variant proteins can be used to screen a compound for the ability to stimulate or inhibit interaction between the variant protein and a target molecule that  
25 normally interacts with the protein. The target can be a ligand, a substrate or a binding partner that the protein normally interacts with (for example, epinephrine or norepinephrine). Such assays typically include the steps of combining the variant protein with a candidate compound under conditions that allow the variant protein, or fragment thereof, to interact with the target molecule, and to detect the formation of a  
30 complex between the protein and the target or to detect the biochemical consequence of the interaction with the variant protein and the target, such as any of the associated effects of signal transduction.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')<sub>2</sub>, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the variant protein that competes for ligand binding. Other candidate compounds include mutant proteins or appropriate fragments containing mutations that affect variant protein function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) variant protein activity. The assays typically involve an assay of events in the signal transduction pathway that indicate protein activity. Thus, the expression of genes that are up or down-regulated in response to the variant protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the variant protein, or a variant protein target, could also be measured. Any of the biological or biochemical functions mediated by the variant protein can be used as an endpoint assay. These include all of the biochemical or biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art.

Binding and/or activating compounds can also be screened by using chimeric variant proteins in which an amino terminal extracellular domain or parts thereof, an entire transmembrane domain or subregions, and/or the carboxyl terminal intracellular domain or parts thereof, can be replaced by heterologous domains or subregions. For example, a substrate-binding region can be used that interacts with a different substrate than that which is normally recognized by a variant protein. Accordingly, a different set

of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the variant protein is derived.

5 The variant proteins are also useful in competition binding assays in methods designed to discover compounds that interact with the variant protein. Thus, a compound can be exposed to a variant protein under conditions that allow the compound to bind or to otherwise interact with the variant protein. A binding partner, such as ligand, that normally interacts with the variant protein is also added to the mixture. If the test compound interacts with the variant protein or its binding partner, it decreases the  
10 amount of complex formed or activity from the variant protein. This type of assay is particularly useful in screening for compounds that interact with specific regions of the variant protein (Hodgson, *Bio/technology*, 1992, Sept 10(9), 973-80).

To perform cell-free drug screening assays, it is sometimes desirable to immobilize either the variant protein or a fragment thereof, or its target molecule, to  
15 facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Any method for immobilizing proteins on matrices can be used in drug screening assays. In one embodiment, a fusion protein containing an added domain allows the protein to be bound to a matrix. For example, glutathione-S-transferase/<sup>125</sup>I fusion proteins can be  
20 adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., <sup>35</sup>S-labeled) and a candidate compound, such as a drug candidate, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads can be washed to remove  
25 any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of bound material found in the bead fraction quantitated from the gel using standard electrophoretic techniques.

30 Either the variant protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Alternatively, antibodies reactive with the variant protein but which do not interfere with binding of the variant protein to its target molecule can be derivatized to the wells of the plate, and the variant protein trapped in

the wells by antibody conjugation. Preparations of the target molecule and a candidate compound are incubated in the variant protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include

5 immunodetection of complexes using antibodies reactive with the protein target molecule, or which are reactive with variant protein and compete with the target molecule, and enzyme-linked assays that rely on detecting an enzymatic activity associated with the target molecule.

Modulators of variant protein activity identified according to these

10 drug screening assays can be used to treat a subject with a disorder mediated by the protein pathway, such as myocardial infarction. These methods of treatment typically include the steps of administering the modulators of protein activity in a pharmaceutical composition to a subject in need of such treatment.

15 The variant proteins, or fragments thereof, disclosed herein can themselves be directly used to treat a disorder characterized by an absence of, inappropriate, or unwanted expression or activity of the variant protein. Accordingly, methods for treatment include the use of a variant protein disclosed herein or fragments thereof.

20 In yet another aspect of the invention, variant proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696;

25 and Brent WO94/10300) to identify other proteins that bind to or interact with the variant protein and are involved in variant protein activity. Such variant protein-binding proteins are also likely to be involved in the propagation of signals by the variant proteins or variant protein targets as, for example, elements of a protein-mediated signaling pathway.

30 Alternatively, such variant protein-binding proteins are inhibitors of the variant protein.

The two-hybrid system is based on the modular nature of most transcription factors, which typically consist of separable DNA-binding

and activation domains. Briefly, the assay typically utilizes two different DNA constructs. In one construct, the gene that codes for a variant protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence,  
5 from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a variant protein-dependent complex, the DNA-binding and activation domains of the transcription factor are  
10 brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected, and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that  
15 encodes the protein that interacts with the variant protein.

#### Antibodies Directed to Variant Proteins

The present invention also provides antibodies that selectively bind to the variant proteins disclosed herein and fragments thereof. Such antibodies may be used to  
20 quantitatively or qualitatively detect the variant proteins of the present invention. As used herein, an antibody selectively binds a target variant protein when it binds the variant protein and does not significantly bind to non-variant proteins, i.e., the antibody does not significantly bind to normal, wild-type, or art-known proteins that do not contain a variant amino acid sequence due to one or more SNPs of the present invention  
25 (variant amino acid sequences may be due to, for example, nonsynonymous cSNPs, nonsense SNPs that create a stop codon, thereby causing a truncation of a polypeptide or SNPs that cause read-through mutations resulting in an extension of a polypeptide).

As used herein, an antibody is defined in terms consistent with that recognized in the art: they are multi-subunit proteins produced by an organism in response to an  
30 antigen challenge. The antibodies of the present invention include both monoclonal antibodies and polyclonal antibodies, as well as antigen-reactive proteolytic fragments of such antibodies, such as Fab, F(ab)'<sub>2</sub>, and Fv fragments. In addition, an antibody of the present invention further includes any of a variety of engineered antigen-binding

molecules such as a chimeric antibody (U.S. Patent Nos. 4,816,567 and 4,816,397; Morrison *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:6851, 1984; Neuberger *et al.*, *Nature* 312:604, 1984), a humanized antibody (U.S. Patent Nos. 5,693,762; 5,585,089; and 5,565,332), a single-chain Fv (U.S. Patent No. 4,946,778; Ward *et al.*, *Nature* 334:544, 5 1989), a bispecific antibody with two binding specificities (Segal *et al.*, *J. Immunol. Methods* 248:1, 2001; Carter, *J. Immunol. Methods* 248:7, 2001), a diabody, a triabody, and a tetrabody (Todorovska *et al.*, *J. Immunol. Methods*, 248:47, 2001), as well as a Fab conjugate (dimer or trimer), and a minibody.

Many methods are known in the art for generating and/or identifying antibodies 10 to a given target antigen (Harlow, *Antibodies*, Cold Spring Harbor Press, (1989)). In general, an isolated peptide (e.g., a variant protein of the present invention) is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit, hamster or mouse. Either a full-length protein, an antigenic peptide fragment (e.g., a peptide 15 fragment containing a region that varies between a variant protein and a corresponding wild-type protein), or a fusion protein can be used. A protein used as an immunogen may be naturally-occurring, synthetic or recombinantly produced, and may be administered in combination with an adjuvant, including but not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substance such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet 20 hemocyanin, dinitrophenol, and the like.

Monoclonal antibodies can be produced by hybridoma technology (Kohler and Milstein, *Nature*, 256:495, 1975), which immortalizes cells secreting a specific monoclonal antibody. The immortalized cell lines can be created *in vitro* by fusing 25 two different cell types, typically lymphocytes, and tumor cells. The hybridoma cells may be cultivated *in vitro* or *in vivo*. Additionally, fully human antibodies can be generated by transgenic animals (He *et al.*, *J. Immunol.*, 169:595, 2002). Fd phage and Fd phagemid technologies may be used to generate and select recombinant antibodies *in vitro* (Hoogenboom and Chames, *Immunol. Today* 21:371, 2000; Liu *et al.*, *J. Mol. Biol.* 315:1063, 2002). The complementarity-determining regions of an antibody can 30 be identified, and synthetic peptides corresponding to such regions may be used to mediate antigen binding (U.S. Patent No. 5,637,677).

Antibodies are preferably prepared against regions or discrete fragments of a variant protein containing a variant amino acid sequence as compared to the corresponding wild-type protein (e.g., a region of a variant protein that includes an

amino acid encoded by a nonsynonymous cSNP, a region affected by truncation caused by a nonsense SNP that creates a stop codon, or a region resulting from the destruction of a stop codon due to read-through mutation caused by a SNP).

Furthermore, preferred regions will include those involved in function/activity and/or  
5 protein/binding partner interaction. Such fragments can be selected on a physical property, such as fragments corresponding to regions that are located on the surface of the protein, e.g., hydrophilic regions, or can be selected based on sequence uniqueness, or based on the position of the variant amino acid residue(s) encoded by the SNPs provided by the present invention. An antigenic fragment will typically comprise at least  
10 about 8-10 contiguous amino acid residues in which at least one of the amino acid residues is an amino acid affected by a SNP disclosed herein. The antigenic peptide can comprise, however, at least 12, 14, 16, 20, 25, 50, 100 (or any other number in-between) or more amino acid residues, provided that at least one amino acid is affected by a SNP disclosed herein.

15 Detection of an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody or an antigen-reactive fragment thereof to a detectable substance. Detectable substances include, but are not limited to, various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish  
20 peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol;  
25 examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

Antibodies, particularly the use of antibodies as therapeutic agents, are reviewed in: Morgan, "Antibody therapy for Alzheimer's disease", *Expert Rev Vaccines*. 2003 Feb;2(1):53-9; Ross et al., "Anticancer antibodies", *Am J Clin Pathol*. 2003  
30 Apr;119(4):472-85; Goldenberg, "Advancing role of radiolabeled antibodies in the therapy of cancer", *Cancer Immunol Immunother*. 2003 May;52(5):281-96. Epub 2003 Mar 11; Ross et al., "Antibody-based therapeutics in oncology", *Expert Rev Anticancer Ther*. 2003 Feb;3(1):107-21; Cao et al., "Bispecific antibody conjugates in therapeutics",

- 5 *Adv Drug Deliv Rev.* 2003 Feb 10;55(2):171-97; von Mehren et al., "Monoclonal antibody therapy for cancer", *Annu Rev Med.* 2003;54:343-69. Epub 2001 Dec 03; Hudson et al., "Engineered antibodies", *Nat Med.* 2003 Jan;9(1):129-34; Brekke et al., "Therapeutic antibodies for human diseases at the dawn of the twenty-first century", *Nat*  
 10 *Rev Drug Discov.* 2003 Jan;2(1):52-62 (Erratum in: *Nat Rev Drug Discov.* 2003 Mar;2(3):240); Houdebine, "Antibody manufacture in transgenic animals and comparisons with other systems", *Curr Opin Biotechnol.* 2002 Dec;13(6):625-9; Andreakos et al., "Monoclonal antibodies in immune and inflammatory diseases", *Curr Opin Biotechnol.* 2002 Dec;13(6):615-20; Kellermann et al., "Antibody discovery: the  
 15 use of transgenic mice to generate human monoclonal antibodies for therapeutics", *Curr Opin Biotechnol.* 2002 Dec;13(6):593-7; Pini et al., "Phage display and colony filter screening for high-throughput selection of antibody libraries", *Comb Chem High Throughput Screen.* 2002 Nov;5(7):503-10; Batra et al., "Pharmacokinetics and biodistribution of genetically engineered antibodies", *Curr Opin Biotechnol.* 2002  
 20 Dec;13(6):603-8; and Tangri et al., "Rationally engineered proteins or antibodies with absent or reduced immunogenicity", *Curr Med Chem.* 2002 Dec;9(24):2191-9.

### Uses of Antibodies

20 Antibodies can be used to isolate the variant proteins of the present invention from a natural cell source or from recombinant host cells by standard techniques, such as affinity chromatography or immunoprecipitation. In addition, antibodies are useful for detecting the presence of a variant protein of the present invention in cells or tissues to determine the pattern of expression of the variant protein among various tissues in an organism and over the course of normal development or disease progression. Further,  
 25 antibodies can be used to detect variant protein *in situ*, *in vitro*, in a bodily fluid, or in a cell lysate or supernatant in order to evaluate the amount and pattern of expression. Also, antibodies can be used to assess abnormal tissue distribution, abnormal expression during development, or expression in an abnormal condition, such as myocardial infarction. Additionally, antibody detection of circulating fragments of the full-length  
 30 variant protein can be used to identify turnover.

Antibodies to the variant proteins of the present invention are also useful in pharmacogenomic analysis. Thus, antibodies against variant proteins encoded by alternative SNP alleles can be used to identify individuals that require modified treatment modalities.

Further, antibodies can be used to assess expression of the variant protein in disease states such as in active stages of the disease or in an individual with a predisposition to a disease related to the protein's function, particularly myocardial infarction. Antibodies specific for a variant protein encoded by a SNP-containing nucleic acid molecule of the present invention can be used to assay for the presence of the variant protein, such as to screen for predisposition to myocardial infarction as indicated by the presence of the variant protein.

Antibodies are also useful as diagnostic tools for evaluating the variant proteins in conjunction with analysis by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays well known in the art.

Antibodies are also useful for tissue typing. Thus, where a specific variant protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

Antibodies can also be used to assess aberrant subcellular localization of a variant protein in cells in various tissues. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting the expression level or the presence of variant protein or aberrant tissue distribution or developmental expression of a variant protein, antibodies directed against the variant protein or relevant fragments can be used to monitor therapeutic efficacy.

The antibodies are also useful for inhibiting variant protein function, for example, by blocking the binding of a variant protein to a binding partner. These uses can also be applied in a therapeutic context in which treatment involves inhibiting a variant protein's function. An antibody can be used, for example, to block or competitively inhibit binding, thus modulating (agonizing or antagonizing) the activity of a variant protein. Antibodies can be prepared against specific variant protein fragments containing sites required for function or against an intact variant protein that is associated with a cell or cell membrane. For *in vivo* administration, an antibody may be linked with an additional therapeutic payload such as a radionuclide, an enzyme, an immunogenic epitope, or a cytotoxic agent. Suitable cytotoxic agents include, but are not limited to, bacterial toxin such as diphtheria, and plant toxin such as ricin. The *in vivo* half-life of an antibody or a fragment thereof may be lengthened by pegylation through conjugation to polyethylene glycol (Leong *et al.*, *Cytokine* 16:106, 2001).

The invention also encompasses kits for using antibodies, such as kits for detecting the presence of a variant protein in a test sample. An exemplary kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting variant proteins in a biological sample; means for determining the amount, or  
5 presence/absence of variant protein in the sample; means for comparing the amount of variant protein in the sample with a standard; and instructions for use.

### Vectors and Host Cells

The present invention also provides vectors containing the SNP-containing  
10 nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport a SNP-containing nucleic acid molecule. When the vector is a nucleic acid molecule, the SNP-containing nucleic acid molecule can be covalently linked to the vector nucleic acid. Such vectors include, but are not limited to, a plasmid, single or double stranded phage, a single or double stranded  
15 RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, or MAC.

A vector can be maintained in a host cell as an extrachromosomal element where it replicates and produces additional copies of the SNP-containing nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and  
20 produce additional copies of the SNP-containing nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the SNP-containing nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

25 Expression vectors typically contain cis-acting regulatory regions that are operably linked in the vector to the SNP-containing nucleic acid molecules such that transcription of the SNP-containing nucleic acid molecules is allowed in a host cell. The SNP-containing nucleic acid molecules can also be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second  
30 nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the SNP-containing nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is

understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequences to which the SNP-containing nucleic acid molecules described herein can be operably linked include promoters for directing mRNA  
5 transcription. These include, but are not limited to, the left promoter from bacteriophage  $\lambda$ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may  
10 also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression  
vectors can also contain sequences necessary for transcription termination and, in the  
15 transcribed region, a ribosome-binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. A person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors (see, e.g., Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor  
20 Laboratory Press, Cold Spring Harbor, NY).

A variety of expression vectors can be used to express a SNP-containing nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example, vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes,  
25 from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors can also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g., cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook and  
30 Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The regulatory sequence in a vector may provide constitutive expression in one or more host cells (e.g., tissue specific expression) or may provide for inducible

expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor, e.g., a hormone or other ligand. A variety of vectors that provide constitutive or inducible expression of a nucleic acid sequence in prokaryotic and eukaryotic host cells are well known to those of ordinary skill in the art.

5 A SNP-containing nucleic acid molecule can be inserted into the vector by methodology well-known in the art. Generally, the SNP-containing nucleic acid molecule that will ultimately be expressed is joined to an expression vector by cleaving the SNP-containing nucleic acid molecule and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction  
10 enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial host cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic host cells include, but are not limited to, yeast, insect cells  
15 such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the variant peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the variant peptides. Fusion vectors can, for example, increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the  
20 purification of the protein by acting, for example, as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired variant peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes suitable for such use include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith et al., *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E  
25 binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., *Gene* 69:301-315 (1988)) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).  
30

Recombinant protein expression can be maximized in a bacterial host by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California

(1990) 119-128). Alternatively, the sequence of the SNP-containing nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example, *E. coli* (Wada et al., *Nucleic Acids Res.* 20:2111-2118 (1992)).

5 The SNP-containing nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast (e.g., *S. cerevisiae*) include pYepSec1 (Baldari, et al., *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan et al., *Cell* 30:933-943(1982)), pJRY88 (Schultz et al., *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

10 The SNP-containing nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow et al., *Virology* 170:31-39 (1989)).

15 In certain embodiments of the invention, the SNP-containing nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, *B. Nature* 329:840(1987)) and pMT2PC (Kaufman et al., *EMBO J.* 6:187-195 (1987)).

20 The invention also encompasses vectors in which the SNP-containing nucleic acid molecules described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to the SNP-containing nucleic acid sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, 25 tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include, for example, prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

30 The recombinant host cells can be prepared by introducing the vector constructs described herein into the cells by techniques readily available to persons of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those described in

Sambrook and Russell, 2000, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Host cells can contain more than one vector. Thus, different SNP-containing nucleotide sequences can be introduced in different vectors into the same cell. Similarly, the SNP-containing nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the SNP-containing nucleic acid molecules, such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced, or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication can occur in host cells that provide functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be inserted in the same vector that contains the SNP-containing nucleic acid molecules described herein or may be in a separate vector. Markers include, for example, tetracycline or ampicillin-resistance genes for prokaryotic host cells, and dihydrofolate reductase or neomycin resistance genes for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait can be effective.

While the mature variant proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these variant proteins using RNA derived from the DNA constructs described herein.

Where secretion of the variant protein is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as G-protein-coupled receptors (GPCRs), appropriate secretion signals can be incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the variant protein is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze/thaw, sonication, mechanical disruption, use of lysing agents, and the like. The variant protein can then be recovered and purified by well-known purification methods including, for example, ammonium sulfate precipitation, acid extraction, anion or cationic exchange

chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that, depending upon the host cell in which  
5 recombinant production of the variant proteins described herein occurs, they can have various glycosylation patterns, or may be non-glycosylated, as when produced in bacteria. In addition, the variant proteins may include an initial modified methionine in some cases as a result of a host-mediated process.

10 For further information regarding vectors and host cells, see *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y.

#### Uses of Vectors and Host Cells, and Transgenic Animals

Recombinant host cells that express the variant proteins described herein have a  
15 variety of uses. For example, the cells are useful for producing a variant protein that can be further purified into a preparation of desired amounts of the variant protein or fragments thereof. Thus, host cells containing expression vectors are useful for variant protein production.

Host cells are also useful for conducting cell-based assays involving  
20 the variant protein or variant protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a variant protein is useful for assaying compounds that stimulate or inhibit variant protein function. Such an ability of a compound to modulate variant protein function may not be apparent from  
25 assays of the compound on the native/wild-type protein, or from cell-free assays of the compound. Recombinant host cells are also useful for assaying functional alterations in the variant proteins as compared with a known function.

Genetically-engineered host cells can be further used to produce non-human  
30 transgenic animals. A transgenic animal is preferably a non-human mammal, for example, a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA containing a SNP of the present invention which is integrated into the genome of a cell from which a transgenic animal

develops and which remains in the genome of the mature animal in one or more of its cell types or tissues. Such animals are useful for studying the function of a variant protein *in vivo*, and identifying and evaluating modulators of variant protein activity.

Other examples of transgenic animals include, but are not limited to, non-human  
5 primates, sheep, dogs, cows, goats, chickens, and amphibians. Transgenic non-human mammals such as cows and goats can be used to produce variant proteins which can be secreted in the animal's milk and then recovered.

A transgenic animal can be produced by introducing a SNP-containing nucleic acid molecule into the male pronuclei of a fertilized oocyte, e.g., by microinjection or  
10 retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any nucleic acid molecules that contain one or more SNPs of the present invention can potentially be introduced as a transgene into the genome of a non-human animal.

Any of the regulatory or other sequences useful in expression vectors can form  
15 part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the variant protein in particular cells or tissues.

Methods for generating transgenic animals via embryo manipulation and  
20 microinjection, particularly animals such as mice, have become conventional in the art and are described in, for example, U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Patent No. 4,873,191 by Wagner et al., and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A  
25 transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes a  
30 non-human animal in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1

(Lakso et al. *PNAS* 89:6232-6236 (1992)). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman et al. *Science* 251:1351-1355 (1991)). If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a  
5 selected protein are generally needed. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected variant protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be  
10 produced according to the methods described in, for example, Wilmut, I. et al. *Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell (e.g., a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be  
15 fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell (e.g., a somatic cell) is isolated.

Transgenic animals containing recombinant cells that express the variant proteins  
20 described herein are useful for conducting the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could influence ligand or substrate binding, variant protein activation, signal transduction, or other processes or interactions, may not be evident from *in vitro* cell-free or cell-based assays. Thus, non-human transgenic animals of the present invention may  
25 be used to assay *in vivo* variant protein function as well as the activities of a therapeutic agent or compound that modulates variant protein function/activity or expression. Such animals are also suitable for assessing the effects of null mutations (i.e., mutations that substantially or completely eliminate one or more variant protein functions).

For further information regarding transgenic animals, see Houdebine, "Antibody  
30 manufacture in transgenic animals and comparisons with other systems", *Curr Opin Biotechnol.* 2002 Dec;13(6):625-9; Petters et al., "Transgenic animals as models for human disease", *Transgenic Res.* 2000;9(4-5):347-51; discussion 345-6; Wolf et al., "Use of transgenic animals in understanding molecular mechanisms of toxicity", *J Pharm Pharmacol.* 1998 Jun;50(6):567-74; Echelard, "Recombinant protein production

in transgenic animals”, *Curr Opin Biotechnol.* 1996 Oct;7(5):536-40; Houdebine, “Transgenic animal bioreactors”, *Transgenic Res.* 2000;9(4-5):305-20; Purity et al., “Embryonic stem cells, creating transgenic animals”, *Methods Cell Biol.* 1998;57:279-93; and Robl et al., “Artificial chromosome vectors and expression of complex proteins in transgenic animals”, *Theriogenology.* 2003 Jan 1;59(1):107-13.

### **COMPUTER-RELATED EMBODIMENTS**

The SNPs provided in the present invention may be "provided" in a variety of mediums to facilitate use thereof. As used in this section, “provided” refers to a manufacture, other than an isolated nucleic acid molecule, that contains SNP information of the present invention. Such a manufacture provides the SNP information in a form that allows a skilled artisan to examine the manufacture using means not directly applicable to examining the SNPs or a subset thereof as they exist in nature or in purified form. The SNP information that may be provided in such a form includes any of the SNP information provided by the present invention such as, for example, polymorphic nucleic acid and/or amino acid sequence information such as SEQ ID NOS:1-26, SEQ ID NOS:27-46, SEQ ID NOS:67-73, SEQ ID NOS:47-66, and SEQ ID NOS:74-80; information about observed SNP alleles, alternative codons, populations, allele frequencies, SNP types, and/or affected proteins; or any other information provided by the present invention in Tables 1-2 and/or the Sequence Listing.

In one application of this embodiment, the SNPs of the present invention can be recorded on a computer readable medium. As used herein, "computer readable medium" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable media can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. One such medium is provided with the present application, namely, the present application contains computer readable medium (CD-R) that has nucleic acid sequences (and encoded protein sequences) containing SNPs provided/recorded thereon in ASCII text format in a Sequence

Listing along with accompanying Tables that contain detailed SNP and sequence information (transcript sequences are provided as SEQ ID NOS:1-26, protein sequences are provided as SEQ ID NOS:27-46, genomic sequences are provided as SEQ ID NOS:67-73, transcript-based context sequences are provided as SEQ ID NOS:47-66, and genomic-based context sequences are provided as SEQ ID NOS:74-80).

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the SNP information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide/amino acid sequence information of the present invention on computer readable medium. For example, the sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, represented in the form of an ASCII file, or stored in a database application, such as OB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the SNP information of the present invention.

By providing the SNPs of the present invention in computer readable form, a skilled artisan can routinely access the SNP information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Examples of publicly available computer software include BLAST (Altschul et al, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag et al, *Comp. Chem.* 17:203-207 (1993)) search algorithms.

The present invention further provides systems, particularly computer-based systems, which contain the SNP information described herein. Such systems may be designed to store and/or analyze information on, for example, a large number of SNP positions, or information on SNP genotypes from a large number of individuals. The

SNP information of the present invention represents a valuable information source. The SNP information of the present invention stored/analyzed in a computer-based system may be used for such computer-intensive applications as determining or analyzing SNP allele frequencies in a population, mapping disease genes, genotype-phenotype association studies, grouping SNPs into haplotypes, correlating SNP haplotypes with response to particular drugs, or for various other bioinformatic, pharmacogenomic, drug development, or human identification/forensic applications.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the SNP information of the present invention. The minimum hardware means of the computer-based systems of the present invention typically comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. Such a system can be changed into a system of the present invention by utilizing the SNP information provided on the CD-R, or a subset thereof, without any experimentation.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein SNPs of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store SNP information of the present invention, or a memory access means which can access manufactures having recorded thereon the SNP information of the present invention.

As used herein, "search means" refers to one or more programs or algorithms that are implemented on the computer-based system to identify or analyze SNPs in a target sequence based on the SNP information stored within the data storage means. Search means can be used to determine which nucleotide is present at a particular SNP position in the target sequence. As used herein, a "target sequence" can be any DNA sequence containing the SNP position(s) to be searched or queried.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences containing a SNP position in which the sequence(s) is chosen based on a three-dimensional configuration that is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic

active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures, and inducible expression elements (protein binding sequences).

5 A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. An exemplary format for an output means is a display that depicts the presence or absence of specified nucleotides (alleles) at particular SNP positions of interest. Such presentation can provide a rapid, binary scoring system for many SNPs simultaneously.

10 One exemplary embodiment of a computer-based system comprising SNP information of the present invention is provided in Figure 1. Figure 1 provides a block diagram of a computer system 102 that can be used to implement the present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented  
15 as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, etc. A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, etc.) containing control logic and/or data  
20 recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable storage medium 116 once inserted in the removable medium storage device 114.

The SNP information of the present invention may be stored in a well-known  
25 manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. Software for accessing and processing the SNP information (such as SNP scoring tools, search tools, comparing tools, etc.) preferably resides in main memory 108 during execution.

30 **EXAMPLES: STATISTICAL ANALYSIS OF SNP ASSOCIATION  
WITH MYOCARDIAL INFARCTION AND RECURRENT  
MYOCARDIAL INFARCTION**

Myocardial Infarction studies (see Table 6)

A case-control genetic study to determine the association of SNPs in the human genome with MI was carried out using genomic DNA extracted from 3 independently collected case-control sample sets.

Study S0012 had 1400 samples, in which patients (cases) had self-reported history of MI, and controls had no history of MI or of acute angina lasting more than 1 hour. Study S0028 had 1500 samples, in which patients (cases) had clinical evidence of history of MI, and controls had no history of MI. Study V0001 had 1288 samples, in which patients (cases) had clinical evidence of history of MI, and controls had no history of MI. All individuals who were included in each study had signed a written informed consent form. The study protocol was IRB approved.

DNA was extracted from blood samples using conventional DNA extraction methods such as the QIA-amp kit from Qiagen. SNP markers in the extracted DNA were analyzed by genotyping. While some samples were individually genotyped, the same samples were also used for pooling studies, in which DNA from about 50 individuals was pooled, and allele frequencies were determined in pooled DNA. For studies S0012(M) and S0012(F), only male or female cases and controls were used in pooling studies. Genotypes and pool allele frequencies were obtained using a PRISM 7900HT sequence detection PCR system (Applied Biosystems, Foster City, CA) by allele-specific PCR, similar to the method described by Germer et al (Germer S., Holland M.J., Higuchi R. 2000, *Genome Res.* 10: 258-266). Primers for the allele-specific PCR reactions are provided in Table 5.

Summary statistics for demographic and environmental traits, history of vascular disease, and allele frequencies for the tested SNPs were obtained and compared between cases and controls. No multiple testing corrections were made.

Tests of association were calculated for both non-stratified and stratified settings: 1) Fisher's exact test of allelic association, and 2) asymptotic chi-square test of genotypic association, taking two different modes of inheritance into account (dominant, and recessive).

Effect sizes were estimated through allelic odds ratios, including 95% confidence intervals. The reported Allele1 may be under-represented in cases (with a lower allele frequency in cases than in controls, indicating that the reported Allele1 is associated with decreased risk and the other allele is a risk factor for disease) or over-represented in cases (indicating that the reported Allele1 is a risk factor in the development of disease).

A SNP was considered to be a significant genetic marker if it exhibited a p-value  $< 0.05$  in the allelic association test or in any of the 2 genotypic tests (dominant, recessive). The association of a marker with MI was considered replicated if the marker exhibited an allelic or genotypic association test p-value  $< 0.05$  in one of the sample sets and the same test and strata (or substrata) were significant ( $p < 0.05$ ) in another independent sample set.

An example of a replicated marker, where the reported Allele1 is associated with decreased risk for MI, is hCV8851074 (Table 6). hCV8851074 shows significant association with all individuals (strata = "ALL") of study S0028 and the non-smoking strata (Stratification = "SMOKE", Strata = "N") of study S0012. The odds ratio in both studies is less than 1 (0.84 and 0.78 respectively), using the same allele (Allele1 = "A") for analysis.

An example of a replicated marker, where the reported Allele1 is associated with increased risk for MI, is hCV2716008 (Table 6). hCV2716008 shows significant association with all individuals (strata = "ALL") of study S0028 and all males of study S0012 (strata = "ALL", study="S0012(M)"). The odds ratio in both studies is greater than 1 (1.42 and 1.32 respectively), using the same allele (Allele1 = "C") for analysis.

#### 20 Recurrent Myocardial Infarction (RMI) studies (see Table 7)

In order to identify genetic markers associated with recurrent myocardial infarction (RMI), samples from the Cholesterol and Recurrent Events (CARE) study were genotyped utilizing 864 assays for functional SNPs in 500 candidate genes. A well-documented MI was one of the enrollment criteria for the CARE study. Patients were followed up for 5 years and rates of recurrent MI were recorded in Pravastatin treated and placebo groups.

In the initial analysis (CARE), SNP genotype frequencies were compared in a group of 264 patients who had another MI (second, third, or fourth) during the 5 years of CARE follow-up (cases) versus the frequencies in the group of 1255 CARE patients who had not experienced a second MI (controls).

To replicate the initial findings, a second group of 394 CARE patients were analyzed who had a history of an MI prior to the MI at CARE enrollment but who had not experienced an MI during trial follow-up (cases), and 1221 CARE MI patients

without a second MI were used as controls (Pre-CARE Study). No patients from the CARE Study were used in the Pre-CARE Study.

The SNPs replicated between CARE and Pre-CARE Studies were also tested for primary MI in study S0012 (UCSF). Study S0012 had 1400 samples. MI patients (cases) in this study had a self-reported history of MI, and controls had no history of MI or of acute angina lasting more than 1 hour. Allele frequencies in this study were detected in pooling experiments, in which DNA from about 50 individuals was pooled, and allele frequencies were determined in pooled DNA. For study S0012, only male cases and controls were used in pooling studies.

DNA was extracted from blood samples using conventional DNA extraction methods like the QIA-amp kit from Qiagen. Genotypes were obtained on a PRISM 7900HT sequence detection PCR system (Applied Biosystems) by allele-specific PCR, similar to the method described by Germer et al (Germer S., Holland M.J., Higuchi R. 2000, *Genome Res.* 10: 258-266). Primers for the allele-specific PCR reactions are provided in Table 5.

Statistical analysis was done using asymptotic chi square test for allelic, dominant, or recessive association, or Armitage trend test for additive genotypic association in the non-stratified as well as in the strata. Replicated SNPs are provided in Table 7. A SNP is considered replicated if the at-risk alleles in both sample sets are identical, the p-values are less than 0.05 in both sample sets, and the significant association is seen in the same stratum in both sets or one stratum is inclusive of the other.

Effect sizes were estimated through allelic odds ratios and odds ratios for dominant and recessive models, including 95% confidence intervals. Homogeneity of Cochran-Mantel-Haenszel odds ratios was tested across different strata using the Breslow-Day test. A SNP was considered to be a significant genetic marker if it exhibited a p-value  $< 0.05$  in the allelic association test or in any of the 3 genotypic tests (dominant, recessive, additive). Haldane Odds Ratios were used if either case or control count was zero. SNPs with significant HWE violations in both cases and controls ( $p < 1 \times 10^{-4}$  in both tests) were not considered for further analysis, since significant deviation from HWE in both cases and controls for individual markers can be indicative of genotyping errors. The association of a marker with RMI was considered replicated if the marker exhibits an allelic or genotypic association test p-

value  $<0.05$  in one of the sample sets and the same test and strata are significant ( $p < 0.05$ ) in either one or two other independent sample sets.

Various modifications and variations of the described compositions, methods and  
5 systems of the invention will be apparent to those skilled in the art without departing from the scope of the invention. Although the invention has been described in connection with specific preferred embodiments and certain working examples, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention that are  
10 obvious to those skilled in the field of molecular biology, genetics and related fields are intended to be within the scope of the following claims.

**TABLE 1**

Gene Number: 8  
Celera Gene: hCG15848 - 146000220474494  
Celera Transcript: hCT6878 - 146000220474495  
Public Transcript Accession: NM\_007348  
Celera Protein: hCP35549 - 197000064932342  
Public Protein Accession: NP\_031374  
Gene Symbol: ATF6  
Protein Name: activating transcription factor 6  
Celera Genomic Axis: GA\_x5YUV32W2NG(94208..292030)  
Chromosome: Chr1  
OMIM number: 605537  
OMIM Information: ACTIVATING TRANSCRIPTION FACTOR 6;ATF6

Transcript Sequence (SEQ ID NO:1):  
GCARAAGTAGTTTGTCTTTAMTAGGCCACCGTCTCGTCAGCGTTACGGAGTATTTTGTCCGCCTGCCGCCGCCGT  
CCCAGATATTAATCACGGAGTTCCA  
GGGAGAAGGAACTTGTGAAATGGGGGAGCCGGCTGGGGTTGCCGGCACCATGGAGTCACCTTTTAGCCCCGGG  
ACTCTYTCACAGGCTGGATGAAGATTGG  
GATTCTGCTCTCTTTGCTGAACTYGGTTATTTACAGACACTGATGAGCTGCAATTGGAAGCAGCAAATGAGACG  
TATGAAAACAATTTTGATAATCTTG  
ATTTTGATTTGGATTTGRTGCCTTGGGAGTCAGACATTTGGGACATCAACAACCAAATCTGTACAGTTAAAGATA  
TTAAGGCAGAACCYCAGCCACTTTC  
TCCAGCCTCCTCAAGTTATTCAGTCTCRTCTCCTCGGTCAGTGGACTCTTATTCTTCAACTCAGCATGTTCTGAGG  
AGTTGGATTTGTCTTCTAGTTCT  
CAGATGTCTCCCCTTTCCTTATATGGTGAAAACCTCTAATAGTCTCTCTTCASCGGAGCCACTGAAGGAAGATAAGC  
CTGTCACTGGTYCTAGGAACAAGA  
CTGAAAATGGACTGACTCCAAAGAAAAAAATTCAGGTGAATTCAAACCTTCAATTCAGCCCAAGCCTTTATTGC  
TTCCAGCAGCACCCAAGACTCAAAC  
AAACTCCAGTGTTCCAGCAAAAACCATCATTATTCAGACAGTACCAACGCTTATGCCATTGGCAAAGCAGCAACC  
AATTATCAGTTTACAACCTGCACCC  
ACTAAAGGCCAGACGGTTTTGCTGTCTCAGCCTACTGTGGTACAACCTTCAAGCACCTGGAGTTCTGCCCTCTGCT  
CAGCCAGTCCTTGCTGTTGCTGGGG  
GAGTCACACAGCTCCCTAATCACGTGGTGAATGTGGTACCAGCCCCTTCAGCGAATAGCCCAGTGAATGGAAAA  
CTTTCCGTGACTAAACCTGTCCTACA  
AAGTACCATGAGAAATGTCGGTTCAGATATTGCTGTGCTAAGGAGACAGCAACGTATGATAAAAAATCGAGAAT  
CCGCTTGTCAGTCTCGCAAGAAGAAG  
AAAGAATATATGCTAGGGTTAGAGGCGAGATTAAAGGCTGCCCTCTCASAAAACGAGCAACTGAAGAAAGAAA  
ATGGAACACTGAAGCGGCAGCTGGATG  
AAGTTGTGTCAGAGAACCAGAGGCTTAAAGTCCCTAGTCCAAAGCGAAGAGTTGTCTGTGTGATGATAGTATTG  
GCATTTATAATACTGAACTATGGACC  
TATGAGCATGTTGGAACAGGATTCCAGGAGAATGAACCCTAGTGTGAGCCCTGCAAATCAAAGGAGGCACCTTC  
TAGGATTTTCTGCTAAAGAGGCACAG  
GACACATCAGATGGTATTATCCAGAAAAACAGCTACAGATATGATCATTCTGTTTCAAATGACAAAGCCCTGATG  
GTGCTAACTGAAGAACCATTGCTTT

ACATTCCTCCACCTCCTTGTGTCAGCCCCTAATTAAYACAACAGAGTCTCTCAGGTTAAATCATGAACTTCGAGGATG  
 GGTTCATAGACATGAAGTAGAAAG  
 GACCAAGTCAAGAAGAATGACAAATAATCAACAGAAAACCCGTATTCTTCAGGGTGCTCTGGAACAGGGGCTCAA  
 ATTCTCAGCTGATGGCTGTTCAATAC  
 ACAGAAACCACTAGTAGTATCAGCAGGAACTCAGGGAGTGAGCTACAAGTGTATTATGCTTCACCCAGAAGTTA  
 TCAAGACTTTTTTTGAAGCCATCCGCA  
 GAAGGGGAGACACATTTTATGTTGTGTCATTTCTGAAGGGATCACCTGCTGTTACCAGCTACCACCCATAACAAGA  
 CCACAAGACCAAAAATGTCAATTGT  
 GTTACCAGCAATAAACATAAATGAGAATGTGATCAATGGGCAGGACTACGAAGTGATGATGCAGATTGACTGTC  
 AGGTGATGGACACCAGGATCCTCCAT  
 ATCAAAAGTTCGTCRGTTCTCTTACCTCCGAGATCAGCAGAGGAATCAAACCAACACCTTCTTTGGCTCCCCTC  
 CCGCAGCCACAGAGGCAACCCACG  
 TTGTCAGCACCATCCCTGAGTCATTACAATAGCACCTGCAGCTATGCTGGAAAACCTGAGCGTGGGACCCTGCCA  
 GACTGAAGAGCAGGTGAGCAAAATG  
 CTGCTTTCTGCCTTGGTGGCAGGCAGAGAACTGTCTCGTACTAGAATTCAAGGAGRAAAGAAGAAGAAATAAAA  
 GAAGCTGCTCCATTTTTTCATCATCTA  
 CCCATCTATTTGGAAAGCACTGGAATTCAGATGCAAGAGAAACAATGTTTCTTCAGTGGCAAATGTAGCCCTGCAT  
 CCTCCAGTGTTACCTGGTGTAGATT  
 TTTTTTCTGTACCTTTCTAAACCTCTCTTCCCTCTGTGATGGTTTTGTGTTTAAACARTCATCTTCTTTTAAATAATA  
 TCCACCTCTCCTTTTTTGCCAT  
 TTCACTTATTGATTCATAAAGTGAATTTTATTTAAAGCTATGCCACACATGCATGTTCAAATGGTTTCCACTGATTC  
 GATTTTTTCATTCATTTAATGCAA  
 ACCCATTTCTGGATATTGTGCTTATTTGAGAAAACACATTTCAAACCCAGAAAAGCCAAAAACACTCCAAAAACAA  
 GCAAAACAATTTGGAGCTTTAGATA  
 AAAGGAAAAACTCCCAGTTGGTAAAGTTTATCTTTACTTAGGATTTGTGGCTCACACCTAAACAAAGGGGGGTCAG  
 GGAGTGGGTACAAATTTGAGAAAAT  
 AGAAGGGTAAGGGAAGGGCCAGTGGTGGGGTTTGGAGAGAGGAGATAGCTCCATTAATACACATGTTTAAAA  
 GATGGAAAGTTCACGCCTGTAATCCCAG  
 CACTTTGGGAGGCGGAGGCGGGCGGATCACGAGGTCAGGAGATCAAGACCATCCCGGCTAAAACGGTGAAACC  
 CCGTCTCTACTAAAAATACAAAAAATT  
 AGCCGGGCGTAGTGACGGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATGGCGTGAACCC  
 GGGAGGCGGAGCTTGCAGTGAGCCGAGA  
 TCCCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAGATGGA  
 AAGTTCGATGTGACTGCAGTATGAGAT  
 TAAAGCCACAACCTATTGTTTATTTTGGGGACTCTAGGCCACCAAGTATTAGCACACATACTTATGTTTTCTCTACTA  
 ATCTGGTCCAGGTCCTCATGGAC  
 CACAGGACAAAGCTTTTCAATTTTCAATTCATTCTTATTGAAATTATACCAAATTCAGCTGAGGAATATGGAAGTAA  
 CTTTAGACTTAAACAAGACAAAAG  
 TTTTTTCACTGAAGAATTGACAAGTATTTGCTCCTTAAACAACGCAGATTAGTGAACGTGGATTCTGCTGAGG  
 GAGTGCATCCCATAATATGGCAATA  
 ATTTTCAGTTTCTCCAACGAAAAGATAGTGAAGGAATTAATCTTTTGTCTCCCATGGTTAAAAAAAAAAAAAAAAA  
 GCTGTGTTTCAATTTTACTGTACTAT  
 GCCTCTTTTTTACCATAGTAGACAATTATGTTTCATTTGATGAATTCATAGAACTGGATCTCATACAGCGATGTC  
 CTCTCTAATGTTCTACCTTTCAGT

TTCTAAAGTGAGTCTTCCTCCCTCTCCTACAAAACCTTTTCAATTTTTTGATGTAACATCTACAAATACTGTTTCTT  
 ACCCCAGTTGACTTGCCTTTGT  
 CAGATTTCTTCTTGTTCCACACTATAGCAATCAATTTCTTCTTCCTTACAAGAAAGGGAACGAGAAATTGTAGC  
 AACCTCTCAAGGATTATATGCAGC  
 TAGTTAGTTTTCTGCCTGTGAAATTAGGTCTGGCTCCTAAATAATTTTAAAGAACCATCAGCACTTCTAACTCTCT  
 GGACAGGTGCCTCTTTGTCCAAGC  
 TAGTTAAATGCTTTCCAAGGAAATCAGTTCAACTTTTGTGAGCGGGGAAAAGCAGGGCTTTATTGTTGTGTTACC  
 TGGGAGTCTGGAGTTTGAAAAGTGC  
 TAATTAACCTTCCTCTTTTTCCACATTACAAACCTTTTTAAGCAGCGCAGCACTCCCYTTAGATTTGGCTATCCTGG  
 GTGATTTTCAGACAAGAACCATT  
 TTCTCTGGGGACCATTCTTCTGCTGGGTGCCAAGGAATATAAGGCAAATGCCCAGAAGACCTTCAGGTGACTGG  
 GCAGTCTTATCATGGGATATTTCTTC  
 TGGCCCTGCCCCCTCCATTCTGTAATGTGAATTAGCCACACCAGAGGCTGTGACCATGGCTAGTAGACAGTGGC  
 AACATAGTCATCCCCAAGATGCTAA  
 TCTTCTGCTGGAACGTGCATACGTTATCATGGTCAATGTAAACCTGGTTTGTGTGGGGTGATTATAAATAGAGTTT  
 CCTCCTCTCTGTGACAGAATCAC  
 AGGAGAAGGACCCATCTCGTGGCCTTCTTGTTCTTAGCGCTTCACTTTTACTTCATCCCTCGATTCCCAGCTTTTTTC  
 TATCATCATTTTRCCAACCTCCTC  
 AGATGCAAGACTTTGGTTATGTCATACTCACCAACGTTAGTCCCTCTCTTCCAGGTGAAAAGGTGGGTAGCGGTT  
 GGGAGGGAGTCTCCACTGAAGAGCA  
 GGAAGGTGGTAGCAGGGCCGGCAGCTCTGCCACAGAGCTAGGGGTGCCTGTAAGGTGCCGCCTAGAGCAGCCT  
 GGGAGCTTTGCCTTCTTTGTCTCTCA  
 CTAGCCCTTCTACTCTTTGTCAATTGCCTGTTCTTGAGTGGATCTTTGAGATGAGGGGACAGGATTCTCCTAAGGGT  
 AGAGTTTCAGGAAATGAGTGAAAG  
 GCAATTGACAAATGCAAAGAAGTAGTCACTTTTTAAATTGCGGGCAAAGCTATAATTAATCCCTAGGCACAATTG  
 TAGTTTTTATTTAATGTTTGTATG  
 CACAAGGCCCTTTAGGAAATGAGAAGTTGCCATGCCAGATTAATTTTTTTTTTTTTTTTTTTGGTGGGATTGCCTTTT  
 GGGGGTTGCAGCCAGAAATTGTGG  
 GTAATGTGTGTATTTTTTTATTTATTAAATTTTAAACAGGATTGTGCAAGCTTATGAGACAATTAGATAAACTCAT  
 GGAGGAGGCAGGTCCTCCTGTTAT  
 TAGATGATTTTGTGCTCTTTGGGCTGACAATAATACTCTTGGGAAGTGATGGTAGAGACTGATGGGAATAGT  
 CTTTCTGCCTGGTTGCAAGTCCCAA  
 TTTTTAAGGSTTAATGGAAGTAAGTGGATGTTTCCTCATGTAACTACTGAATCAGATGTTAGGAGCTTGTCCCTT  
 TGGGGTTGACTTATGCCCAGCAGT  
 ACAGGGACACAGCTTCATTAGAGTGTTAGTGTAATACTAACTCCAAAGTTAGGAGTTAATGTGAAAGGATCATCC  
 TTGAAACAARTCTGCTGTTTGCCATG  
 CTTGTAGTACAGAACTTCACATGGAGTTTTGGGTGGGATTTGTGTTTTTACAAGTAAAAAATCCCTCACGATTA  
 TAAACTCAGAGCATCATCTAATTT  
 TTTTTTTAATGACTACAAGTTCCAGCACAAAACCTGGCATTCTTTGCCATTTCTTGCCAGTAAGAAGTTGACACG  
 GAGGTATTTGAAAGCAATGTTATG  
 TGAGTCATTCTTAAGTGTTCCAAGTAAGTTTAGAAACAGAAAAGGAACTTGGGATTCAAATTGATTTTTCAAATC  
 ATTTTAAAGAGACATCATCCTGAC  
 TAAATCTTAGCCTGAACCTTCCTCCCCTGTGTGTATTCCCCGGTAGTCACCGCAGCGAGATGCTGGTGAGACTGC  
 CGTGGTGGCATTTAGCATCGTTAAA

ACTGGAAACTCTCAAGCTCTTTGCCACTTTCCTACTATTTTTTGATTCTTGCCATTTTACCAAGCTTAGGTTGTGA  
 AACTTGACAGAAATGTATTACAG  
 GAAAARCTTATAATTGTATTTGACTTTCTAACACATTGCAAAGTTTCAAAGTGACTTTCACTTTCAACAACATATTA  
 GAAGTAACCACTTTTGCTTTCAC  
 AGCCTGAAGAGTTAGAGCCTGATCTGATGCCCCCTTCACTCTGAAGTCATGGGAAATTTTCCAGCCATGAAAGC  
 CCTCTTCCACTGCATACTGATGGG  
 CTGACTCAGCTTCCTTCAGCCGACTGAGATCTTTTCATACTATTGGCTATTTTCATACCAATTAACCTCTTAAATAAG  
 ATTGTGAATTGCCAAAATTGATA  
 GACACTTATTACCACCTGTGGACTCCATATTCCTTACCACAAATGTTATTTTCATCAGTCCTGAGTCATTTTAACTT  
 ACAGAAATTAGGATTGTTGCTGC  
 TAATATGAATACCAATTATAACTTTTAGAAACAAGAATAAAGCCTAAAAGAGAATGAAATATAAGAAATGTTTCGT  
 TCCCACCCCTAATAACATTTGGAAG  
 TGAATATTCCCATTTTCTTCCACCCACAGGGATTGGGATTGATTTTAAATTCCTAGGAAACAATACTAGACTACC  
 CAAAAGATGTTGCCAGAATCCAA  
 AAGGAACTATGCTCGTAAAAGAAATGCAGTTTTCTCCTACCTAAAAAAAAGAAAGTAAAGTGTGTTCTGTTCTTA  
 TCTTTTAATGACTAAGCTTTAAAC  
 AGTTTATTTTGGGTAAGACTAGAACTTTCAGCCATTTGTTCTAATATCTGTGTTATTAGATGCAATAGAATTTATG  
 AAAAGAAGAATGACAAAGGTATCT  
 GATTAGAAAATTTGATCTTACGCATGAATCCATGTCATGGCCAGCCACTGTCACATAGTGGGTGCCATTCTCAAC  
 ATATTGGTTTGCTAACTTTAAGCAT  
 TAGGGATTTAGCACACTAAAATACTTTTAATTATATTAGGTTTGGTAACTAAGGAGTAAATAAATCATAATTTATC  
 ATTTGCCAAGGCCAACAAACAACA  
 CTATTGTGCTGTTTGCTCTCAATGAAGTTGAATAAACCAGGAGGCTTGGCATATCCCCTTTATGTTAATCCCAGCT  
 AGAGATTAGTAGGTTGACTTTCAC  
 AGCAATTGTATATTGATCCATTTTAACTCATCCTTGCCATAATTTCCAGGCCAGTCACCAGGACAGAGGAGATGA  
 TGGGGAAACAGAGCTTTAGATGAAA  
 ACTACTATGCACTACTAGCCTTAGAGGCACTGGTTTCCTGTTACCACTTTGGCAAGTATGGATGGTCTAAGTCCA  
 GTAGGGCTTCATCCATGGAGCCATT  
 AGAACTGAGGGGGGAGTGTTAGAGATGCCATTTACCAGGATCTTTTTGCTCAGGTTGTACCCATGCCAATTGA  
 AGAACGTGTTAAAGATGAGGAGGAGA  
 GATGTACCATTCTCTCCCTTAATAATGATGTTGGTTTGCAAAACCTAAAGAAATAATAACAACAGACTATTTTCATA  
 CTTTCAAGCAAGTCTTTATACTAC  
 CTGTTATTTCTCTAAAATTCAAATAAAGAATTTTAAACTTA

Protein Sequence (SEQ ID NO:27):

MGEPAGVAGTMESPFSPGLFHRLDEDWDSALFAELGYFTDTDELQLEAANETYENNFDNLDFDLMPWESDIWDI  
 NNQICTVKDIKAEPQPLSPASSY  
 SVSSPRSVDSYSSTQHVP EELDLSSSSQMSPLSLYGENSNSLSAEPLKEDK PVTGPRNKTE NGLTPKKKIQVNSKPSIQ  
 PKPLLLPAAPKTQTNSSVPA  
 KTIHQTVPTLMPLAKQQPIISLQPAPTKGQTVLLSQPTVVQLQAPGVLPSAQPV LAVAGGVTQLPNHVNVVPAPSA  
 NSPVNGKLSVTKPVLQSTM RNV  
 GSDIAVLRRQQRMIKNRESACQSRKKKKEYMLGLEARLKAALSENEQLKKENGTLKRQLDEVVSENQRLKVPSPKRRV  
 VCVMIVLAFIILNYGPM SML EQ  
 DSRRMNPSVSPANQRRHLLGFS AKEAQDTS DGIIQKNSYRYDHSVSN DKALMVLTEEPLLYIPPPCQPLINTTESLRL  
 NHELRGWVHRHEVERTKSRRM

TNNQQKTRILQGALEQGSNSQLMAVQYTETTSSISRNSGSELQVYYASPRSYQDFFEAIRRRGDTFYVVSFRRDHLLLP  
ATTHNKTTTPKMSIVLPAINI  
NENVINGQDYEVMMQIDCQVMDTRILHIKSSSVPPYLRDQQRNQTNFFGSPPAATEATHVVSTIPESLQ

SNP Information

Context (SEQ ID NO:47):  
GTCTTCTAGTTCTCAGATGTCTCCCCTTTCTTATATGGTGAAAACCTCTAATAGTCTCTCTTCAGCGGAGCCACTGA  
AGGAAGATAAGCCTGTCACTGGT  
Y  
CTAGGAACAAGACTGAAAATGGACTGACTCCAAAGAAAAAAATTCAGGTGAATTCAAACCTTCAATTCAAGCCC  
AAGCCTTTATTGCTTCCAGCAGCACC  
Celera SNP ID: hCV25631989  
SNP Position Transcript: 588  
SNP Source: Applera  
Population(AAllele,Count): african american(C,28|T,4) caucasian(C,35|T,5) total(C,63|T,9)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:27, 9, (P,CCT) (S,TCT)

Gene Number: 8  
Celera Gene: hCG15848  
Celera Transcript: CRA|335000489562722  
Public Transcript Accession: NM\_007348  
Celera Protein: CRA|224000006548847  
Public Protein Accession: NP\_031374  
Gene Symbol: ATF6  
Protein Name: activating transcription factor 6  
Celera Genomic Axis: GA\_x5YUV32W2NG(94208..292030)  
Chromosome: Chr1  
OMIM number: 605537  
OMIM Information: ACTIVATING TRANSCRIPTION FACTOR 6;ATF6

Transcript Sequence (SEQ ID NO:2):  
AAGATATTAATCACGGAGTTCCAGGGAAAAGGAACTTGTGAAATGGGGGAGCCGGCTGGGGTTGCCGGCACCA  
TGGAGTCACCTTTTAGCCCGGGACTCT  
YTCACAGGCTGGATGAAGATTGGGATTCTGCTCTCTTTGCTGAACTYGGTTATTTACAGACACTGATGAGCTGC  
AATTGGAAGCAGCAAATGAGACGTA  
TGAAAACAATTTTGATAATCTTGATTTTGATTTGGATTTGRTACCTTGGGAGTCAGACATTTGGGACATCAACAAC  
CAAATCTGTACAGTTAAAGATATT  
AAGGCAGAACCYCAGCCACTTTCTCCAGCCTCCTCAAGTTATTCAGTCTCRTCTCCTCGGTCAGTGGACTCTTATTC  
TTCAACTCAGCATGTTCTGAGG

AGTTGGATTTGTCTTCTAGTTCTCAGATGTCTCCCCTTTCCTTATATGGTGAAAACCTCTAATAGTCTCTCTTCASCG  
 GAGCCACTGAAGGAAGATAAGCC  
 TGTCAGTGGTYCTAGGAACAAGACTGAAAATGGACTGACTCCAAAGAAAAAAATTCAGGTGAATTCAAAACCTT  
 CAATTCAGCCCAAGCCTTTATTGCTT  
 CCAGCAGCACCCAAGACTCAAACAACTCCAGTGTTCCAGCAAAAACCATCATTATTCAGACAGTACCAACGCTT  
 ATGCCATTGGCAAAGCAGCAACCAA  
 TTATCAGTTTACAACCTGCACCCACTAAAGGCCAGACGGTTTTGCTGTCTCAGCCTACTGTGGTACAACCTTCAAGC  
 ACCTGGAGTTCTGCCCTCTGCTCA  
 GCCAGTCCTTGCTGTTGCTGGGGGAGTCACACAGCTCCCTAATCACGTGGTGAATGTGGTACCAGCCCCTTCAGC  
 GAATAGCCCAGTGAATGGAAAACCTT  
 TCCGTGACTAAACCTGTCCTACAAAGTACCATGAGAAATGTCGGTTCAGATATTGCTGTGCTAAGGAGACAGCA  
 ACGTATGATAAAAAAATCGAGAATCCG  
 CTTGTCAGTCTCGCAAGAAGAAGAAAGAATATATGCTAGGGTTAGAGGCGAGATTAAAGGCTGCCCTCTCASAA  
 AACGAGCAACTGAAGAAAGAAAATGG  
 AACACTGAAGCGGCAGCTGGATGAAGTTGTGTCAGAGAACCAGAGGCTTAAAGTCCCTAGTCCAAGCGAAGA  
 GTTGTCTGTGTGATGATAGTATTGGCA  
 TTTATAATACTGAACTATGGACCTATGAGCATGTTGGAACAGGATTCCAGGAGAATGAACCCTAGTGTGGGACC  
 TGCAAATCAAAGGAGGCACCTTCTAG  
 GATTTTCTGCTAAAGAGGCACAGGACACATCAGATGGTATTATCCAGAAAAACAGCTACAGATATGATCATTCTG  
 TTTCAAATGACAAAGCCCTGATGGT  
 GCTAACTGAAGAACCATTGCTTTACATTCCCCCACCTCCTTGTCAGCCCCTAATTAAYACAACAGAGTCTCTCAGG  
 TTAATCATGAACCTTCGAGGATGG  
 GTTCATAGACATGAAGTAGAAAGGACCAAGTCTAGAAGAATGACAAATAATCAACAGAAAACCCGTATTCTTCA  
 GGGTGTTGTGGAACAGGGCTCAAATT  
 CTCAGCTGATGGCTGTTCAATACACAGAAACCACTAGTAGTATCAGCAGGAACTCAGGGAGTGAGCTACAAGTG  
 TATTATGCTTCACCCAGAAGTTATCA  
 AGACTTTTTTTGAAGCCATCCGCAGAAGGGGAGACACATTTTATGTTGTGTCATTTCTGAAGGGATCACCTGCTGTT  
 ACCAGCTACCACCCATAACAAGACC  
 ACAAGACCAAAAATGTCAATTGTGTTACCAGCAATAAACATAAATGAGAATGTGATCAATGGGCAGGACTACGA  
 AGTGATGATGCAGATTGACTGTCAGG  
 TGATGGACACCAGGATCCTCCATATCAAAAGTTCGTCRGTTCCCTCCTTACCTCCGAGATCAGCAGAGGAATCAAA  
 CCAACACCTTCTTTGGCTCCCCCTCC  
 CGCAGCCACAGAGGCAACCCACGTTGTCAGCACCATCCCTGAGTCATTACAATAGCACCCGCAGCTATGTGGAA  
 AACTGAGCGTGGGACCCCCAGACTGA  
 AGAGCAGGTGAGCAAAATGCTGCTTTTCTTGGTGGCAGGCAGAGAACTGTTCTGACTAGAATTCAAGGAGAAA  
 AGRAGAAGAAATAAAAGAAGCTGCTC  
 CATTTTTTCATCATCTACCCATCTATTTGGAAAGCACTGGAATTCAGATGCAAGAGAACAAATGTTTCTTCAGTGGCA  
 AATGTAGCCCTGCATCCTCCAGTG  
 TTACCTGGTGTAGATTTTTTTTTCTGTACCTTTCTAAACCTCTCTTCCCTCTGTGATGGTTTTGTGTTTAAACARTCA  
 TCTTCTTTTAAATAATATCCAC  
 CTCTCCTTTTTGCCATTTCACTTATTGATTCATAAAGTGAATTTTATTTAAAGCTAAAAAAAAAAAAAAAAAAAAA

Protein Sequence (SEQ ID NO:28):

MGEPAGVAGTMESPFSPGLFHRLDEDWDSALFAELGYFTDTDELQLEAANETYENNFDNLDFDLPLPWESDIWDI  
 NNQICTVKDIKAEPQPLSPASSY

SVSSPRSVDSYSSTQHVPEELDLSSSSQMSPLSLYGENSNSLSPEPLKEDKPV TGSRNKTENGLTPKKKIQVNSKPSIQP  
KPLLLPAAPKTQTNSSVPA  
KTIIIQTVPTLMPLAKQQPIISLQPAPT KGQTVLLSQPTVVQLQAPGVLPSAQPVLAVAGGVTQLPNHVNVVPAPSA  
NSPVNGKLSVTKPVLQSTMRNV  
GSDIAVLRRQQRM IKNRESACQSRKKKKEYMLGLEARLKAALSENEQLKKENGTLKRQLDEVVSENQRLKVPSPKRRV  
VCVMIVLAFIILNYGPMSMLEQ  
DSRRMNPSVGPANQRRHLLGFSAKEAQDTS DGIIQKNSYRYDHSVSN DKALMVLTEEPLLYIPPPPCQPLINTTESLRL  
NHELRGWVHRHEVERTKSRRM  
TNNQQKTRILQGVVEQGSNSQLMAVQYTETTSSISRNSGSELQVYYASPRSYQDFFEAIRRRGDTFYVVSFRRDHLLP  
ATTHNKTTTPKMSIVLPAINI  
NENVINGQDYEVMMQIDCQVMDTRILHIKSSSVPPYLRDQQRNQNTNFFGSPPAATEATHVVSTIPESLQ

SNP Information

Context (SEQ ID NO:48):  
GTCTTCTAGTTCTCAGATGTCTCCCCTTTCCTTATATGGTGAAA ACTCTAATAGTCTCTCTTCACCGGAGCCACTGA  
AGGAAGATAAGCCTGTCACTGGT  
Y  
CTAGGAACAAGACTGAAAATGGACTGACTCCAAAGAAAAAAATTCAGGTGAATTCAA AACCTTCAATTCAGCCC  
AAGCCTTTATTGCTTCCAGCAGCACC

Celera SNP ID: hCV25631989  
SNP Position Transcript: 511  
SNP Source: Applera  
Population(Allele,Count): african american(C,28|T,4) caucasian(C,35|T,5) total(C,63|T,9)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:28, 9, (P,CCT) (S,TCT)

Gene Number: 20  
Celera Gene: hCG1644130 - 146000219722637  
Celera Transcript: hCT2285762 - 146000219722661  
Public Transcript Accession: NM\_033554  
Celera Protein: hCP1854305 - 197000069367439  
Public Protein Accession: NP\_291032  
Gene Symbol: HLA-DPA1  
Protein Name: major histocompatibility complex, class II, DP alpha 1;HLA-DP1A;HLADP;HLASB  
Celera Genomic Axis: GA\_x5YUV32W6W6(6045879..6073199)  
Chromosome: Chr6  
OMIM number: 142858  
OMIM Information: MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS II, DP BETA-1;HLA-DPB1

Transcript Sequence (SEQ ID NO:3):  
STYATGTTGGGTTTTGATTATTTGTGGTGCTAAGAAGAAGGGAACATGGTTGGAGGCC CAGTGAGAGAAACAGT  
GCTTTGAATCAAAGAGCAGAAYGATA

GAAACTGACTTCAGAGCAACTTCTTGGCAGCAGTATCCWMTTKGGAAGYTGARGGTCTGTCCTGGARCCAGRT  
 GYTAACRAAACACAGCAAATGCTTTTC  
 CTAAGGCACAWTAGTCTTTTCRGTGAGSTCAGGAACCCTGTTTATGCAGATCCTCGTTGAACTTTCTTGCTCCTCC  
 TGTGCATGAAGATGCCCACTCCAC  
 AGATGATGAGCCCCAGCACGAAGCCCCCAGCTCCCGTCAATGTCTTACTCCGGGCAGAATCAGACTGTGCCTTCC  
 ACTCCACGGTGACAGGACTRTCCAK  
 GCTGGTGTGCTCCACTTGGCAGRTGTAGACRTCTCCCTGCTGGGGGGTCAATTCAGCATCWCCAGGATCTGGA  
 AGGTCCAGTCTCCATTACGGATCAGG  
 TTGGTGGACACGACCCCAGCTGTTTCCTCCTGTCCATTGAGGAASCATCGGACTTGAATGCTGYCTGGGTAGAAA  
 TCYGTACGTGGYAGACAAGCAGGT  
 TGTGGTGTGCTGCARGGGCYCCTCTCGTAGTTGTGTCTGCAYAYCCTGTCCGGCACTGCCCCGYCTCCTCCAGGAKG  
 TCCTTCTGGCTGTTCCAGTASTCC  
 KCAKAGGCCGCCCCAGCTCCGTACCCGCCCGGAACTCYCCCACGTGCTGTGGAAGCGCRCGWRCTCCTSCCG  
 GTTGTAGATGTATCTCTCCAGGAAGC  
 GCTGTGTCCCATTAACGCGTAGCATTCTGCCGTMMCTGGWAMASTCATCAATTATAGACCCYACARCATGC  
 GCCCWGAAGACAGAATGTTCCATATCA  
 GAGCTGTGATCTTGAGAGCCCTCTCCTTGGCTTTCCTGCTGAGTCTCCGAGGAGCTGGGGCCATCAAGGCGGAC  
 CATGTGTCAACTTATGCCRYGTTTGT  
 ACAGACSCATAGAMCAACAGGRGAGTTTAKGTTTGARTTTGATGAHGATGAGMWGTTCTATGTGGATCTGGAY  
 AARAAGGAGACCGTCTGGCATCTGGAG  
 GAGTTTGGCCRARCCTTTTCCTTTGAGGCTCAGGGCGGGCTGGCTAACATTGCTATATYGAACAACAACCTTGAAT  
 ACCTTGATCCAGCGTTCCAACCACA  
 CTCAGGCCRCCAAYGATCCCCCTGAGGTGACCGTGTTTCCCAAGGAGCCTGTGGAGCTGGGGCCAGCCCAACACC  
 CTCATCTGCCACATTGACARGTTCTT  
 CCCACCAGTGCTCAAYGTCACRTGGCTGTGCAAYGGGRAGCYRGTCCTGAGGGTGTGCTGAGAGCYTCTTCCT  
 GCCCAGAACAGATTACAGCTTCCAC  
 AAGTTCCATTACCTGACCTTTGTRCCCTCAGCAGAGGACKTCTATGACTGCAGGGTGGAGCACTGGGGCTTGGA  
 CCAGCCGCTCCTCAAGCACTGGGAGG  
 CCCAAGAGCCAATCCAGATGCCTGAGACAACGGAGACTGTGCTCTGTGCCCTGGGCCTGGTGCTGGGCCTAGTS  
 GGCATCATCGTGGGCACCGTCCTCAT  
 CATAAAGTCTCTGCGTTCTGGCCATGACCCCCGGGCCAGGGGACCCTGTGAAATACTGTAAAGGTGGGAATGT  
 AAAGAGGAGGCCCTAGGATTTGTAGA  
 ATGTGACAAAATATCTGAACAGAAGAGGACTTAGGAGAGATYTGAACTCCAGCTGCCCTACAACTCCRTCWCA  
 GCTTTTCTTCTCACTTCATGTGAAAA  
 CTACYCCAGTGGCTGACTGAATTGCTGACCCTTCAAGCTCTGTCCTTATCCRTTACCTHAAAGCAGTCATTCCTTA  
 GTAAAGTTTCCAACAAATAGAAAT  
 TAATGACACTTTGRTAGCACTRATAYGGAGATTATCCTTTCATTGAGCCTTTTATCCTCTGKTYTYCTTTGAARARC  
 CCCTCACTGTCACCTTCCYGAGA  
 ATACYCTAAGACCAATAAATACTTCAGTATTTTCAGAGCR

Protein Sequence (SEQ ID NO:29):

MRPEDRMFHIRAVILRALSLAFLLSLRGAGAIKADHVSTYAAFVQTHRPTGEFMFEFDEDEM FYVDLDKKETVWHLE  
 EFGQAFSFEAQGGLANIAILNNN  
 LNTLIQRSNHTQATNDPPEVTVFPKEPVELGQPNTLICHIDKFFPPVLNVTWLCNGELVTEGVAESLFLPRTDYSFHKF  
 HYLTFVPSAEDFYDCRVEHWG

LDQPLLKHWEAQEPIQMPETTETVLCALGLVLGLVGIIVGTVLIIKSLRSGHDPRAQGTL

SNP Information

Context (SEQ ID NO:49):  
TGTGGTGCTGCAAGGGCCCCTCTCGTAGTTGTGTCTGCACATCCTGTCCGGCACTGCCCCGCTTCTCCTCCAGGAT  
GTCCTTCTGGCTGTTCCAGTACTCC  
K  
CAGCAGGCCGCCCCAGCTCCGTCACCGCCCGGAACTCCCCCACGTCGCTGTCGAAGCGCGCGAACTCCTCCCCGG  
TTGTAGATGTATCTCTCCAGGAAGCG  
Celera SNP ID: hCV8851074  
SNP Position Transcript: 701  
SNP Source: Applera  
Population(AAllele,Count): african american(G,15|T,17) caucasian(G,9|T,5) total(G,24|T,22)  
SNP Type: UTR 5  
Protein Coding: SEQ ID NO:29, None

SNP Source: HGBASE;dbSNP  
Population(AAllele,Count): no\_pop(T,-|G,-) ;no\_pop(T,-|G,-)  
SNP Type: UTR 5  
Protein Coding: SEQ ID NO:29, None

Gene Number: 20  
Celera Gene: hCG1644130 - 146000219722637  
Celera Transcript: hCT1967720 - 146000219722647  
Public Transcript Accession: NM\_033554  
Celera Protein: hCP1779945 - 197000069367438  
Public Protein Accession: NP\_291032  
Gene Symbol: HLA-DPA1  
Protein Name: major histocompatibility complex, class II, DP alpha 1;HLA-DP1A;HLADP;HLASB  
Celera Genomic Axis: GA\_x5YUV32W6W6(6045879..6073199)  
Chromosome: Chr6  
OMIM number: 142858  
OMIM Information: MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS II, DP BETA-1;HLA-DPB1

Transcript Sequence (SEQ ID NO:4):  
STYATGTTGGGTTTTGATTATTTGTGGTGCTAAGAAGAAGGGAACATGGTTGGAGGCCAGTGAGAGAAACAGT  
GCTTTGAATCAAAGAGCAGAAAYGATA  
GAAACTGACTTCAGAGCAACTTCTTGGCAGCAGTATCCWMTTKGGAAGYTGARGGTCTGTCCTGGARCCAGRT  
GYTAACRAAACACAGCAAATGCTTTTC  
CTAAGGCACAWTAGTCTTTTCRGTTGAGSTCAGGAACCCTGTTTATGCAGATCCTCGTTGAACTTTCTTGCTCCTCC  
TGTGCATGAAGATGCCCACTCCAC  
AGATGATGAGCCCCAGCACGAAGCCCCAGCTCCCGTCAATGTCTTACTCCGGGCAGAATCAGACTGTGCCTTCC  
ACTCCACGGTGACAGGACTRTCCA

GCTGGTGTGCTCCACTTGGCAGRTGTAGACRTCTCCCTGCTGGGGGGTCATTTCCAGCATCWCCAGGATCTGGA  
 AGGTCCAGTCTCCATTACGGATCAGG  
 TTGGTGGACACGACCCAGCTGTTTCCTCCTGTCCATTAGGAASCATCGGACTTGAATGCTGYCTGGGTAGAAA  
 TCYGTACGTGGYAGACAAGCAGGT  
 TGTGGTGTGCTGCARGGGCYCCTCTCGTAGTTGTGTCTGCAYAYCCTGTCCGGCACTGCCCCGYCTCCTCCAGGAKG  
 TCCTTCTGGCTGTTCCAGTASTCC  
 KCAKCAGGCCGCCAGCTCCGTCACCGCCCGGAACTCYCCACGTCGCTGTCGAAGCGCRCGWRCTCCTSCCG  
 GTTGTAGATGTATCTCTCCAGGAAGC  
 GCTGTGTCCATTAAACGCGTAGCATTCTGCCGTMCTGGWAMASGTGAATCCCAGCCATGCTGATTCTCTC  
 CACCCATTSCAGTGCTAGAGGCCCA  
 CAGTTTCAGTCTCATCTGCCTCCACTCGGCCTCAGTTCCTYATCRCTGYTCCTGTGCTCACAGTCATCAATTATAGA  
 CCCYACARCATGCGCCCWGAAGA  
 CAGAATGTTCCATATCAGAGCTGTGATCTTGAGAGCCCTCTCCTTGGCTTTCCTGCTGAGTCTCCGAGGAGCTGG  
 GGCCATCAAGGCGGACCATGTGTCA  
 ACTTATGCCRYGTTTGTACAGACSCATAGAMCAACAGGRGAGTTTAKGTTTGARTTTGATGAHGATGAGMWGT  
 TCTATGTGGATCTGGAYAARAAGGAGA  
 CCGTCTGGCATCTGGAGGAGTTTGGCCRARCCTTTTCCTTTGAGGCTCAGGGCGGGCTGGCTAACATTGCTATAT  
 YGAACAACAACCTGAATACCTTGAT  
 CCAGCGTTCCAACCACACTCAGGCCRCCAAYGATCCCCCTGAGGTGACCGTGTTTCCAAGGAGCCTGTGGAGCT  
 GGGCCAGCCCAACACCCTCATCTGC  
 CACATTGACARGTTCTTCCCACAGTGCTCAAYGTCACRTGGCTGTGCAAYGGGRAGCYRGTCAGTGGGGTGTC  
 GCTGAGAGCYTCTTCTGCCCAGAA  
 CAGATTACAGCTTCCACAAGTTCCATTACCTGACCTTTGTRCCCTCAGCAGAGGACKTCTATGACTGCAGGGTGG  
 AGCACTGGGGCTTGGACCAGCCGCT  
 CCTCAAGCACTGGGAGGCCCAAGAGCCAATCCAGATGCCTGAGACAACGGAGACTGTGCTCTGTGCCCTGGGCC  
 TGGTGTGGGCCTAGTSGGCATCATC  
 GTGGGCACCGTCCTCATCATAAAGTCTCTGCGTTCTGGCCATGACCCCCGGGCCAGGGGACCCTGTGAAATACT  
 GTAAAGGTGACAAAATATCTGAACA  
 GAAGAGGACTTAGGAGAGATYTGAAGTCCAGCTGCCCTACAACTCCRTCWCAGCTTTTCTTCTCACTTCATGTG  
 AAAACTACYCCAGTGGCTGACTGAA  
 TTGCTGACCCTTCAAGCTCTGTCCTTATCCRTTACCTHAAAGCAGTCATTCCTTAGTAAAGTTTCCAACAAATAGA  
 AATTAATGACACTTTGRTAGCACT  
 RATAYGGAGATTATCCTTTCATTGAGCCTTTTATCCTCTGKTYTYCTTTGAARARCCCCTCACTGTCACCTTCCYGAG  
 AATACYCTAAGACCAATAAATA  
 CTTCAGTATTTCAAGAGCR

Protein Sequence (SEQ ID NO:30):

MRPEDRMFHIRAVILRALSLAFLSLRGAGAIKADHVSTYAAFVQTHRPTGEFMFEFDEDEM FYVDLDKKETVWHLE  
 EFGQAFSFEAQGGLANIAILNNN  
 LNTLIQRSNHTQATNDPPEVTVFPKEPVELGQPNTLICHIDKFFPPVLNVTWLCNGELVTEGVAESLFLPRTDYSFHKF  
 HYLTFVPSAEDFYDCRVEHWG  
 LDQPLLKHWEAQEPIQMPETTETVLCALGLVLGLVGIIVGTVLIKSLRSGHDPRAQGTL

SNP Information

Context (SEQ ID NO:50):

TGTGGTGCTGCAAGGGCCCCTCTCGTAGTTGTGTCTGCACATCCTGTCCGGCACTGCCCCGCTTCTCCTCCAGGAT  
GTCCTTCTGGCTGTTCCAGTACTCC

K

CAGCAGGCCGCCCCAGCTCCGTCACCGCCCGGAACTCCCCACGTCGCTGTCTGAAGCGCGCGAACTCCTCCCGG  
TTGTAGATGTATCTCTCCAGGAAGCG

Celera SNP ID: hCV8851074

SNP Position Transcript: 701

SNP Source: Applera

Population(AAllele,Count): african american(G,15|T,17) caucasian(G,9|T,5) total(G,24|T,22)

SNP Type: UTR 5

Protein Coding: SEQ ID NO:30, None

SNP Source: HGBASE;dbSNP

Population(AAllele,Count): no\_pop(T,-|G,-) ;no\_pop(T,-|G,-)

SNP Type: UTR 5

Protein Coding: SEQ ID NO:30, None

Gene Number: 45

Celera Gene: hCG17514 - 146000219707780

Celera Transcript: hCT8565 - 146000219707781

Public Transcript Accession: NM\_002121

Celera Protein: hCP37473 - 197000069367536

Public Protein Accession: NP\_002112

Gene Symbol: HLA-DPB1

Protein Name: major histocompatibility complex, class II, DP beta 1;HLA-DP1B

Celera Genomic Axis: GA\_x5YUV32W6W6(6056841..6073623)

Chromosome: Chr6

OMIM number: 142858

OMIM Information: MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS II, DP BETA-1;HLA-DPB1

Transcript Sequence (SEQ ID NO:10):

AGCTCCCTTTAGCGAGTCCTTCTTTTCCTGACTGCAGCTCTTTYCATTTTGCCATCCTTTTCCAGCTCCATGATGGTT  
CTGCAGGTTTCTGCGGCCCCC

GGACAGTGGCTCTGACGGCGTTACTGATGGTGCTGCTCACATCTGTGGTCCAGGGCAGGGCCACTCCAGAGAAT  
TACSTKTWCCAGKKACGGCAGGAATG

CTACGCGTTTAATGGGACACAGCGCTTCCTGGAGAGATACATCTACAACCGGSAGGAGYWCGYGCGCTTCGACA  
GCGACGTGGGRGAGTTCCGGGCGGTG

ACGGAGCTGGGGCGGCCTGMTGMGGASTACTGGAACAGCCAGAAGGACMTCCTGGAGGAGRRGCGGGCAGT  
GCCGGACAGGRTRTGACAGACAACTACG

AGCTGGDCGRGSCSRTGACCCTGCAGCGCCGAGTCCAGCCTARGGTGAAYGTTTCCCCCTCCAAGAAGGRGCCC  
YTGACGACCAACCTGCTTGTCTR

CCACGTGACRGATTTCTACCCAGRCAGCATTCAAGTCCGATGSTTCCTGAATGGACAGGAGGAAACAGCTGGGG  
TCGTGTCCACCAACCTGATCCGTAAT

GGAGACTGGACCTTCCAGATCCTGGWGATGCTGGAAATGACCCCCCAGCAGGGAGAYGTCTACAYCTGCCAAG  
 TGGAGCACACCAGCMTGGAYAGTCCTG  
 TCACCGTGGAGTGGAAGGCACAGTCTGATTCTGCCCGGAGTAAGACATTGACGGGAGCTGGGGGCTTCGTGCT  
 GGGGCTCATCATCTGTGGAGTGGGCAT  
 CTTTCATGCACAGGAGGAGCAAGAAAGTTCAACGAGGATCTGCATAAACAGGGTTCCTGASCTCACYGAAAAGAC  
 TAWTGTGCCTTAGGAAAAGCATTTGC  
 TGTGTTTYGTTARCACTGGYTCCAGGACAGACCYTCARCTTCCMAAKWGGATACTGCTGCCAAGAAGTTGCTCT  
 GAAGTCAGTTTCTATCRTTCTGCTC  
 TTTGATTCAAAGCACTGTTTCTCTCACTGGGCCTCCAACCATGTTCCCTTCTTCTTAGCACCACAAATAATCAAAC  
 CCAACATRASTGTTTGYTTTCCT  
 TTAAAAATATGCAYCAAATCRTCTCTCATYACTTTTCTCTGAGGGTTTTAGTAGACAGTAGGAGTTAATAAGAAG  
 TTCATTTTGGTTTAAACATAGGAA  
 AGAAGAGAACCATGAAAATGGGGATATGTAACTATTGTATAATGGGGCCTGTTACACATGACACTCTTCTGAAT  
 TGAAGTATTTTCAAGTGAAGCTGCCCC  
 CAAATCAAGTTTAGTGCCCTCATCCATTTATGTCTCAGACCACTATTCTTAACTATTCAATGGTGAGCAGACTGCA  
 AATCTGCCTGATAGGACCCATATT  
 CCCACAGCACTAATTCAACATATACCTTACTGAGAGCATGTTTTATCATTACCATTAAKAAGTTAAATGAACATCA  
 GAATTTAAAATCATAAATATAATC  
 TAATACACTTT

Protein Sequence (SEQ ID NO:31):

MMVLQVSAAPRTVALTALLMVLLTSVVQGRATPENYLFQGRQECYAFNGTQRFLERYIYNREEFARFDSVDVGEFRAV  
 TELGRPAAEYWNSQKDILEEKRA  
 VPDRMCRHNYELGGPMTLQRRVQPRVNVSPSKKGPLQHNNLLVCHVTDFYPGSIQVRWFLNGQEETAGVVSTNLI  
 RNGDWTFQILVMLEMTPQQGDVYTC  
 QVEHTSLDSPVTVEWKAQSDSARSKTLTGAGGFVLGLIICGVGIFMHRRSKKVQRGSA

SNP Information

Context (SEQ ID NO:52):

CGCTTCCTGGAGAGATACATCTACAACCGGGAGGAGTTCGCGCGCTTCGACAGCGACGTGGGGGAGTTCCGGG  
 CGGTGACGGAGCTGGGGCGGCCTGCTG  
 M

GGAGTACTGGAACAGCCAGAAGGACATCCTGGAGGAGAAGCGGGCAGTGCCGGACAGGATGTGCAGACACAA  
 CTACGAGCTGGGCGGGCCCATGACCCTG

Celera SNP ID: hCV8851074

SNP Position Transcript: 323

SNP Source: Applera

Population(AAllele,Count): african american(A,17|C,15) caucasian(A,5|C,9) total(A,22|C,24)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:31, 7, (E,GAG) (A,GCG)

SNP Source: HGBASE;dbSNP

Population(AAllele,Count): no\_pop(A,-|C,-) ;no\_pop(A,-|C,-)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:31, 7, (E,GAG) (A,GCG)

Gene Number: 45  
 Celera Gene: hCG17514 - 146000219707780  
 Celera Transcript: hCT2283126 - 146000219707830  
 Public Transcript Accession: NM\_002121  
 Celera Protein: hCP1854407 - 197000069367541  
 Public Protein Accession: NP\_002112  
 Gene Symbol: HLA-DPB1  
 Protein Name: major histocompatibility complex, class II, DP beta 1;HLA-DP1B  
 Celera Genomic Axis: GA\_x5YUV32W6W6(6056841..6073197)  
 Chromosome: Chr6  
 OMIM number: 142858  
 OMIM Information: MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS II, DP BETA-1;HLA-DPB1

Transcript Sequence (SEQ ID NO:11):

AGCTCCCTTTAGCGAGTCCTTCTTTTCCTGACTGCAGCTCTTTYCATTTTGCCATCCTTTCCAGCTCCATGATGGTT  
 CTGCAGGTTTCTGCGGCCCCC  
 GGACAGTGGCTCTGACGGCGTTACTGATGGTGCTGCTCACATCTGTGGTCCAGGGCAGGGCCACTCCAGAGAAT  
 TACSTKTWCCAGKKACGGCAGGAATG  
 CTACGCGTTTAATGGGACACAGCGCTTCCTGGAGAGATACATCTACAACCGGSAGGAGYWCGYGCGCTTCGACA  
 GCGACGTGGGRGAGTTCCGGGCGGTG  
 ACGGAGCTGGGGCGGCCTGMTGMGGASTACTGGAACAGCCAGAAGGACMTCCTGGAGGAGRRGCGGGCAGT  
 GCCGGACAGGRTRTGACAGACAACTACG  
 AGCTGGDCGRGSCSRTGACCCTGCAGCGCCGAGTCCAGCCTARGGTGAAYGTTTCCCCCTCCAAGAAGGRGCCC  
 YTGACGACCAACCTGCTTGCTR  
 CCACGTGACRGATTTCTACCCAGRCAGCATTCAAGTCCGATGSTTCCTGAATGGACAGGAGGAAACAGCTGGGG  
 TCGTGTCCACCAACCTGATCCGTAAT  
 GGAGACTGGACCTTCCAGATCCTGGWGATGCTGGAAATGACCCCCCAGCAGGGAGAYGTCTACAYCTGCCAAG  
 TGGAGCACACCAGCMTGGAYAGTCCTG  
 TCACCGTGGAGTGGAAGGCACAGTCTGATTCTGCCCCGAGTAAGACATTGACGGGAGCTGGGGGCTTCGTGCT  
 GGGGCTCATCATCTGTGGAGTGGGCAT  
 CTTTCATGCACAGGAGGAGCAAGAAAGTTCAACGAGGATCTGCATAAACAGGGTTCCTGASCTCACYGAAAAGAC  
 TAWTGTGCCTTAGGAAAANKWGGATAC  
 TGCTGCCAAGAAGTTGCTCTGAAGTCAGTTTCTATCRTTCTGCTCTTTGATTCAAAGCACTGTTTCTCTCACTGGGC  
 CTCCAACCATGTTCCCTTCTTCT  
 TAGCACCACAAATAATCAAAACCCAACATR

Protein Sequence (SEQ ID NO:32):

MMVLQVSAAPRTVALTALLMVLLTSVVQGRATPENYLFQGRQECYAFNGTQRFLERYIYNREEFARFDSVDVGEFRAV  
 TELGRPAAEYWNSQKDILEEKRA  
 VPDRMCRHNYELGGPMTLQRRVQPRVNVSPSKKGPLQHNNLLVCHVTDFYPGSIQVRWFLNGQEETAGVVSTNLI  
 RNGDWTFQILVMLEMTTPQQGDVYTC

QVEHTSLDSPVTVEWKAQSDSARSKTLTGAGGFVLGLIICGVGIFMHRRSKKVQRGSA

#### SNP Information

Context (SEQ ID NO:53):

CGCTTCCTGGAGAGATACATCTACAACCGGGAGGAGTTCGCGCGCTTCGACAGCGACGTGGGGGAGTTCCGGG  
CGGTGACGGAGCTGGGGCGGCCTGCTG

M

GGAGTACTGGAACAGCCAGAAGGACATCCTGGAGGAGAAGCGGGCAGTGCCGGACAGGATGTGCAGACACAA  
CTACGAGCTGGGGCGGGCCCATGACCCTG

Celera SNP ID: hCV8851074

SNP Position Transcript: 323

SNP Source: Applera

Population(AAllele,Count): african american(A,17|C,15) caucasian(A,5|C,9) total(A,22|C,24)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:32, 7, (E,GAG) (A,GCG)

SNP Source: HGBASE;dbSNP

Population(AAllele,Count): no\_pop(A,-|C,-) ;no\_pop(A,-|C,-)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:32, 7, (E,GAG) (A,GCG)

Gene Number: 45

Celera Gene: hCG17514 - 146000219707780

Celera Transcript: hCT1967009 - 146000219707799

Public Transcript Accession: NM\_002121

Celera Protein: hCP1779974 - 197000069367538

Public Protein Accession: NP\_002112

Gene Symbol: HLA-DPB1

Protein Name: major histocompatibility complex, class II, DP beta 1;HLA-DP1B

Celera Genomic Axis: GA\_x5YUV32W6W6(6056841..6073197)

Chromosome: Chr6

OMIM number: 142858

OMIM Information: MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS II, DP BETA-1;HLA-DPB1

Transcript Sequence (SEQ ID NO:12):

AGCTCCCTTTAGCGAGTCCTTCTTTTCTGACTGCAGCTCTTTYCATTTTGCCATCCTTTTCCAGCTCCATGATGGTT  
CTGCAGGTTTCTGCGGCCCCCC

GGACAGTGGCTCTGACGGCGTTACTGATGGTGCTGCTCACATCTGTGGTCCAGGGCAGGGCCACTCCAGAGAAT  
TACSTKTWCCAGKKACGGCAGGAATG

CTACGCGTTTAATGGGACACAGCGCTTCCTGGAGAGATACATCTACAACCGGSAGGAGYWCGYGCGCTTCGACA  
GCGACGTGGGRGAGTTCCGGGCGGTG

ACGGAGCTGGGGCGGCCTGMTGMGGASTACTGGAACAGCCAGAAGGACMTCCTGGAGGAGRRGCGGGCAGT  
GCCGGACAGGRTRTGCAGACACAACACTACG

AGCTGGDCGRGSCSRTGACCCTGCAGCGCCGAGGCTGGAGTGCTGTGGCACCATCATGGCTCACTGCAGCCTCA  
 ACCTCCTGGGCTCAAGTGATCCTCCT  
 GCCTCAGCCTCCCATGTAGCTAGAACTACAGATACACGTACCACCATGTCTGGCTAATTTATTTTCTTTTAGAG  
 ATGGGTTCTCACTATGTTGCCAG  
 GCCGGTCTCAAAACCCTGGGCTCAAGTGATCCTCATGCCTCAACCTCCCAAAGTGCTAAGATTATAGGCATGACC  
 ACCATGCCTGGCCTTTTCTGCTTTC  
 TGAGGAGGAAAAAGGTACTGGTGGCAGAGATCCAAAAGAAAAGTTGCCAGTGGCAGTGTGGAAATTCACCTGA  
 GAACAACAGGACAAGCTGGGGGCACAAA  
 TGCAAAGATGCAGAGGGAGGCAACACCTGGTCATCTGTGAGACCTTCATGGGACCTGAAGACGCAGCACAGAG  
 GAGGAACTTGAAAAAGGACGGGATTTC  
 TACTACTCAAGCATGTAGGAGCTCAGGATATTCTTCCAGCCTARGGTGAAYGTTTCCCCCTCCAAGAAGGRGCCC  
 YTGACGACCAACCTGCTTGTCT  
 RCCACGTGACRGATTTCTACCCAGRCAGCATTCAAGTCCGATGSTTCCTGAATGGACAGGAGGAAACAGCTGGG  
 GTCGTGTCCACCAACCTGATCCGTAA  
 TGGAGACTGGACCTTCCAGATCCTGGWGATGCTGGAAATGACCCCCCAGCAGGGAGAYGTCTACAYCTGCCAA  
 GTGGAGCACACCAGCMTGAYAGTCCT  
 GTCACCGTGGAGTGGAAGGCACAGTCTGATTCTGCCCCGAGTAAGACATTGACGGGAGCTGGGGGCTTCGTGC  
 TGGGGCTCATCATCTGTGGAGTGGGCA  
 TCTTCATGCACAGGAGGAGCAAGAAAGTTCAACGAGGATCTGCATAAACAGGGTTCCTGASCTCACYGAAAAGA  
 CTAWTGTGCCTTAGGAAAAGCATTTG  
 CTGTGTTTYGTTARCACTGGYTCCAGGACAGACCYTCARCTTCCMAAKWGGATACTGCTGCCAAGAAGTTGCTC  
 TGAAGTCAGTTTCTATCRTTCTGCT  
 CTTTGATTCAAAGCACTGTTTCTCTCACTGGGCCTCCAACCATGTTCCCTTCTTCTTAGCACCACAAATAATCAAAA  
 CCCAACATR

Protein Sequence (SEQ ID NO:33):

MMVLQVSAAPRTVALTALLMVLLTSVVQGRATPENYLFQGRQECYAFNGTQRFLERYIYNREEFARFDSVDVGEFRAV  
 TELGRPAAEYWNSQKDILEEKRA  
 VPDRMCRHNYELGGPMTLQRRGWSAVAPSWLTAASTSWAQVILLPQPPM

SNP Information

Context (SEQ ID NO:54):

CGCTTCCTGGAGAGATACATCTACAACCGGGAGGAGTTCGCGCGCTTCGACAGCGACGTGGGGGAGTTCCGGG  
 CGGTGACGGAGCTGGGGCGGCCTGCTG

M

GGAGTACTGGAACAGCCAGAAGGACATCCTGGAGGAGAAGCGGGCAGTGCCGGACAGGATGTGCAGACACAA  
 CTACGAGCTGGGCGGGCCCATGACCCTG

Celera SNP ID: hCV8851074

SNP Position Transcript: 323

SNP Source: Applera

Population(AAllele,Count): african american(A,17|C,15) caucasian(A,5|C,9) total(A,22|C,24)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:33, 7, (E,GAG) (A,GCG)

SNP Source: HGBASE;dbSNP  
 Population(A,Count): no\_pop(A,-|C,-) ;no\_pop(A,-|C,-)  
 SNP Type: MISSENSE MUTATION  
 Protein Coding: SEQ ID NO:33, 7, (E,GAG) (A,GCG)

Gene Number: 45  
 Celera Gene: hCG17514 - 146000219707780  
 Celera Transcript: hCT2283130 - 146000219707791  
 Public Transcript Accession: NM\_002121  
 Celera Protein: hCP1854403 - 197000069367537  
 Public Protein Accession: NP\_002112  
 Gene Symbol: HLA-DPB1  
 Protein Name: major histocompatibility complex, class II, DP beta 1;HLA-DP1B  
 Celera Genomic Axis: GA\_x5YUV32W6W6(6056837..6073197)  
 Chromosome: Chr6  
 OMIM number: 142858  
 OMIM Information: MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS II, DP BETA-1;HLA-DPB1

Transcript Sequence (SEQ ID NO:13):

CAGGAGCTCCCTTTAGCGAGTCCTTCTTTTCCTGACTGCAGCTCTTTYCATTTTGCCATCCTTTTCCAGCTCCATGAT  
 GGTCTGCGAGTTTCTGCGGCC  
 CCCCCGACAGTGGCTCTGACGGCGTTACTGATGGTGCTGCTCACATCTGTGGTCCAGGGCAGGGCCACTCCAGK  
 KACGGCAGGAATGCTACGCGTTTAAT  
 GGGACACAGCGCTTCCTGGAGAGATACATCTACAACCGGSAGGAGYWCGYGCGCTTCGACAGCGACGTGGGRG  
 AGTTCCGGGCGGTGACGGAGCTGGGGC  
 GGCCTGMTGMGGASTACTGGAACAGCCAGAAGGACMTCTGAGGAGRRGCGGGCAGTGCCGGACAGGRTR  
 TGCAGACACAACACTACGAGCTGGDCGRGSC  
 SRTGACCCTGCAGCGCCGAGTCCAGCCTARGGTGAAYGTTTCCCCCTCCAAGAAGGRGCCCYTGCAGCACCAAA  
 CCTGCTTGCTRCCACGTGACRGAT  
 TTCTACCCAGRCAGCATTCAAGTCCGATGSTTCCTGAATGGACAGGAGGAAACAGCTGGGGTCGTGTCCACCAA  
 CCTGATCCGTAATGGAGACTGGACCT  
 TCCAGATCCTGGWGATGCTGGAAATGACCCCCCAGCAGGGAGAYGTCTACAYCTGCCAAGTGGAGCACACCAG  
 CMTGGAYAGTCCTGTACCGTGGAGTG  
 GAAGGCACAGTCTGATTCTGCCCGGAGTAAGACATTGACGGGAGCTGGGGGCTTCGTGCTGGGGCTCATCATCT  
 GTGGAGTGGGCATCTTCATGCACAGG  
 AGGAGCAAGAAAGGTGAGAAAGCCTGCAGGGTGAGCGGGACTTACCTTCCCCTGGCATATTCACACTTATTCCA  
 CGATGAGGGGTYWGACAGAAAAGAAA  
 TGTCAGAAAGCTCTAGAGGCCACTGAWATCAGATARTCGGGGAACAAACATGACCTATRGCGAGAGDGGGATC  
 CCAGGCTGGGATCTTAATGCAGCCAGA  
 TGCATGAGGTCCCAAGTACTCAGGCTCCTGCGGAGCGTCCATTGAGTGATGGGCAATGGAATTTGGTGGGATG  
 GAAATGTTTCTCTAATTATCTGAGGTG  
 GTTTCATGGCTGATTATATAACCTTTCGTCTTTCATTTCAAGTTCAACGAGGATCTGCATAAACAGGTAATATTCCT  
 GCTTTGATTTCTTGTGGGGTGG

GTTGCAGGAGGATATGAGTCCTTTCTGTGCATTGTAACACTGAGGCTCCTCCAGGAAGGGAATCTCAGGCATGA  
 ACCCCTCTTTCAATGTCAGCCTTCAG  
 GCAAGTGGGGAAAGAGCATTGCTTGGCTCCATTGCTGAAGGAAGCAGAGATCAACTCTGTTATTTATCAGCCTG  
 AGACGCATCCTCTCACCATAATTTTT  
 CTCTCCTGGACTTACAGGAAGGAGGCTGGCAACCTGGGATAACTTGTCTTTTACCCCCACAGGGTTCCTGASCTC  
 ACYGAAAAGACTAWTGTGCCTTAGG  
 AAAAGCATTTGCTGTGTTTYGTTARCACTGGYTCCAGGACAGACCYTCARCTTCCMAAKWGGATACTGCTGCCA  
 AGAAGTTGCTCTGAAGTCAGTTTCT  
 ATCRTTCTGCTCTTTGATTCAAAGCACTGTTTCTCTCACTGGGCCTCCAACCATGTTCCCTTCTTCTTAGCACCACA  
 AATAATCAAAACCCAACATR

Protein Sequence (SEQ ID NO:34):

MCRHNYELGGPMTLQRRVQPRNVSPSKKGPLQHNNLLVCHVTDFYPGSIQVRWFLNGQEETAGVVSTNLIRNGD  
 WTFQILVMLEMTTPQQGDVYTCQVEH  
 TSLDSPVTVEWKAQSDSARSKTLTGAGGFVLGLIICGVGIFMHRRSKKGEKACRVSGTYLPLAYSHLFHDEGFDRKEM  
 SESSRGH

SNP Information

Context (SEQ ID NO:55):

CGCTTCCTGGAGAGATACATCTACAACCGGGAGGAGTTCGCGCGCTTCGACAGCGACGTGGGGGAGTTCCGGG  
 CGGTGACGGAGCTGGGGCGGCCTGCTG

M

GGAGTACTGGAACAGCCAGAAGGACATCCTGGAGGAGAAGCGGGCAGTGCCGGACAGGATGTGCAGACACAA  
 CTACGAGCTGGGCGGGCCCATGACCCTG

Celera SNP ID: hCV8851074

SNP Position Transcript: 310

SNP Source: Applera

Population(AAllele,Count): african american(A,17|C,15) caucasian(A,5|C,9) total(A,22|C,24)

SNP Type: UTR 5

Protein Coding: SEQ ID NO:34, None

SNP Source: HGBASE;dbSNP

Population(AAllele,Count): no\_pop(A,-|C,-) ;no\_pop(A,-|C,-)

SNP Type: UTR 5

Protein Coding: SEQ ID NO:34, None

Gene Number: 45

Celera Gene: hCG17514 - 146000219707780

Celera Transcript: hCT1967010 - 146000219707810

Public Transcript Accession: NM\_002121

Celera Protein: hCP1779913 - 197000069367539

Public Protein Accession: NP\_002112

Gene Symbol: HLA-DPB1  
Protein Name: major histocompatibility complex, class II, DP beta 1;HLA-DP1B  
Celera Genomic Axis: GA\_x5YUV32W6W6(6056841..6073197)  
Chromosome: Chr6  
OMIM number: 142858  
OMIM Information: MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS II, DP BETA-1;HLA-DPB1

Transcript Sequence (SEQ ID NO:14):  
AGCTCCCTTTAGCGAGTCCTTCTTTTCCTGACTGCAGCTCTTTYCATTTTGCCATCCTTTTCCAGCTCCATGATGGTT  
CTGCAGGTTTCTGCGGCCCCC  
GGACAGTGGCTCTGACGGCGTTACTGATGGTGCTGCTCACATCTGTGGTCCAGGGCAGGGCCACTCCAGAGAAT  
TACSTKTWCCAGKKACGGCAGGAATG  
CTACGCGTTTAATGGGACACAGCGCTTCCTGGAGAGATACATCTACAACCGGSAGGAGYWCGYGCGCTTCGACA  
GCGACGTGGGRGAGTTCCGGGCGGTG  
ACGGAGCTGGGGCGGCCTGMTGMGGASTACTGGAACAGCCAGAAGGACMTCCTGGAGGAGRRGCGGGCAGT  
GCCGGACAGGRTRTGCAGACACAACACTACG  
AGCTGGDCGRGSCSRTGACCCTGCAGCGCCGAGAGGCACAGTCTGATTCTGCCCGGAGTAAGACATTGACGGG  
AGCTGGGGGCTTCGTGCTGGGGCTCAT  
CATCTGTGGAGTGGGCATCTTCATGCACAGGAGGAGCAAGAAAGTTCAACGAGGATCTGCATAAACAGGGTTC  
CTGASCTCACYGAAAAGACTAWTGTGC  
CTTAGGAAAKWGGATACTGCTGCCAAGAAGTTGCTCTGAAGTCAGTTTCTATCRTTCTGCTCTTTGATTCAAAG  
CACTGTTTCTCTCACTGGGCCTCCA  
ACCATGTTCCCTTCTTCTTAGCACCACAAATAATCAAAACCCAACATR

Protein Sequence (SEQ ID NO:35):  
MMVLQVSAAPRTVALTALLMVLLTSVVQGRATPENYLFQGRQECYAFNGTQRFLERYIYNREEFARFDSVDVGEFRAV  
TELGRPAAEYWNSQKDILEEKRA  
VPDRMCRHNYELGGPMTLQRREAQSDSARSKTLTGAGGFVLGLICGVGIFMHRRSKKVQRGSA

SNP Information

Context (SEQ ID NO:56):  
CGCTTCCTGGAGAGATACATCTACAACCGGGAGGAGTTCGCGCGCTTCGACAGCGACGTGGGGGAGTTCCGGG  
CGGTGACGGAGCTGGGGCGGCCTGCTG  
M  
GGAGTACTGGAACAGCCAGAAGGACATCCTGGAGGAGAAGCGGGCAGTGCCGGACAGGATGTGCAGACACAA  
CTACGAGCTGGGCGGGCCCATGACCCTG  
Celera SNP ID: hCV8851074  
SNP Position Transcript: 323  
SNP Source: Applera  
Population(Allele,Count): african american(A,17|C,15) caucasian(A,5|C,9) total(A,22|C,24)  
SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:35, 7, (E,GAG) (A,GCG)

SNP Source: HGBASE;dbSNP

Population(AAllele,Count): no\_pop(A,-|C,-) ;no\_pop(A,-|C,-)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:35, 7, (E,GAG) (A,GCG)

Gene Number: 45

Celera Gene: hCG17514 - 146000219707780

Celera Transcript: hCT1967011 - 146000219707820

Public Transcript Accession: NM\_002121

Celera Protein: hCP1779938 - 197000069367540

Public Protein Accession: NP\_002112

Gene Symbol: HLA-DPB1

Protein Name: major histocompatibility complex, class II, DP beta 1;HLA-DP1B

Celera Genomic Axis: GA\_x5YUV32W6W6(6059943..6073578)

Chromosome: Chr6

OMIM number: 142858

OMIM Information: MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS II, DP BETA-1;HLA-DPB1

Transcript Sequence (SEQ ID NO:15):

ATGGTGCCCTTGAGCCCAGCCCTACCCCATCTCCACTATCCTCTGCCACCAGCTGTGCAACTTCTGCWAGGGGTG  
 AGGTTAATAAACTGGAGAAAGAATT  
 ACSTKTWCCAGKKACGGCAGGAATGCTACGCGTTTAATGGGACACAGCGCTTCCTGGAGAGATACATCTACAAC  
 CGGSAGGAGYWCGYGCGCTTCGACAG  
 CGACGTGGGRGAGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTGMTGMGGASTACTGGAACAGCCAGAAGG  
 ACMTCTTGAGGAGRRGCGGGCAGTGCCG  
 GACAGGRTRTGCAGACACAACACTACGAGCTGGDCGRGSCSRTGACCCTGCAGCGCCGAGTCCAGCCTARGGTGAA  
 YGTTTCCCCCTCCAAGAAGGRGCCCY  
 TGCAGCACCACAACCTGCTTGTCTRCCACGTGACRGATTTCTACCCAGRCAGCATTCAAGTCCGATGSTTCCTGAA  
 TGGACAGGAGGAAACAGCTGGGGT  
 CGTGTCCACCAACCTGATCCGTAATGGAGACTGGACCTTCAGATCCTGGWGATGCTGGAAATGACCCCCCAGC  
 AGGGAGAYGTCTACAYCTGCCAAGTG  
 GAGCACACCAGCMTGGAYAGTCCTGTACCGTGGAGTGGAAGGCACAGTCTGATTCTGCCCCGAGTAAGACAT  
 TGACGGGAGCTGGGGGCTTCGTGCTGG  
 GGCTCATCATCTGTGGAGTGGGCATCTTCATGCACAGGAGGAGCAAGAAAGTTCAACGAGGATCTGCATAAACA  
 GGGTTCCTGASCTCACYGAAAAGACT  
 AWTGTGCCTTAGGAAAAGCATTTGCTGTGTTTYGTTARCACTGGYTCCAGGACAGACCYTCARCTTCCMAAKW  
 GGATACTGCTGCCAAGAAGTTGCTCT  
 GAAGTCAGTTTCTATCRTTCTGCTCTTTGATTCAAAGCACTGTTTCTCTCACTGGGCCTCCAACCATGTTCCCTTCTT  
 CTTAGCACCACAAATAATCAAA  
 ACCCAACATRASTGTTTGYTTTCCTTTAAAAATATGCAYCAAATCRTCTCTCATYACTTTTCTCTGAGGGTTTTAGTA  
 GACAGTAGGAGTTAATAAAGAA

GTTCAATTTTGGTTTAAACATAGGAAAGAAGAGAACCATGAAAATGGGGATATGTAACTATTGTATAATGGGGC  
 CTGTTACACATGACACTCTTCTGAAT  
 TGACTGTATTTTCAGTGAGCTGCCCCCAAATCAAGTTTAGTGCCCTCATCCATTTATGTCTCAGACCACTATTCTTAA  
 CTATTCAATGGTGAGCAGACTGC  
 AAATCTGCCTGATAGGACCCATATTCCCACAGCACTAATTCAACATATACCTTACTGAGAGCATGTTTTATCATT  
 CCATTAACAAGTTAA

Protein Sequence (SEQ ID NO:36):

MVPLSPALPHLHYPLPPAVQLLLGVRLINWRKNYLFQGRQECYAFNGTQRFLERYIYNREEFARFDSVDVGEFRAVTEL  
 GRPAAEYWNSQKDILEEKRAVP  
 DRMCRHNYELGGPMTLQRRVQPRVNVSPSKKGPLQHNNLLVCHVTDFYPGSIQVRWFLNGQEETAGVVSTNLIRN  
 GDWTFQILVMLEMTPQQGDVYTCQV  
 EHTSLDSPVTVEWKAQSDSARSKTLTGAGGFVLGLICGVGIFMHRRSKKVQRGSA

SNP Information

Context (SEQ ID NO:57):

CGCTTCCTGGAGAGATACATCTACAACCGGGAGGAGTTCGCGCGCTTCGACAGCGACGTGGGGGAGTTCCGGG  
 CGGTGACGGAGCTGGGGCGGCCTGCTG  
 M

GGAGTACTGGAACAGCCAGAAGGACATCCTGGAGGAGAAGCGGGCAGTGCCGGACAGGATGTGCAGACACAA  
 CTACGAGCTGGGCGGGCCCATGACCCTG

Celera SNP ID: hCV8851074

SNP Position Transcript: 248

SNP Source: Applera

Population(AAllele,Count): african american(A,17|C,15) caucasian(A,5|C,9) total(A,22|C,24)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:36, 6, (E,GAG) (A,GCG)

SNP Source: HGBASE;dbSNP

Population(AAllele,Count): no\_pop(A,-|C,-) ;no\_pop(A,-|C,-)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:36, 6, (E,GAG) (A,GCG)

Gene Number: 52

Celera Gene: hCG1779944 - 209000105681216

Celera Transcript: hCT1818766 - 209000105681228

Public Transcript Accession: NM\_018256

Celera Protein: hCP1728848 - 209000105681203

Public Protein Accession: NP\_060726

Gene Symbol: WDR12

Protein Name: WD repeat domain 12;FLJ10881;FLJ12719;FLJ12720;YTM1

Celera Genomic Axis: GA\_x5YUV32VWPT(8412535..8444406)

Chromosome: Chr2

OMIM number:

## OMIM Information:

Transcript Sequence (SEQ ID NO:16):

GGGGGTTGGGAGGTAGCGTCCGGTAGGTCTGCGCAGACAGCGCTGCCAGACCTCTGACTCCCCCAGTCCTGCC  
 GCCGCCGATAGCCCCACAAGGCGACAT  
 GGGAAGTCCGCCGGAGCAGAGGGGAGAAAATTCAAAACGAGTCTACTCAACCTCGTCTTGGCTCCGAGGCCCC  
 GCCATGGAGACCAGCATAATGCGCAGG  
 CGCGCCCCACCGGGAGCGCGCAGAGCAGTCCCACGACAGTAGGACCACTATCGGCCACCGTCGACCGTTTCCTC  
 TCTCATCCGTTCTTTTCTTTGTATTT  
 CCGCCTCTCGCCTCTCTCTAAAAGCCGCGAGTTAGAGGCGAGATTTAGGAAAAACCTCTGCCGAGTGAGCCTCTGG  
 TTGGGAATATGTATGAGAAAAAAA  
 ACTGGCAAGGCGTTAGTCAAGCAAAGCTGAAGGCAGAGGAAATTTGATATCTGGCTGGAGTCTAGAGGATTTA  
 ATGCAAATAAGATACTCTGAGGGCAGC  
 GTGGCAAAAAAAGACTACAATTCCCGGTGGTCACAGCGTTTGAGAAGCGATGCTTTCTGAGACTTGTAGTAACT  
 AGGAGCTGTGTTTGAAGTATCCAGGC  
 TCAGGACAGCCTCTTGAAAAAAATTTTTTATTAATAAAGCGGATTTGAGTGGGATCTTTTCTMATCGATTAC  
 GGGCCACACGTATGGGAAGAATTC  
 TAACAATGATTAAAGGGACATGCTACCTTTACGACTATCCTTTTCTAATCGATGACTCCTAAATCTAGGAGTAGGT  
 AGTCGATGTTTGTGGTCTGGGCGT  
 CTGTAGAAGGGCAACCTCGTGCTTTCTGCAGAGGAGACCGGAGGGCAGAAGGCAGAGTCCAGGCTTAGACTGC  
 AGTTCCTCGCTTACCTGTGCAGTCTAA  
 TTTTGAGCTGCCTCTTTGTAGTCTTAAAAGGCAGGAGCTTCGTGTWGTGGGTCTGCTAACCCGTACGTTTCCGTG  
 GGCAAGTCGTGTGTACTCCTCGCCA  
 TGGCTCAGCTCCAAACACGCTTCTACACTGATAACAAGAAATATGCCGTAGATGATGTTCCCTTCTCAATCCCTGC  
 TGCCTCTGAAATTGCCGACCTTAG  
 TAACATCATCAATAAACTACTAAAGGACAAAAATGAGTTCCACAAACATGTGGAGTTTGATTTCCTTATTAAGGG  
 CCAGTTTCTGCGAATGCCCTTGGAC  
 AAACACATGGAAATGGAGAACRTCTCATCAGAAGAAGTTGTGGAAATAGAATACGTGGAGAAGTATACTGCAC  
 CCCAGCCAGAGCAATGCATGTTCCATG  
 ATGACTGGATCAGTTCAATTAAAGGGGCGAGAGGAATGGATCTTGACTGGTTCTTATGATAAGACTTCTCGGATCT  
 GGTCTTGGAAGGAAAGTCAATAAT  
 GACAATTGTGGGACATACGGATGTTGTAAAAGATGTGGCCTGGGTGAAAAAAGATAGTTTGTCTGCTTATTAT  
 TGAGTGCTTCTATGGATCAGACTATT  
 CTCTTATGGGAGTGGAATGTAGAGMGAAACAAAGTGAAAGCCCTACACTGCTGTAGAGGTCATGCTGGAAGTG  
 TAGATTCTATAGCTGTTGATGGCTCAG  
 GAACTAAATTTTGCAGTGGCTCCTGGGATAAGATGCTAAAGATCTGGTCTACAGTCCCTACAGATGAAGAAGAT  
 GAAATGGAGGAGTCCACAAATCGACC  
 AAGAAAGAAACAGAAGACAGAACAGTTGGGACTAACAAGGACTCCCATAGTGACCCTCTCTGGCCACATGGAG  
 GCAGTTTCCTCAGTTCTGTGGTCAGAT  
 GCTGAAGAAATCTGCAGTGCATCTTGGGACCATAACAATTAGAGTGTGGGATGTTGAGTCTGGCAGTCTTAAGTC  
 AACTTTGACAGGAAATAAAGTGTTTA  
 ATTGTATTTCTATTCTCCACTTTGTAAACGTTTAGCATCTGGAAGCACAGATAGGCATATCAGACTGTGGGATCC  
 CCGAACTAAAGATGGTTCTTTGGT  
 GTCGCTGTCCCTAACGTACATACTGGTTGGGTGACATCAGTAAATGGTCTCCTACCCATGAACAGCAGCTGAT  
 TTCAGGATCTTTAGATAACATTGTT

AAGCTGTGGGATACAAGAAGTTGTAAGGCTCCTCTCTATGATCTGGCTGCTCATGAAGACAAAGTTCTGAGTGT  
 AGACTGGACAGACACAGGGCTACTTC  
 TGAGTGGAGGAGCAGACAATAAATTGTATTCCTACAGATATTCACCTACCACTTCCCATGTTGGGGCATGAAAGT  
 GAACAATAATTTGACTATAGAGATT  
 ATTTCTGTAAATGAAATTGGTAGAGAACCATGAAATTACATAGATGCAGATGCAGAAAGCAGCCTTTTGAAGTTT  
 ATATAATGTTTTTACCCTTCATAAC  
 AGCTAACGTATCACTTTTTCTTATTTTGTATTTATAATAAGATAGGTTGTGTTTATAAAATACAACTGTGGCATAC  
 ATTCTCTATACAACTTGAAATT  
 AAACTGAGTTTTACATTTCTCTTTAAAGGTATTG

Protein Sequence (SEQ ID NO:37):

MAQLQTRFYTDNKKYAVDDVPFSIPAASEIADLSNIINKLLKDKNEFHKHVEFDLIKQQLRMPDKHMEMENISSEE  
 VVEIEYVEKYTAPQPEQCMFH  
 DDWISSIKGAEWILTGSYDKTSRIWSLEGKSIMTIVGHTDVVKDVAWVKKDSLSCLLLSASMDQTILLWEWNVERN  
 KVKALHCCRGHAGSVDSIAVDGS  
 GTKFCSGSWDKMLKIWSTVPTDEEDEMEESTNRPRKKQKTEQLGLTRTPIVTLSGHMEAVSSVLWSDAEEICSASW  
 DHTIRVWDVESGSLKSTLTGNKVF  
 NCISYSPICKRLASGSTDRHIRLWDPRTKDGLVSLSLTSHTGWVTSVKWSPTHEQQLISGSLDNIVKLWDTRSCAPL  
 YDLAAHEDKVLSVDWTDGTL  
 LSGGADNKLYSYRYSPTTSHVGA

SNP Information

Context (SEQ ID NO:58):

AAAGGACAAAAATGAGTTCCACAAACATGTGGAGTTTGATTTCCTTATTAAGGGCCAGTTTCTGCGAATGCCCTT  
 GGACAAACACATGGAAATGGAGAAC

R

TCTCATCAGAAGAAGTTGTGGAAATAGAATACGTGGAGAAGTATACTGCACCCCAGCCAGAGCAATGCATGTTC  
 CATGATGACTGGATCAGTTCAATTAA

Celera SNP ID: hCV25653599

SNP Position Transcript: 1222

SNP Source: Applera

Population(A,Count): african american(A,37|G,1) caucasian(A,39|G,1) total(A,76|G,2)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:37, 5, (I,ATC) (V,GTC)

Gene Number: 52

Celera Gene: hCG1779944

Celera Transcript: CRA|161000002536323

Public Transcript Accession: NM\_018256

Celera Protein: CRA|224000006551765

Public Protein Accession: NP\_060726

Gene Symbol: WDR12

Protein Name: WD repeat domain 12;FLJ10881;FLJ12719;FLJ12720;YTM1

Celera Genomic Axis: GA\_x5YUV32VWPT(8412535..8444406)

Chromosome: Chr2  
OMIM number:  
OMIM Information:

## Transcript Sequence (SEQ ID NO:17):

AGAAGGGCAACCTCGTGCTTTCTGCAGAGGAGACCGGAGGGCAGAAGGCAGAGTCCAGGCTTAGACTGCAGTT  
CCTCGCTTACCTGTGCAGTCTAATTTT  
GAGCTGCCTCTTTGTAGTCTTAAAAGGCAGGAGCTTCGTGTWGTGGGTCTGCTAACCCGTACGTTTCCGTGGGC  
AAGTCGTGTGTACTCCTCGCCATGGC  
TCAGCTCCAAACACGCTTCTACACTGATAACAAGAAATATGCCGTAGATGATGTTCCCTTCTCAATCCCTGCTGCC  
TCTGAAATTGCCGACCTTAGTAAC  
ATCATCAATAAACTACTAAAGGACAAAAATGAGTTCCACAAACATGTGGAGTTTGATTTCCTTATTAAGGGCCAG  
TTTCTGCGAATGCCCTTGGACAAAC  
ACATGGAAGTGGAGAACRTCTCATCAGAAGAAGTTGTGGAAATAGAATACGTGGAGAAGTGTACCGCACCCCA  
GCCAGAGCAATGCATGTTCCATGATGA  
CTGGATCAGTTCAATTAAAGGGGCAGAGGAATGGATCTTGACTGGTTCTTATGATAAGACTTCTCGGATCTGGTC  
CTTGGAAGGAAAGTCAATAATGACA  
ATTGTGGGACATACGGATGTTGTAAAAGATGTGGCCTGGGTGAAAAAAGATAGTTTGCCTGCTTATTATTGAG  
TGCTTCTATGGATCAGACTATTCTCT  
TATGGGAGTGGAATGTAGAGMGAAACAAAGTGAAAGCCCTACACTGCTGTAGAGGTCATGCTGGAAGTGTAG  
ATTCTATAGCTGTTGATGGCTCAGGAAC  
TAAATTTTGCAGTGGCTCCTGGGATAAGATGCTAAAGATCTGGTCTACAGTCCCTACAGATGAAGAAGATGAAA  
TGGAGGAGTCCACAAATCGACCAAGA  
AAGAAACAGAAGACAGAACAGTTGGGACTAACAAGGACTCCCATAGTGACCCTCTCTGGCCACATGGAGGCAG  
TTTCCTCAGTTCTGTGGTCAGATGCTG  
AAGAAATCTGCAGTGCATCTTGGGACCATAACAATTAGAGTGTGGGATGTTGGGTCTGGCAGTCTTAAGTCAACT  
TTGACAGGAAATAAAGTGTTTAATTG  
TATTTCTATTCTCCACTTTGTAAACGTTTAGCATCTGGAAGCACAGATAGGCATATCAGACTGTGGGATCCCCGA  
ACTAAAGATGGTTCTTTGGTGTCG  
CTGTCCCTAACGTCACATACTGGTTGGGTGACATCAGTAAAATGGTCTCCTACCCATGAACAGCAGCTGATTTCA  
GGATCTTTAGATAACATTGTTAAGC  
TGTGGGATACAAGAAGTTGTAAGGCTCCTCTCTATGATCTGGCTGCTCATGAAGACAAAGTTCTGAGTGTAGACT  
GGACAGACACAGGGCTACTTCTGAG  
TGGAGGAGCAGACAATAAATTGTATTCCTACAGATATTCACCTACCACTTCCCATGTTGGGGCATGAAAGTGAAC  
AATAATTTGACTATAGAGATTATTT  
CTGTAAATGAAATTGGTAGAGAACCATGAAATTACATAGATGCAGATGCAGAAAGCAGCCTTTTGAAGTTTATA  
TAATGTTTTTACCCTTCATAACAGCT  
AACGTATCACTTTTTCTTATTTTGTATTTATAATAAGATAGGTTGTGTTTATGAAATACAAACTGTGGCATAACATTC  
TCTATACAAACTTGAAATTAAC  
TGAGTTTTACATTTCTCTTT

## Protein Sequence (SEQ ID NO:38):

MAQLQTRFYTDNKKYAVDDVPFSIPAASEIADLSNIINKLLKDKNEFHKHVEFDLIKQQLRMPLDKHMEVENISSEE  
VVEIEYVEKCTAPQPEQCMFH

DDWISSIKGAEEWILTGSYDKTSRIWSLEGKSIMTIVGHTDVVKDVAWVKKDSLSCLLLSASMDQTILLWEWNVERN  
 KVKALHCCRGHAGSVDSIAVDGS  
 GTKFCSGSWDKMLKIWSTVPTDEEDEMEESTNRPRKKQKTEQLGLTRTPIVTLSGHMEAVSSVLWSDAEEICSASW  
 DHTIRVWDVVGSGSLKSTLTGNKVF  
 NCISYSPLCKRLASGSTDRHIRLWDPRTKDGSLVSLSLTSHTGWVTSVKWSPTHEQQLISGSLDNIVKLWDTRSCAPL  
 YDLAAHEDKVLSVDWTDGGLL  
 LSGGADNKLYSYRYSPTTSHVGA

#### SNP Information

Context (SEQ ID NO:59):

AAAGGACAAAATGAGTTCCACAAACATGTGGAGTTTGATTTCCTTATTAAGGGCCAGTTTCTGCGAATGCCCTT  
 GGACAAACACATGGAAGTGGAGAAC  
 R  
 TCTCATCAGAAGAAGTTGTGGAAATAGAATACGTGGAGAAGTGTACCGCACCCCAGCCAGAGCAATGCATGTTC  
 CATGATGACTGGATCAGTTCAATTAA

Celera SNP ID: hCV25653599

SNP Position Transcript: 418

SNP Source: Applera

Population(AAllele,Count): african american(A,37|G,1) caucasian(A,39|G,1) total(A,76|G,2)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:38, 5, (I,ATC) (V,GTC)

Gene Number: 66

Celera Gene: hCG1810785 - 146000220705953

Celera Transcript: hCT1950078 - 146000220705954

Public Transcript Accession: NM\_005577

Celera Protein: hCP1765809 - 197000069410643

Public Protein Accession: NP\_005568

Gene Symbol: LPA

Protein Name: lipoprotein, Lp(a);AK38;LP

Celera Genomic Axis: GA\_x54KRFTF0F9(10032912..10114024)

Chromosome: Chr6

OMIM number: 152200

OMIM Information: APOLIPOPROTEIN(a);LPA

Transcript Sequence (SEQ ID NO:18):

ATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGA  
 GTACTGCAACCTGACGCAATGCTCAG  
 ACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCA  
 CCAACTGAGCAAAGGCCTGGGGTGCA  
 GGAGTGCTACCACGGAAATGGACAGAGTTATCAAGGCACATACTTCATTACTGTCACAGGAAGAACCTGCCAAG  
 CTTGGTCATCTATGACACCACTCG

CATAGTCGGACCCCAGCATACTACCCAAATGCTGGCTTGATCAAGAACTACTGCCGAAATCCAGATCCTGTGGCA  
 GCCCCTTGGTGTTATACAACAGATC  
 CCAGTGTCAGGTGGGAGTACTGCAACCTGACACGATGCTCAGATGCAGAATGGACTGCCTTCGTCCCTCCGAAT  
 GTTATTCTGGCTCCAAGCCTAGAGGC  
 TTTTTTTGAACAAGCACTGACTGAGGAAACCCCCGGGGTACAGGACTGCTACTACCATTATGGACAGAGTTACCG  
 AGGCACATACTCCACCACTGTCACA  
 GGAAGAACTTGCCAAGCTTGGTCATCTATGACACCACACCAGCATAGTCGGACCCCAGAAAACCTACCCAAATGC  
 TGGCCTGASCAGGAACTACTGCAGGA  
 ATCCAGATGCTGAGATTGCCCCTTGGTGTTACACCATGGATCCCAGTGTCAGGTGGGAGTACTGCAACCTGACAC  
 AATGCCTGGTGACAGAATCAAGTGT  
 CCTTGCAACTCTCACGGTGGTCCCAGAYCCAAGCACAGAGGCTTCTTCTGAAGAAGCACCAACGGAGCAAAGCC  
 CCGGGGTCCAGGATTGCTACCATGGT  
 GATGGACAGAGTTATCGAGGCTCATTCTCTACCACTGTCACAGGAAGGACATGTCAGTCTTGGTYCTCTATGACA  
 CCACACTGGCATCAGAGGACAACAG  
 AATATTATCCAAATGGTGGCCTGACCAGGAACTACTGCAGGAATCCAGATGCTGAGATTAGTCCTTGGTGTTATA  
 CCATGGATCCCAATGTCAGATGGGA  
 GTACTGCAACCTGACACAATGTCCAGTGACAGAATCAAGTGTCCTTGCGACGTCCACGGCTGTTTCTGAACAAGC  
 ACCAACGGAGCAAAGCCCCACAGTC  
 CAGGACTGCTACCATGGTGATGGACAGAGTTATCGAGGCTCATTCTCCACCACTGTTACAGGAAGGACATGTCA  
 GTCTTGGTCTCTATGACACCACACT  
 GGCATCAGAGAACCACAGAATACTACCCAAATGGTGGCCTGACCAGGAACTACTGCAGGAATCCAGATGCTGA  
 GATTCGCCCTTGGTGTTATACCATGGA  
 TCCCAGTGTCAGATGGGAGTACTGCAACCTGACRCAATGTCCAGTGATGGAATCAACTSTCCTCACAACCTCCCAC  
 GGTGGTCCCAGTTCCAAGCACAGAG  
 STTCCTTCTGAAGAAGCACCAACTGAAAACAGCACTGGGGTCCAGGACTGCTACCGAGGTGATRGACAGAGTTA  
 TCGAGGCMCACTCTCCACCACTATCA  
 CAGGAAGAACATGTCAGTCTTGGTCGTCTATGACACCACATTGGCATCGGAGGATCCCATTATACTATCCAAATG  
 RTGGCCTGACCAGGAACTACTGCAG  
 GAATCCAGATGCTGAGATTCGCCCTTGGTGTTACACCATGGATCCCAGTGTCAGGTGGGAGTACTGCAACCTGA  
 CACGATGTCCAGTGACAGAATCGAGT  
 GTCCTCACAACCTCCCACAGTGGCCCCGGTTCCAAGCACAGAGGCTCCTTCTGAACAAGCACCACTGAGAAAAG  
 CCCTGTGGTCCAGGATTGCTACCATG  
 GTGATGGAYGGAGTTATCGAGGCATATCCTCCACCACTGTCACAGGAAGGACCTGTCAATCTTGGTCATCTATGA  
 TACCACACTGGYATCAGAGGACCCC  
 AGAAAACCTACCCAAATGCTGGCCTGACCGAGAACTACTGCAGGAATCCAGATTCTGGGAAACAACCCTGGTGTT  
 ACACAACCGATCCGTGTGTGAGGTGG  
 GAGTACTGCAATCTGACACAATGCTCAGAAACAGAATCAGGTGTCCTAGAGACTCCCACTGTTGTTCCAGTTCCA  
 AGCAYGGAGGCTCATTCTGAAGCAG  
 CACCAACTGAGCAAACCCCTGTGGTCCGGCAGTGCTACCATGGTAATGGCCAGAGTTATCGAGGCACATTCTCC  
 ACCACTGTCACAGGAAGGACATGTCA  
 ATCTTGGTCATCCATGACACCACACCGGCATCAGAGGACCCCAGAAAACCTACCCAAATGATGGCCTGACAATGA  
 ACTACTGCAGGAATCCAGATGCCGAT  
 ACAGGCCCTTGGTGTTTTACCAYGGACCCCAGCATCAGGTGGGAGTACTGCAACCTGACGCGATGCTCAGACAC  
 AGAAGGGACTGTGGTCGCTCCTCCGA

CTGTCATCCAGGTTCCAAGCCTAGGGCCTCCTTCTGAACAAGAYTGTATGTTTGGGAATGGGAAAGGATACCGG  
 GGCAAGAARGCAACCACTGTTACTGG  
 GACGCCATGCCAGGAATGGGCTKCCCAGGAGCCCCATAGACACAGCACRTTCATTCCAGGGACAAATAAATGGG  
 CAGGTCTGGAAAAAATTACTGCCGT  
 AACCTGATGGTGACATCAATGGTCCCTGGTGCTACACAATGAATCCAAGAAAACCTTTTGGACTACTGTGATATC  
 CCTCTCTGTGCATCCTCTTCATTTG  
 ATTGTGGGAAGCCTCAAGTGGAGCCGAAGAAATGTCCTGGAAGCATTGTAGSGGKGTGTGTGGCCCACCCACAT  
 TCCTGGCCCTGGCAAGTCAGTCTCAG  
 AACAAGGTTTGGAAAGCACTTCTGTGGAGGCACCTTAATATCCCCAGAGTGGGTGCTGACTGCTGCTCACTGCTT  
 GAAGAAGTCCTCAAGGCCTTCATCC  
 TACAAGGTCATCCTGGGTGCACACCAAGAAGTGAACCTCGAATCTCATGTTCAGGAAATRGAAGTGTCTAGGCT  
 GTTCTTGGAGCCCACACAAGCAGATA  
 TTGCCTTGCTAAAGCTAAGCAGGCCTGCCGTCATCACTGACAAAGTAATGCCAGCTTGTCTGCCATCCCCAGACT  
 ACATGGTCACCGCCAGGACTGAATG  
 TTACATCACTGGCTGGGGAGAAACCCAAGGTACCTTTGGGACTGGCCTTCTCAAGGAAGCCCAGCTCCTTGTTAT  
 TGAGAATGAAGTGTGCAATCACTAT  
 AAGTATATTTGTGCTGAGCATTGCGCCAGAGGCACTGACAGTTGCCAGGGTGACAGTGGAGGGCCTCTGGTTTG  
 CTTGAGAAGGACAAATACATTTTAC  
 AAGGAGTCACTTCTTGGGGTCTTGGCTGTGCAYGCCCCAATAAGCCTGGTGTCTATGCTCGTGTTTCAAGGTTTG  
 TTAATTGATTGAGGGAATGATGAG  
 AAATAATTAATTGGACGGGAGACAGAGTGAAGCATCAACCTACTTAGAAGCTGAAACGTGGGTAAGGATTTAG  
 CATGCTGGAAATAATAGACAGCAATCA  
 AACGAAGACACTGTTCCAGCTACCAGCTATGCCAAACCTTGGCATTGTTTGGTATTTTGTGTATAAGCTTTTAAG  
 GTCTGACTGACAAATTCTGTATTA  
 AGGTGTCATAGCTATGACATTTGTTAAAAATAAACTCTGCACTTATTTTGATTTGA

Protein Sequence (SEQ ID NO:40):

MNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPQGVQECYHGNGQ  
 SYQGTFFITVTGRTCQAWSSMTPHS  
 HSRTPAYYPNAGLIKNYCRNPDPVAAPWCYTDDPSVRWEYCNLTRCSDAEWTAFFPPNVILAPSLEAFFEQALTEETP  
 GVQDCYHYHGQSYRGTYSTTVT  
 GRTCQAWSSMTPHQHSRTPENYPNAGLTRNYCRNPDAEIRPWCYTMDPSVRWEYCNLTQCLVTESSVLATLTVVP  
 DPSTEASSEEAPTEQSPGVQDCYHG  
 DGQSYRGSFSTTVTGRTCQSWSSMTPHWHQRTTEYYPNGGLTRNYCRNPDAEISPWCYTMDPNVRWEYCNLTQC  
 PVTESVLATSTAVSEQAPTEQSPTV  
 QDCYHGDGQSYRGSFSTTVTGRTCQSWSSMTPHWHQRTTEYYPNGGLTRNYCRNPDAEIRPWCYTMDPSVRWEY  
 CNLTQCPVMESTLLTPTVPVPSTE  
 LPSEEAPTENSTGVQDCYRGDGQSYRGTLSTTITGRTCQSWSSMTPHWHRRIPLYYPNAGLTRNYCRNPDAEIRPWC  
 YTMPSVRWEYCNLTRCPVTESS  
 VLTTPTVAPVPSTEAPSEQAPPEKSPVVQDCYHGDGRSYRGISSTTVTGRTCQSWSSMIPHWHQRTPENYPNAGLTE  
 NYCRNPDSGKQPWCYTDDPCVRW  
 EYCNLTQCSETESGVLETPTVPVPVPSMEAHSEAAPEQTPVVRQCYHGNGQSYRGTFSTTVTGRTCQSWSSMTPHR  
 HQRTPENYPNDGLTMNYCRNPDA  
 TGPWCFTMDPSIRWEYCNLTRCSDTEGTVVAPPTVIQVPSLGPPSEQDCMFGNGKGYRGKKATTVTGTPCQEWAA  
 QEPHRHSTFIPGTNKGWAGLEKNYCR

NPDGDINGPWCYTMNPRKLFDYCDIPLCASSSFDCGKPQVEPKKCPGSIVGGCVAHPHSWPWQVSLRTRFGKHFCG  
GTLISPEWVLTAAHCLKKSSRPSS  
YKVILGAHQEVNLESHVQEIEVSRLFLEPTQADIALLKLSRPAVITDKVMPACLPSPDYMVTARTECYITGWGETQGTF  
GTGLLKEAQLLVIENEVCNH  
KYICAEHLARGTDSCQGDSGGPLVCFEKDKYILQGVTSWGLGCARPKNKPGVYARVSRFVTWIEGMMRNN

SNP Information

Context (SEQ ID NO:60):  
CTGCTGCTCACTGCTGAAGAAGTCCTCAAGGCCTTCATCCTACAAGGTCATCCTGGGTGCACACCAAGAAGTGA  
ACCTCGAATCTCATGTTCAAGGAAAT  
R  
GAAGTGTCTAGGCTGTTCTTGGAGCCCCACACAAGCAGATATTGCCTTGCTAAAGCTAAGCAGGCCTGCCGTCATC  
ACTGACAAAGTAATGCCAGCTTGTC  
Celera SNP ID: hCV25930271  
SNP Position Transcript: 3060  
SNP Source: Applera  
Population(AAllele,Count): african american(A,37|G,1) caucasian(A,36|G,2) total(A,73|G,3)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:40, 39, (I,ATA) (M,ATG)

SNP Source: dbSNP  
Population(AAllele,Count): no\_pop(G,-|A,-)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:40, 39, (I,ATA) (M,ATG)

Gene Number: 66  
Celera Gene: hCG1810785  
Celera Transcript: CRA|335000489558673  
Public Transcript Accession: NM\_005577  
Celera Protein: CRA|224000006540879  
Public Protein Accession: NP\_005568  
Gene Symbol: LPA  
Protein Name: lipoprotein, Lp(a);AK38;LP  
Celera Genomic Axis: GA\_x54KRFTF0F9(10032912..10114024)  
Chromosome: Chr6  
OMIM number: 152200  
OMIM Information: APOLIPOPROTEIN(a);LPA

Transcript Sequence (SEQ ID NO:19):  
CTGGGATTGGGACACACTTTCTGGACACTGCTGGCCAGTCCCAAATGGAACATAAGGAAGTGGTTCTTCTACTT  
CTTTTATTTCTGAAATCAGCAGCAC  
CTGAGCAAAGCCATGTGGTCCAGGATTGCTACCATGGTGATGGACAGAGTTATCGAGGCACGTACTCCACCACT  
GTCACAGGAAGGACCTGCCAAGCTTG

GTCATCTATGACACCACATCAACATAATAGGACCACAGAAAACCTACCCAAATGCTGGCTTGATCATGAACTACTG  
 CAGGAATCCAGATGCTGTGGCAGCT  
 CCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGG  
 GACTGCCGTCGCGCCTCCGACTGTTA  
 CCCCAGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTAC  
 CATGGTAATGGACAGAGTTATCGAGG  
 CACATACTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGAC  
 CCCAGAATACTACCCAAATGCTGGC  
 TTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGG  
 TGGGAGTACTGCAACCTGACGCAAT  
 GCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAA  
 CAAGCACCGACTGAGCAAAGGCCTGG  
 GGTGCAGGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACAGGAAGAACCT  
 GCCAAGCTTGGTCATCTATGACACCA  
 CACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCT  
 GTGGCAGCTCCTTATTGTTATACGA  
 GGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCC  
 TCCGACTGTTACCCCGGTTCCAAGCCT  
 AGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGGACAG  
 AGTTATCGAGGCACATACTCCACCACT  
 GTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCA  
 AATGCTGGCTTGATCATGAACTACT  
 GCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAAC  
 CTGACGCAATGCTCAGACGCAGAAGG  
 GACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCA  
 AAGGCCTGGGGTGCAGGAGTGCTAC  
 CATGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATC  
 TATGACACCACACTCGCATAGTCGGA  
 CCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATT  
 GTTATACGAGGGATCCCGGTGTCAG  
 GTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGG  
 TTCCAAGCCTAGAGGCTCCTTCCGAA  
 CAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGCACAT  
 ACTCCACCACTGTCACAGGAAGAACCT  
 GCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCA  
 TGAATACTGCAGGAATCCAGATGC  
 TGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAG  
 ACGCAGAAGGGACTGCCGTCGCGCCT  
 CCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAG  
 GGAGTGCTACCATGGTAATGGACAGA  
 GTTATCGAGGCACATACTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGC  
 ATAGTCGGACCCCAGAATACTACCC  
 AAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCC  
 CGGTGTCAGGTGGGAGTACTGCAAC

CTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGC  
 TCCTTCCGAACAAGCACCGACTGAGC  
 AAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACA  
 GGAAGAACCTGCCAAGCTTGGTCATC  
 TATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAA  
 TCCAGATGCTGTGGCAGCTCCTTAT  
 TGTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTG  
 CCGTCGCGCCTCCGACTGTTACCCCGG  
 TTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGT  
 AATGGACAGAGTTATCGAGGCACATA  
 CTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGA  
 AATACTACCCAAATGCTGGCTTGATC  
 ATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGA  
 GTACTGCAACCTGACGCAATGCTCAG  
 ACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCA  
 CCGACTGAGCAAAGGCCTGGGGTGCA  
 GGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACAGGAAGAACCTGCCAAG  
 CTTGGTCATCTATGACACCACACTCG  
 CATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGC  
 AGCTCCTTATTGTTATACGAGGGATC  
 CCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGAC  
 TGTTACCCCGGTTCCAAGCCTAGAGGC  
 TCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGGACAGAGTTATC  
 GAGGCACATACTCCACCACTGTCACA  
 GGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCT  
 GGCTTGATCATGAACTACTGCAGGA  
 ATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACG  
 CAATGCTCAGACGCAGAAGGGACTGC  
 CGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGC  
 CTGGGGTGCAGGAGTGCTACCATGGT  
 AATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGAC  
 ACCACACTCGCATAGTCGGACCCCAG  
 AATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATA  
 CGAGGGATCCCGGTGTCAGGTGGGA  
 GTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAA  
 GCCTAGAGGCTCCTTCCGAACAAGCA  
 CCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGCACATACTCCA  
 CCACTGTCACAGGAAGAACCTGCCAAG  
 CTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACT  
 ACTGCAGGAATCCAGATGCTGTGGC  
 AGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAG  
 AAGGGACTGCCGTCGCGCCTCCGACT  
 GTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGT  
 GCTACCATGGTAATGGACAGAGTTATC

GAGGCACATACTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTC  
GGACCCCAGAATACTACCCAAATGC  
TGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGT  
CAGGTGGGAGTACTGCAACCTGACG  
CAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCC  
GAACAAGCACCGACTGAGCAAAGGC  
CTGGGGTGCAGGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACAGGAAG  
AACCTGCCAAGCTTGGTCATCTATGAC  
ACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGA  
TGCTGTGGCAGCTCCTTATTGTTAT  
ACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCG  
CGCCTCCGACTGTTACCCCGGTTCCAA  
GCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGG  
ACAGAGTTATCGAGGCACATACTCCAC  
CACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTA  
CCCAAATGCTGGCTTGATCATGAAC  
TACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTG  
CAACCTGACGCAATGCTCAGACGCAG  
AAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACT  
GAGCAAAGGCCTGGGGTGCAGGAGTG  
CTACCATGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGT  
CATCTATGACACCACACTCGCATAGT  
CGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCT  
TATTGTTATACGAGGGATCCCGGTG  
TCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACC  
CCGGTTCCAAGCCTAGAGGCTCCTTC  
CGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGC  
ACATACTCCACCACTGTCACAGGAAGA  
ACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTG  
ATCATGAACTACTGCAGGAATCCAG  
ATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGC  
TCAGACGCAGAAGGGACTGCCGTCGC  
GCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGG  
TGCAGGAGTGCTACCATGGTAATGGA  
CAGAGTTATCGAGGCACATACTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACAC  
TCGCATAGTCGGACCCCAGAATACT  
ACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGG  
ATCCCGGTGTCAGGTGGGAGTACTG  
CAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAG  
AGGCTCCTTCCGAACAAGCACCGACT  
GAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTG  
TCACAGGAAGAACCTGCCAAGCTTGGT  
CATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCA  
GGAATCCAGATGCTGTGGCAGCTCC

TTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGG  
 ACTGCCGTCGCGCCTCCGACTGTTACC  
 CCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACC  
 ATGGTAATGGACAGAGTTATCGAGGCA  
 CATACTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCC  
 CAGAATACTACCCAAATGCTGGCTT  
 GATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGT  
 GGGAGTACTGCAACCTGACGCAATGC  
 TCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACA  
 AGCACCGACTGAGCAAAGGCCTGGGG  
 TGCAGGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACAGGAAGAACCTGC  
 CAAGCTTGGTCATCTATGACACCACA  
 CTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGT  
 GGCAGCTCCTTATTGTTATACGAGG  
 GATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTC  
 CGACTGTTACCCCGGTTCCAAGCCTAG  
 AGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGGACAGAG  
 TTATCGAGGCACATACTCCACCACTGT  
 CACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAA  
 TGCTGGCTTGATCATGAACTACTGC  
 AGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCT  
 GACGCAATGCTCAGACGCAGAAGGGA  
 CTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAGA  
 GGCCTGGGGTGCAGGAGTGCTACCA  
 CGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACTGGAAGAACCTGCCAAGCTTGGTCATCTAT  
 GACACCACACTCGCATAGTCGGACC  
 CCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGT  
 TATACGAGGGATCCCGGTGTCAGGT  
 GGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTT  
 CCAAGCCTAGAGGCTCCTTCCGAACA  
 AGCACCGACTGAGCAAAGGCCTGGGGTGCAGGAGTGCTACCATGGTAATGGACAGAGTTATCGAGGCACATAC  
 TCCACCACTGTCACAGGAAGAACCTGC  
 CAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATG  
 AACTACTGCAGGAATCCAGATGCTG  
 TGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGAC  
 GCAGAAGGGACTGCCGTCGCGCCTCC  
 GACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAAAGGCCTGGGGTGCAGG  
 AGTGCTACCATGGTAATGGACAGAGT  
 TATCGAGGCACATACTCCACCACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCAT  
 AGTCGGACCCCAGAATACTACCCAA  
 ATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCG  
 GTGTCAGGTGGGAGTACTGCAACCT  
 GACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTC  
 CTTCCGAACAAGCACCGACTGAGCAG

AGGCCTGGGGTGCAGGAGTGCTACCACGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACTGG  
 AAGAACCTGCCAAGCTTGGTCATCTA  
 TGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATC  
 CAGATGCTGTGGCAGCTCCTTATTG  
 TTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCC  
 GTCGCGCCTCCGACTGTTACCCCGGTT  
 CCAAGCCTAGAGGCTCCTTCCGAACAAGCACCGACTGAGCAGAGGCCTGGGGTGCAGGAGTGCTACCACGGTA  
 ATGGACAGAGTTATCGAGGCACATACT  
 CCACCACTGTCACTGGAAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAAT  
 ACTACCCAAATGCTGGCTTGATCAT  
 GAACTACTGCAGGAATCCAGATGCTGTGGCAGCTCCTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGT  
 ACTGCAACCTGACGCAATGCTCAGAC  
 GCAGAAGGGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAAGCACC  
 GACTGAGCAGAGGCCTGGGGTGCAGG  
 AGTGCTACCACGGTAATGGACAGAGTTATCGAGGCACATACTCCACCACTGTCACTGGAAGAACCTGCCAAGCT  
 TGGTCATCTATGACACCACACTCGCA  
 TAGTCGGACCCCAGAATACTACCCAAATGCTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAG  
 CTCCTTATTGTTATACGAGGGATCCC  
 GGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAGGGACTGCCGTCGCGCCTCCGACTG  
 TTACCCCGGTTCCAAGCCTAGAGGCTC  
 CTTCCGAACAAGCACCGACTGAGCAGAGGCCTGGGGTGCAGGAGTGCTACCACGGTAATGGACAGAGTTATCG  
 AGGCACATACTCCACCACTGTCACTGG  
 AAGAACCTGCCAAGCTTGGTCATCTATGACACCACACTCGCATAGTCGGACCCCAGAATACTACCCAAATGCTGG  
 CTTGATCATGAACTACTGCAGGAAT  
 CCAGATCCTGTGGCAGCCCCTTATTGTTATACGAGGGATCCCAGTGTCAGGTGGGAGTACTGCAACCTGACACA  
 ATGCTCAGACGCAGAAGGGACTGCCG  
 TCGCGCCTCCAACCTATTACCCCGATTCCAAGCCTAGAGGCTCCTTCTGAACAAGCACCAACTGAGCAAAGGCCTG  
 GGGTGCAGGAGTGCTACCACGGAAA  
 TGGACAGAGTTATCAAGGCACATACTTCATTACTGTCACAGGAAGAACCTGCCAAGCTTGGTCATCTATGACACC  
 ACACTCGCATAGTCGGACCCCAGCA  
 TACTACCCAAATGCTGGCTTGATCAAGAACTACTGCCGAAATCCAGATCCTGTGGCAGCCCCTTGGTGTTATACA  
 ACAGATCCCAGTGTCAGGTGGGAGT  
 ACTGCAACCTGACACGATGCTCAGATGCAGAATGGACTGCCTTCGTCCCTCCGAATGTTATTCTGGCTCCAAGCC  
 TAGAGGCTTTTTTTGAACAAGCACT  
 GACTGAGGAAACCCCGGGGTACAGGACTGCTACTACCATTATGGACAGAGTTACCGAGGCACATACTCCACCA  
 CTGTCACAGGAAGAAGCTTGCCAAGCT  
 TGGTCATCTATGACACCACACCAGCATAGTCGGACCCCAGAAAATACCCAAATGCTGGCCTGASCAGGAAGTAC  
 TGCAGGAATCCAGATGCTGAGATTC  
 GCCCTTGGTGTTACACCATGGATCCCAGTGTCAGGTGGGAGTACTGCAACCTGACACAATGCCTGGTGACAGAA  
 TCAAGTGTCTTGCAACTCTCACGGT  
 GGTCCCAGAYCCAAGCACAGAGGCTTCTTCTGAAGAAGCACCAACGGAGCAAAGCCCCGGGGTCCAGGATTGC  
 TACCATGGTGATGGACAGAGTTATCGA  
 GGCTCATTCTCTACCACTGTCACAGGAAGGACATGTCAGTCTTGGTYCTCTATGACACCACACTGGCATCAGAGG  
 ACAACAGAATATTATCCAAATGGTG

GCCTGACCAGGAACTACTGCAGGAATCCAGATGCTGAGATTAGTCCTTGGTGTTATACCATGGATCCCAATGTCA  
 GATGGGAGTACTGCAACCTGACACA  
 ATGTCCAGTGACAGAATCAAGTGTCTTGGCGACGTCCACGGCTGTTTCTGAACAAGCACCAACGGAGCAAAGCC  
 CCACAGTCCAGGACTGCTACCATGGT  
 GATGGACAGAGTTATCGAGGCTCATTCTCCACCACTGTTACAGGAAGGACATGTCAGTCTTGGTCCTCTATGACA  
 CCACACTGGCATCAGAGAACCACAG  
 AATACTACCCAAATGGTGGCCTGACCAGGAACTACTGCAGGAATCCAGATGCTGAGATTGCCCCTTGGTGTTATA  
 CCATGGATCCCAGTGTCAGATGGGA  
 GTACTGCAACCTGACRCAATGTCCAGTGATGGAATCAACTSTCCTCACAACCTCCCACGGTGGTCCCAGTTCCAAG  
 CACAGAGSTTCCTTCTGAAGAAGCA  
 CCAACTGAAAACAGCACTGGGGTCCAGGACTGCTACCGAGGTGATRGACAGAGTTATCGAGGCMCACTCTCCA  
 CCACTATCACAGGAAGAACATGTCAGT  
 CTTGGTCGTCTATGACACCACATTGGCATCGGAGGATCCCATTATACTATCCAAATGRTGGCCTGACCAGGAACT  
 ACTGCAGGAATCCAGATGCTGAGAT  
 TCGCCCTTGGTGTTACACCATGGATCCCAGTGTCAGGTGGGAGTACTGCAACCTGACACGATGTCCAGTGACAG  
 AATCGAGTGTCCTCACAACCTCCCACA  
 GTGGCCCCGGTTCCAAGCACAGAGGCTCCTTCTGAACAAGCACCACTGAGAAAAGCCCTGTGGTCCAGGATTG  
 CTACCATGGTGATGGAYGGAGTTATC  
 GAGGCATATCCTCCACCACTGTCACAGGAAGGACCTGTCAATCTTGGTCATCTATGATACCACACTGGYATCAGA  
 GGACCCAGAAAACCTACCCAAATGC  
 TGGCCTGACCGAGAACTACTGCAGGAATCCAGATTCTGGGAAACAACCCTGGTGTTACACAACCGATCCGTGTG  
 TGAGGTGGGAGTACTGCAATCTGACA  
 CAATGCTCAGAAACAGAATCAGGTGTCCTAGAGACTCCCACTGTTGTTCCAGTTCCAAGCAYGGAGGCTCATTCT  
 GAAGCAGCACCAACTGAGCAAACCC  
 CTGTGGTCCGGCAGTGCTACCATGGTAATGGCCAGAGTTATCGAGGCACATTCTCCACCACTGTCACAGGAAGG  
 ACATGTCAATCTTGGTCATCCATGAC  
 ACCACACCGGCATCAGAGGACCCAGAAAACCTACCCAAATGATGGCCTGACAATGAACTACTGCAGGAATCCAG  
 ATGCCGATACAGGCCCTTGGTGTTTT  
 ACCAYGGACCCAGCATCAGGTGGGAGTACTGCAACCTGACGCGATGCTCAGACACAGAAGGGACTGTGGTCTG  
 CTCCTCCGACTGTCATCCAGGTTCCAA  
 GCCTAGGGCCTCCTTCTGAACAAGAYTGTATGTTTGGGAATGGGAAAGGATACCGGGGCAAGAARGCAACCACT  
 GTTACTGGGACGCCATGCCAGGAATG  
 GGCTKCCCAGGAGCCCCATAGACACAGCACRTTCATTCCAGGGACAAATAAATGGGCAGGTCTGGAAAAAAATT  
 ACTGCCGTAACCCTGATGGTGACATC  
 AATGGTCCCTGGTGCTACACAATGAATCCAAGAAAACCTTTTGGACTACTGTGATATCCCTCTCTGTGCATCCTCTT  
 CATTTGATTGTGGGAAGCCTCAAG  
 TGGAGCCGAAGAAATGTCCTGGAAGCATTGTAGSGGKGTGTGTGGCCACCCACATTCCTGGCCCTGGCAAGTC  
 AGTCTCAGAACAAGGTTTGGAAAGCA  
 CTTCTGTGGAGGCACCTTAATATCCCCAGAGTGGGTGCTGACTGCTGCTCACTGCTTGAAGAAGTCCTCAAGGCC  
 TTCATCCTACAAGGTCATCCTGGGT  
 GCACACCAAGAAGTGAACCTCGAATCTCATGTTTCAGGAAATRGAAAGTGTCTAGGCTGTTCTTGGAGCCACACA  
 AGCAGATATTGCCTTGCTAAAGCTAA  
 GCAGGCCTGCCGTCATCACTGACAAAGTAATGCCAGCTTGTCTGCCATCCCCAGACTACATGGTCAACCGCCAGGA  
 CTGAATGTTACATCACTGGCTGGGG

AGAAACCCAAGGTACCTTTGGGACTGGCCTTCTCAAGGAAGCCCAGCTCCTTGTTATTGAGAATGAAGTGTGCA  
 ATCACTATAAGTATATTTGTGCTGAG  
 CATTTGGCCAGAGGCACTGACAGTTGCCAGGGTGACAGTGGAGGGCCTCTGGTTTGCTTCGAGAAGGACAAAT  
 ACATTTTACAAGGAGTCACTTCTTGGG  
 GTCTTGGCTGTGCAYGCCCCAATAAGCCTGGTGTCTATGCTCGTGTTTCAAGGTTTGTTACTTGGATTGAGGGAA  
 TGATGAGAAATAATTAATTGGACGG  
 GAGACAGAGTGAAGCATCAACCTACTTAGAAGCTGAAACGTGGGTAAGGATTTAGCATGCTGGAAATAATAGA  
 CAGCAATCAAACGAAGACACTGTTCCC  
 AGCTACCAGCTATGCCAAACCTTGGCATTITTTGGTATITTTGTGTATAAGCTTTTAAGGTCTGACTGACAAATTCT  
 GTATTAAGGTGTCATAGCTATGAC  
 ATTTGTTAAAAATAAACTCTGCACTTATTTTGATTGA

Protein Sequence (SEQ ID NO:41):

MEHKEVVLLLLFLKSAAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPNAGLIMN  
 YCRNPDAVAAPYCYTRDPGVRWEYC  
 NLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSR  
 TPEYYPNAGLIMNYCRNPDAVAAP  
 YCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRT  
 CQAWSSMTPHSHSRTPEYYPNAGL  
 IMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNG  
 QSYRGTYSTTVTGRTCQAWSSMTPH  
 SHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQ  
 RPGVQECYHGNGQSYRGTYSTTV  
 TGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTP  
 VPSLEAPSEQAPTEQRPGVQECYH  
 GNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQC  
 SDAEGTAVAPPTVTPVPSLEAPSEQ  
 APTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTR  
 DPGVRWEYCNLTQCSDAEGTAVAPP  
 TVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCR  
 NPDAVAAPYCYTRDPGVRWEYCNL  
 TQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPE  
 EYYPNAGLIMNYCRNPDAVAAPYC  
 YTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQ  
 AWSSMTPHSHSRTPEYYPNAGLIM  
 NYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQS  
 YRGTYSTTVTGRTCQAWSSMTPHSH  
 SRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRP  
 GVQECYHGNGQSYRGTYSTTVTG  
 RTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPV  
 PSLEAPSEQAPTEQRPGVQECYHGN  
 GQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDA  
 EGTAVAPPTVTPVPSLEAPSEQAP  
 TEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDP  
 GVRWEYCNLTQCSDAEGTAVAPPTV

TPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNP  
 DAVAAPYCYTRDPGVRWEYCNLTQ  
 CSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEY  
 YPNAGLIMNYCRNPDAVAAPYCYT  
 RDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQA  
 WSSMTPHSHSRTPEYYPNAGLIMNY  
 CRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYR  
 GTYSTTVTGRTCQAWSSMTPHSHSR  
 TPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPG  
 VQECYHGNGQSYRGTYSTTVTGRT  
 CQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPS  
 LEAPSEQAPTEQRPGVQECYHGNGQ  
 SYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAE  
 GTAVAPPTVTPVPSLEAPSEQAPTE  
 QRPVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGV  
 RWEYCNLTQCSDAEGTAVAPPTVTP  
 VPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDA  
 VAAPYCYTRDPGVRWEYCNLTQCS  
 DAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYP  
 NAGLIMNYCRNPDAVAAPYCYTRD  
 PGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSS  
 MTPHSHSRTPEYYPNAGLIMNYCR  
 NPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRG  
 TYSTTVTGRTCQAWSSMTPHSHSRT  
 EYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQ  
 EYHGNGQSYRGTYSTTVTGRTCQ  
 AWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEA  
 PSEQAPTEQRPGVQECYHGNGQSY  
 RGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGT  
 AVAPPTVTPVPSLEAPSEQAPTEQR  
 PGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAVAAPYCYTRDPGVR  
 WEYCNLTQCSDAEGTAVAPPTVTPV  
 SLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNAGLIMNYCRNPDAV  
 AAPYCYTRDPGVRWEYCNLTQCSDA  
 EGTAVAPPTVTPVPSLEAPSEQAPTEQRPGVQECYHGNGQSYRGTYSTTVTGRTCQAWSSMTPHSHSRTPEYYPNA  
 GLIMNYCRNPDPVAAPYCYTRDPS  
 VRWEYCNLTQCSDAEGTAVAPPTITIPISLEAPSEQAPTEQRPGVQECYHGNGQSYQGTYFITVTGRTCQAWSSMTP  
 HSHSRTPAYYPNAGLIKNYCRNP  
 DPVAAPWCYTDPVSVRWEYCNLTRCSDAEWTAFFVPPNVILAPSLEAFFEQALTEETPGVQDCYHYHGQSYRGTYST  
 VTGRTCQAWSSMTPHQHSRTPEN  
 YPNAGLTRNYCRNPDAEIRPWCYTMDPSVRWEYCNLTQCLVTESSVLATLTVVPDPSTEASSEEAPTEQSPGVQDCY  
 HGDGQSYRGSFSTTVTGRTCQSW  
 SSMTPHWHQRTTEYYPNGGLTRNYCRNPDAEISPWCYTMDPNVRWEYCNLTQCPVTESSVLATSTAVSEQAPTEQ  
 SPTVQDCYHGDGQSYRGSFSTTVTG

RTCQSWSSMTPHWHQRTTEYYPNGGLTRNYCRNPDAEIRPWCYTMDPSVRWEYCNLTQCPVMESTLLTTPTVVP  
VPSTELPSEEAPTENSTGVQDCYRGD  
GQSYRGTLSTTITGRTCQSWSSMTPHWHRRIPLYYPNAGLTRNYCRNPDAEIRPWCYTMDPSVRWEYCNLTRCPVT  
ESSVLTTPTVAPVPSTEAPSEQAP  
PEKSPVVQDCYHGDGRSYRGISSTTVTGRTCQSWSSMIPHWHQRTPENYPNAGLTENYCRNPDSGKQPWCYTDDP  
CVRWEYCNLTQCSETESGVLETPTV  
VPVPSMEAHSAAAPTEQTPVVRQCYHGNQGQSYRGTFSTTVTGRTCQSWSSMTPHRHQRTPENYPNDGLTMNYCR  
NPDADTGPWCFTMDPSIRWEYCNLTR  
CSDTEGTVVAPPTVIQVPSLGPPSEQDCMFGNGKGYRGKKATTVTGTPCQEWAAQEPHRHSTFIPGTNKGWAGLEK  
NYCRNPDGDINGPWCYTMNPRKLFD  
YCDIPLCASSSFDCGKPQVEPKKCPGSIVGGCVAHPHWPWQVSLRTRFGKHFCGGTLISPEWVLTAHCLKKSSRPS  
SYKVILGAHQEVNLESHVQEIE  
VSRLFLEPTQADIALLKLSRPAVITDKVMPACLPSPDYMTARTECYITGWGETQGTFGTGLLKEAQLLVIENEVCNHY  
KYICAEHLARGTDSCQGDSGG  
PLVCFEKDKYILQGVTSWGLGCARPKNKPGVYARVSRFVTWIEGMMRNN

SNP Information

Context (SEQ ID NO:61):  
CTGCTGCTCACTGCTGAAGAAGTCCTCAAGGCCTTCATCCTACAAGGTCATCCTGGGTGCACACCAAGAAGTGA  
ACCTCGAATCTCATGTTCAAGGAAAT  
R  
GAAGTGTCTAGGCTGTTCTTGGAGCCCACACAAGCAGATATTGCCTTGCTAAAGCTAAGCAGGCCTGCCGTCATC  
ACTGACAAAGTAATGCCAGCTTGTC

Celera SNP ID: hCV25930271  
SNP Position Transcript: 13242  
SNP Source: Applera  
Population(AAllele,Count): african american(A,37|G,1) caucasian(A,36|G,2) total(A,73|G,3)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:41, 51, (I,ATA) (M,ATG)

SNP Source: dbSNP  
Population(AAllele,Count): no\_pop(G,-|A,-)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:41, 51, (I,ATA) (M,ATG)

Gene Number: 147  
Celera Gene: hCG23557 - 104000116665608  
Celera Transcript: hCT2273323 - 104000116665629  
Public Transcript Accession: NM\_003053  
Celera Protein: hCP1885376 - 197000069428640  
Public Protein Accession: NP\_003044  
Gene Symbol: SLC18A1  
Protein Name: solute carrier family 18 (vesicular monoamine), member 1;CGAT;VAT1;VMAT1

Celera Genomic Axis: GA\_x5YUV32VUUD(23750742..23789095)  
Chromosome: Chr8  
OMIM number: 193002  
OMIM Information: SOLUTE CARRIER FAMILY 18, MEMBER 1;SLC18A1

Transcript Sequence (SEQ ID NO:22):

CACACRCACACATACACAGAATCCTCAGATAACAGGAGGCAATAAATCCAACAGCACATCCACGTTTCAGAGAAC  
AGTGTCCCTGCTGTCTTGCTAACAGC  
TGCCAATACCTCACTGAGTGCCTCACACCAACATGGGCTCCAAGTGAGTTTCCTTCGTCTGGGCAGACTCCCTCCC  
CTCTTCCATAAAGGCTGCAGGAGA  
CCTGTAGCTGTCACAGGACCTTCCSTAAGAGCCCCGCAGGGGAAGACTGCCCCAGTCCGGCCATCACCATGCTCC  
GGMCCATTCTGGATGCTCCCCAGCGG  
TTGCTGAAGGAGGGGAGAGCGTCCCGGCAGCTGGTGGTGGTGGTATTTCGTGCTTTGCTCCTGGACAACAT  
GCTGTTTACTGTGGTGGTGCCAATTG  
TGCCACCTTCCTATATGACATGGAGTTCAAAGAAGTCAACTCTTCTCTGCACCTCGGCCATGCCGGAAGTTCCCC  
ACATGCCCTCGCCTCTCCTGCCTT  
TTCCACCATCTKCTCCTTCTTCAACAACAACACCGTGGCTGTTGAAGAAAGCGTACCTASTGGAATAGCATGGATG  
AATGACACTGCCAGCACCATCCCA  
CCTCCAGCCACTGAAGCCATCTCAGCTCATAAAAACAACTGCTTGCAAGGCACAGGTTTCTTGGAGGAAGAGAYT  
ACCCGGGTCRGGGTCTGTTTGCTT  
CAAAGGCTGTGATGCAACTTCTGGTCAACCCATTCGTGGGCCCTCTCACCAACAGGATTGGATATCATAYCCCCA  
TGTTTGCTGGCTTTGTTATCATGTT  
TCTCTCCACAGTTATGTTTGCTTTTTCTGGGACCTATACTCTACTCTTTGTGGCCCGAACCTTCAAGGCATTGGAT  
CTTCATTTTCATCTGTTGCAGGT  
CTTGAATGCTGGCCAGTGTCTACACTGATGACCATGAGAGAGGACGAGCCATGGGAACTGCTCTGGGGGGCC  
TGGCCTTGGGGTTGCTGGTGGGAGCTC  
CCTTTGGAAGTRTAATGTAYGAGTTTGTTGGGAAGTCTGCACCCTTCCTCATCCTGGCCTTCCTGGCACTACTGGA  
TGGAGCACTCCAGCTTTGCATCCT  
ACAGCCTTCCAAAGTCTCTCCTGAGAGTGCCAAGGGGACTCCCCTCTTTATGCTTCTCAAAGACCCTTACATCCTG  
GTGGCTGCAGGTCTAGCTTTCTTG  
CCTGCCAGTGTGTCCTACCTCATTGGCACCAACCTCTTTGGTGTGTTGGCCAACAAGATGGGTTCGGTGGCTGTGT  
TCCCTAATCGGGATGCTGGTAGTAG  
GTACCAGCTTGCTCTGTGTTCTCTGGCTCACAATATTTTTGGTVTCATTGGCCCCAATGCAGGGCTTGGCCTTGC  
CATAGGCATGGTGGATTCTTCTAT  
GATGCCCATCATGGGGCACCTGGTGGATCTACGCCACACCTCGGTGTATGGGAGTGTCTAYGCCATCGCTGATG  
TGGCTTTTTGCATGGGCTTTGCTATA  
GGTCCATCCACCGGTGGTGGCATTGTAAAGGCCATCGGTTTTCCCTGGCTCATGGTCATCACTGGGGTCATCAAC  
ATCGTCTATGCTCCACTCTGCTACT  
ACCTGCGGAGCCCCCGGCRAAGGAAGAGAAGCTTGCTATTCTGAGTCAGGACTGCCCCATGGAGACCCGGAT  
GTATGCAACCCAGAAGCCACGAAGGA  
ATTTCTCTGGGGGAGGACAGTGATGAGGAGCCTGACCATGAGGAGTAGCAGCAGAAGGTGCTCCTTGAATTM  
ATGATGCCTCAGTGACCACCTCTTTCC  
CTGGGACCAGATCACCATGGCTGAGCCACGGCTCAGTGGGCTTCACATACCTCTGCCTGGGAATCTTCTTTCCT  
CCCCTCCCATGGACACTGTCCCTGA

TACTCTTCTCACCTGTGTAACCTGTAGCTCTTCTCTATGCCTTGGTGCCGCAGTGGCCCATCTTTTATGGGAAGA  
 CAGAGTGATGCACCTTCCCGCTGC  
 TGTGAGGTTGATTAAACTTGAGCTGTGACGGGTTCTGCAAGGGGTGACTCATTGCATAGAGGTGGTAGTGAGTA  
 ATGTGCCCCCTGAAACCAGTGGGGTGA  
 CTGACAAGCCTCTTTAATCTGTTGCCTGATTTTCTCTGGCATAGTCCCAACAGATCGGAAGAGTGTTACCCTCTTT  
 TCCTCAACGTGTTCTTTCCCGGGT  
 TTTCCAGCCGAGTTGAGAAAATGTTCTCAGCATTGTCTTGCTGCCAAATGCCAGCTTGAAGAGTTTTGTTTTGTT  
 TTTTTTCATTTATTTTTTTTTTTA  
 ATAAAGTGAGTGATTTTTCTGTGGCTAAATCTAGAGCTGCTAAAAGGGCTTTACCCTCAGTGAAAAGTGTCTTCT  
 ATTTTCATTATCTTTCAGAAACAGG  
 AGCCCATTTCTCTTCTGCTGGAGTTATTGACATTCTCCTGACCTCCCCTGTGTGTTCTACCTTTTCTGAACCTCTTA  
 GACTCTTAGAAATAAAAGTAGA  
 AGAAAGACAGAAAAAATAACTGATTAGACCCAAGATTTTCATGGGAAGAAGTTAAAAGAAACTGCCTTGAAATCC  
 CTCCTGATTGTAGATTTCTTAAYAGG  
 AGGGGTGTAATGTGACATTGTTCACTTGCTAATAAATACATTATTGCCTAATTCAGAC

Protein Sequence (SEQ ID NO:42):

MLRTILDAPQRLLKEGRASRQLVLVVVFVALLLDNMLFTVVVPIVPTFLYDMEFKEVNSSLHLGHAGSSPHALASPAFS  
 TIFSFNNNTVAVEESVPSGI  
 AWMNDTASTIPPPATEAISAHKNNCLQGTGFLEEETTRVGVLFASKAVMQLLVNPFGVPLTNRIGYHIPMFAGFVIM  
 FLSTVMFAFSGTYTLFVARTLQ  
 GIGSSFSSVAGLGLMLASVYTDHERGRAMGTALGGLALGLLVGAPFGSVMYEFVGKSAPFLILAFLLDGLQLCILQ  
 PSKVSPESAKGTPLFMLLKDP  
 YILVAAGLAFLPASVSYLIGTNLFGVLANKMGRWLCSLIGMLVVGTSLLCVPLAHNIFGLIGNAGLGLAIGMVDSSM  
 MPIMGHLVDLRHTSVYGSVYAI  
 ADVAFCMGFAIGPSTGGAIVKAIGFPWLMVITGVINIVYAPLCYYLRSPPAKEEKLAILSQDCPMETRMATQKPTKEF  
 PLGEDSDEEPDHEE

SNP Information

Context (SEQ ID NO:62):

CATGCCGGAAGTTCCCCACATGCCCTCGCCTCTCCTGCCTTTTCCACCATCTTCTCCTTCTTCAACAACAACACCGT  
 GGCTGTTGAAGAAAGCGTACCTA  
 S

TGGAATAGCATGGATGAATGACACTGCCAGCACCATCCACCTCCAGCCACTGAAGCCATCTCAGCTCATAAAAA  
 CAACTGCTTGCAAGGCACAGGTTTC

Celera SNP ID: hCV2716008

SNP Position Transcript: 560

SNP Source: Applera

Population(AAllele,Count): african american(C,8|G,30) caucasian(C,10|G,30) total(C,18|G,60)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:42, 8, (T,ACT) (S,AGT)

SNP Source: Applera

Population(AAllele,Count): african american(C,2|G,22) caucasian(C,2|G,24) total(C,4|G,46)  
 SNP Type: MISSENSE MUTATION  
 Protein Coding: SEQ ID NO:42, 8, (T,ACT) (S,AGT)

SNP Source: Celera  
 Population(AAllele,Count): Caucasian(G,61|C,249) Chinese(G,6|C,54) Japanese(G,7|C,13)  
 African(G,21|C,75) total(G,95|C,391) no\_pop(C,1|G,5) total(G,5|C,1)  
 SNP Type: MISSENSE MUTATION  
 Protein Coding: SEQ ID NO:42, 8, (T,ACT) (S,AGT)

Gene Number: 147  
 Celera Gene: hCG23557 - 104000116665608  
 Celera Transcript: hCT14664 - 104000116665609  
 Public Transcript Accession: NM\_003053  
 Celera Protein: hCP40951 - 197000069428639  
 Public Protein Accession: NP\_003044  
 Gene Symbol: SLC18A1  
 Protein Name: solute carrier family 18 (vesicular monoamine), member 1;CGAT;VAT1;VMAT1  
 Celera Genomic Axis: GA\_x5YUV32VUUD(23750742..23789095)  
 Chromosome: Chr8  
 OMIM number: 193002  
 OMIM Information: SOLUTE CARRIER FAMILY 18, MEMBER 1;SLC18A1

Transcript Sequence (SEQ ID NO:23):

CACACRCACACATACAGAAATCCTCAGATAACAGGAGGCAATAAATCCAACAGCACATCCACGTTTCAGAGAAC  
 AGTGTCCCTGCTGTCTTGCTAACAGC  
 TGCCAATACCTCACTGAGTGCCTCACACCAACATGGGCTCCAAGTGAGTTTCCTTCGTCTGGGCAGACTCCCTCCC  
 CTCTTCCATAAAGGCTGCAGGAGA  
 CCTGTAGCTGTACAGGACCTTCCSTAAGAGCCCGCAGGGGAAGACTGCCCCAGTCCGGCCATCACCATGCTCC  
 GGMCCATTCTGGATGCTCCCCAGCGG  
 TTGCTGAAGGAGGGGAGAGCGTCCCGGCAGCTGGTGCTGGTGGTATTTCGTCGCTTTGCTCCTGGACAACAT  
 GCTGTTTACTGTGGTGGTGCCAATTG  
 TGCCCACCTTCCTATATGACATGGAGTTCAAAGAAGTCAACTCTTCTCTGCACCTCGGCCATGCCGGAAGTTCCCC  
 ACATGCCCTCGCCTCTCCTGCCTT  
 TTCCACCATCTKCTCCTTCTTCAACAACAACACCGTGGCTGTTGAAGAAAGCGTACCTASTGGAATAGCATGGATG  
 AATGACACTGCCAGCACCATCCCA  
 CCTCCAGCCACTGAAGCCATCTCAGCTCATAAAAACAACTGCTTGCAAGGCACAGGTTTCTTGGAGGAAGAGAYT  
 ACCCGGGTCRGGGTCTGTTTGCTT  
 CAAAGGCTGTGATGCAACTTCTGGTCAACCCATTTCGTGGGCCCTCTACCAACAGGATTGGATATCATAYCCCCA  
 TGTTTGCTGGCTTTGTTATCATGTT  
 TCTCTCCACAGTTATGTTTGCTTTTTCTGGGACCTATACTCTACTCTTTGTGGCCCGAACCCTTCAAGGCATTGGAT  
 CTTCAATTTTCATCTGTTGCAGGT  
 CTTGGAATGCTGGCCAGTGTCTACACTGATGACCATGAGAGAGGACGAGCCATGGGAACTGCTCTGGGGGGGCC  
 TGGCCTTGGGGTTGCTGGTGGGAGCTC

CCTTTGGAAGTRTAATGTAYGAGTTTGTGGGAAGTCTGCACCCTTCCTCATCCTGGCCTTCCTGGCACTACTGGA  
 TGGAGCACTCCAGCTTTGCATCCT  
 ACAGCCTTCCAAAGTCTCTCCTGAGAGTGCCAAGGGGACTCCCCTCTTTATGCTTCTCAAAGACCCTTACATCCTG  
 GTGGCTGCAGGGTCCATCTGCTTT  
 GCCAACATGGGGGTGGCCATCCTGGAGCCCACACTGCCCATCTGGATGATGCAGACCATGTGCTCCCCCAAGTG  
 GCAGCTGGGTCTAGCTTTCTTGCCTG  
 CCAGTGTGTCTACCTCATTGGCACCAACCTCTTTGGTGTGTTGGCCAACAAGATGGGTGCGGTGGCTGTGTTCCC  
 TAATCGGGATGCTGGTAGTAGGTAC  
 CAGCTTGCTCTGTGTTCTCTGGCTCACAATATTTTTGGTVTCATTGGCCCCAATGCAGGGCTTGGCCTTGCCATA  
 GGCATGGTGGATTCTTCTATGATG  
 CCCATCATGGGGCACCTGGTGGATCTACGCCACACCTCGGTGTATGGGAGTGTCTAYGCCATCGCTGATGTGGC  
 TTTTTCATGGGGCTTTGCTATAGGTC  
 CATCCACCGGTGGTGCCATTGTAAAGGCCATCGGTTTTCCCTGGCTCATGGTCATCACTGGGGTCATCAACATCG  
 TCTATGCTCCACTCTGCTACTACCT  
 GCGGAGCCCCCGGCRAAGGAAGAGAAGCTTGCTATTCTGAGTCAGGACTGCCCCATGGAGACCCGGATGTAT  
 GCAACCCAGAAGCCACGAAGGAATTT  
 CCTCTGGGGGAGGACAGTGATGAGGAGCCTGACCATGAGGAGTAGCAGCAGAAGGTGCTCCTTGAATTMATG  
 ATGCCTCAGTGACCACCTCTTCCCTGG  
 GACCAGATCACCATGGCTGAGCCCACGGCTCAGTGGGCTTCACATACCTCTGCCTGGGAATCTTCTTCCCTCCCCT  
 CCCATGGACACTGTCCCTGATACT  
 CTTCTCACCTGTGTAACCTGTAGCTCTTCCTCTATGCCTTGGTGCCGCAGTGGCCCATCTTTTATGGGAAGACAGA  
 GTGATGCACCTTCCCGCTGCTGTG  
 AGGTTGATTAACTTGAGCTGTGACGGGTTCTGCAAGGGGTGACTCATTGCATAGAGGTGGTAGTGAGTAATGT  
 GCCCTGAAACCAGTGGGGTGACTGA  
 CAAGCCTCTTTAATCTGTTGCCTGATTTTCTCTGGCATAGTCCCAACAGATCGGAAGAGTGTTACCCTCTTTTCCTC  
 AACGTGTTCTTTCCCGGGTTTTT  
 CCAGCCGAGTTGAGAAAATGTTCTCAGCATTGTCTTGCTGCCAAATGCCAGCTTGAAGAGTTTTGTTTTGTTTTT  
 TTCATTTATTTTTTTTTTTAATAA  
 AGTGAGTGATTTTTCTGTGGCTAAATCTAGAGCTGCTAAAAGGGCTTTACCCTCAGTGAAAAGTGTCTTCTATTTT  
 CATTATCTTTCAGAAACAGGAGCC  
 CATTCTCTTCTGCTGGAGTTATTGACATTCTCCTGACCTCCCCTGTGTGTTCTACCTTTTCTGAACCTCTTAGACT  
 CTTAGAAATAAAAGTAGAAGAA  
 AGACAGAAAAAATAACTGATTAGACCCAAGATTTTCATGGGAAGAAGTTAAAAGAACTGCCTTGAAATCCCTCC  
 TGATTGTAGATTTCTAAYAGGAGGG  
 GTGTAATGTGACATTGTTCTACTTGCTAATAAATACATTATTGCCTAATTCAGAC

Protein Sequence (SEQ ID NO:43):

MLRTILDAPQRLLKEGRASRQLVLVVVFVALLLDNMLFTVVVPIVPTFLYDMEFKEVNSSLHLGHAGSSPHALASPAFS  
 TIFSFNNNTVAVEESVPSGI  
 AWMNDTASTIPPPATEAISAHKNCLQGTGFLEEETTRVGVLFAASKAVMQLLVNPFGPLTNRIGYHIPMFAGFVIM  
 FLSTVMFAFSGTYTLLFVARTLQ  
 GIGSSFSSVAGLGMLASVYTDDHERGRAMGTALGGLALGLLVGAPFGSVMYEFVGKSAPFLILAFLLDGLQLCILQ  
 PSKVSPESAKGTPLFMLLKDP  
 YILVAAGSICFANMGVAILEPTLPIWMMQTMCSPKWQLGLAFLPASVSYLIGNLFGVLANKMGRWLCSLIGMLVV  
 GTSLLCVPLAHNIFGLIGNAGLG

LAIGMVDSSMMPIMGHLVDLRHTSVYGSVYAIADVAFCMGFAIGPSTGGAIVKAIGFPWLMVITGVINIVYAPLCYYL  
RSPPAKEEKLAILSQDCPMETR  
MYATQKPTKEFPLGEDSDEEPDHEE

SNP Information

Context (SEQ ID NO:63):  
CATGCCGGAAGTTCCCCACATGCCCTCGCCTCTCCTGCCTTTTCCACCATCTTCTCCTTCTTCAACAACAACACCGT  
GGCTGTTGAAGAAAGCGTACCTA  
S  
TGGAATAGCATGGATGAATGACACTGCCAGCACCATCCCACCTCCAGCCACTGAAGCCATCTCAGCTCATAAAAA  
CAACTGCTTGCAAGGCACAGGTTTC  
Celera SNP ID: hCV2716008  
SNP Position Transcript: 560  
SNP Source: Applera  
Population(AAllele,Count): african american(C,8|G,30) caucasian(C,10|G,30) total(C,18|G,60)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:43, 8, (T,ACT) (S,AGT)

SNP Source: Applera  
Population(AAllele,Count): african american(C,2|G,22) caucasian(C,2|G,24) total(C,4|G,46)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:43, 8, (T,ACT) (S,AGT)

SNP Source: Celera  
Population(AAllele,Count): Caucasian(G,61|C,249) Chinese(G,6|C,54) Japanese(G,7|C,13)  
African(G,21|C,75) total(G,95|C,391) no\_pop(C,1|G,5) total(G,5|C,1)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:43, 8, (T,ACT) (S,AGT)

Gene Number: 209  
Celera Gene: hCG37135 - 79000075741507  
Celera Transcript: hCT2315937 - 79000075741516  
Public Transcript Accession: NM\_006404  
Celera Protein: hCP1858145 - 197000069371279  
Public Protein Accession: NP\_006395  
Gene Symbol: PROCR  
Protein Name: protein C receptor, endothelial (EPCR);CCCA;CCD41;EPCR;bA42O4.2  
Celera Genomic Axis: GA\_x5YUV32VYHA(29091543..29108236)  
Chromosome: Chr20  
OMIM number: 600646  
OMIM Information: PROTEIN C RECEPTOR;PROCR

## Transcript Sequence (SEQ ID NO:24):

GAAGAGCTCTGGTCTGCAGCTCCCAGTGAGATCGATGCAGAAGACAGAATCTTGCCGCATGACCCAGGCTAGAA  
 TGCAGTGGTGTGATCTCGGCCCTCTG  
 CAACCTCCACCTCCCAGACGTTTCGCCATATTGGCCAGGCTGGTCTCAAACCTCCTGGCCTCAAGTGATCTGCCCACC  
 TCGGCCTCCCAAAGTGCCAGCAGC  
 ATGCTCGGAGGAGTGACTTTAAAGCTTTTCTACTTGCTTCCTAGAACTGCAGCCAGCGGAGCCCGCAGCCGGCCC  
 GAGCCAGGAACCCAGGTCCGGAGCC  
 TCAACTTCAGGATGTTGACAACATTGCTGCCGATACTGCTGCTGTCTGGCTGGGCCTTTTGTAGCCAAGACGCCT  
 CAGATGGCCTCCAAAGACTTCATAT  
 GCTCCAGATCTCCTACTTCCGCGACCCCTATCACGTGTGGTACCAGGGCAACGCGTCGCTGGGGGGACACCTAA  
 CGCACGTGCTGGAAGGCCCAGACACC  
 AACACCACGATCATTGAGCTGCAGCCCTTGAGGAGCCCGAGAGCTGGGCGCGCACGCAGAGTGGCCTGCAGT  
 CCTACCTGCTCCAGTTCACGGCCTCG  
 TCGCCTGGTGCACCAGGAGCGGACCTTGGCCTTTCCTCTGACCATCCGCTGCTTCCTGGGCTGTGAGCTGCCTC  
 CCGAGGGCTCTAGAGCCCATGTCTT  
 CTTGAAGTGGCTGTGAATGGGAGCTCCTTTGTGAGTTTCCGGCCGGAGAGAGCCTTGTGGCAGGCAGACACCC  
 AGGTCACCTCCGGAGTGGTCACCTTC  
 ACCCTGCAGCAGCTCAATGCCTACAACCGCACTCGGTATGAACTGCGGGAATTCCTGGAGGACACCTGTSTGCA  
 GTATGTGCAGAAACATATTTCCGCGG  
 AAAACACGAAAGGGAGCCAAACAAGCCGCTCCTACACTTCGCTGGTCCTGGGCGTCCTGGTGGGCRGTTTCATC  
 ATTGCTGGTGTGGCTGTAGGCATCTT  
 CCTGTGCACAGGTGGACGGCGATGTTAATTACTCTCCAGCCCCSTCAGAAGGGGCTGGATTGATGGAGGCTGGC  
 AAGGGAAAGTTTCAGCTCACTGTGAA  
 GCCAGACTCCCCAACTGAAACACCAGAAGGTTTGGAGTGACAGCTCCTTTCTTCTCCACATCTGCCCACTGAAG  
 ATTTGAGGGAGGGGARATGGARAGG  
 AGAGGTGGACAAAGTACTTGGTTTGCTAAGAACCTAAGAACGTGTATGCTTTGCTGAATTAGTCTGATAAGTGA  
 ATGTTTATCTATCTTTGTGGAAAACA  
 GATAATGGAGTTGGGGCAGGAAGCCTATGGCCCATCCTCCAAAGACAGACAGAATCACCTGAGGCGTTCAAAA  
 GATATAACCAAATAAACAAGTCATCCM  
 CAATCAAAATACAACATTCAATACTTCCAGGTGTGTCAGACTTGGGATGGGACGCTGATATAATAGGGTAGAAA  
 GAAGTAACACGAAGAAGTGGTGGAAA  
 TGTAATCAAGTCATATGGCAGTGATCAATTATTAATCAATTAATAATAATAAATTTCTTATATTTAAGG

## Protein Sequence (SEQ ID NO:44):

MLGGVTLKLFYLLPRTAASGARSRPEPGTQVRSLNFRMLTLLPILLLSGWAFCSQDASDGLQRLHMLQISYFRDPYH  
 VWYQGNASLGGLHVLHLEGPDT  
 NTTIIQLQPLQEPESWARTQSGLQSYLLQFHGLVRLVHQERTLAFPLTIRCFLGCELPPEGSRAHVFFEVAVNGSSFVSF  
 RPERALWQADTQVTSGVVTF  
 TLQQLNAYNRTRYELREFLEDTCVQYVQKHISAENTKGSQTSRYSYSLVLGVLVGSFIIAGVAVGIFLCTGGRRRC

## SNP Information

Context (SEQ ID NO:64):

CTGTGTGCAGTATGTGCAGAAACATATTTCCGCGGAAAACACGAAAGGGAGCCAAACAAGCCGCTCCTACACTT  
CGCTGGTCCTGGGCGTCCTGGTGGGC

R

GTTTCATCATTGCTGGTGTGGCTGTAGGCATCTTCCTGTGCACAGGTGGACGGCGATGTTAATTACTCTCCAGCC  
CCGTCAGAAGGGGCTGGATTGATGG

Celera SNP ID: hCV25620145

SNP Position Transcript: 966

SNP Source: Applera

Population(AAllele,Count): african american(A,32|G,6) caucasian(A,38|G,2) total(A,70|G,8)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:44, 21, (S,AGT) (G,GGT)

SNP Source: Celera;HGBASE;dbSNP

Population(AAllele,Count): no\_pop(G,2|A,9) total(G,2|A,9);no\_pop(A,-|G,-) ;no\_pop(A,-|G,-)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:44, 21, (S,AGT) (G,GGT)

Gene Number: 209

Celera Gene: hCG37135 - 79000075741507

Celera Transcript: hCT2315936 - 79000075741508

Public Transcript Accession: NM\_006404

Celera Protein: hCP1858144 - 197000069371278

Public Protein Accession: NP\_006395

Gene Symbol: PROCR

Protein Name: protein C receptor, endothelial (EPCR);CCCA;CCD41;EPCR;bA42O4.2

Celera Genomic Axis: GA\_x5YUV32VYHA(29091543..29096814)

Chromosome: Chr20

OMIM number: 600646

OMIM Information: PROTEIN C RECEPTOR;PROCR

Transcript Sequence (SEQ ID NO:25):

ACTGCAGCCAGCGGAGCCCGCAGCCGGCCCGAGCCAGGAACCCAGGTCCGGAGCCTCAACTTCAGGATGTTGA  
CAACATTGCTGCCGATACTGCTGCTGT  
CTGGCTGGGCCTTTTGTAGCCAAGACGCCTCAGATGGCCTCCAAAGACTTCATATGCTCCAGATCTCCTACTTCCG  
CGACCCCTATCACGTGTGGTACCA  
GGGCAACGCGTCGCTGGGGGGACACCTAACGCACGTGCTGGAAGGCCCCAGACACCAACACCACGATCATTAG  
CTGCAGCCCTTGCAAGAGCCCGAGAGC  
TGGGCGCGCACGCAGAGTGGCCTGCAGTCCTACCTGCTCCAGTTCCACGGCCTCGTGCGCCTGGTGCACCAGGA  
GCGGACCTTGGCCTTTCCTCTGACCA  
TCCGCTGCTTCCTGGGCTGTGAGCTGCCTCCCGAGGGCTCTAGAGCCCATGTCTTCTTCGAAGTGGCTGTGAATG  
GGAGCTCCTTTGTGAGTTTCCGGCC  
GGAGAGAGCCTTGTGGCAGGCAGACACCCAGGTCACCTCCGGAGTGGTCACCTTCACCCTGCAGCAGCTCAATG  
CCTACAACCGCACTCGGTATGAACTG

CGGGAATTCCTGGAGGACACCTGTSTGCAGTATGTGCAGAAACATATTTCCGCGGAAAACACGAAAGGGAGCCA  
 AACAAAGCCGCTCCTACACTTCGCTGG  
 TCCTGGGCGTCCTGGTGGGCRGTTTCATCATTGCTGGTGTGGCTGTAGGCATCTTCCTGTGCACAGGTGGACGG  
 CGATGTTAATTACTCTCCAGCCCCST  
 CAGAAGGGGGCTGGATTGATGGAGGCTGGCAAGGGGAAAGTTTCAGCTCACTGTGAAGCCAGACTCCCCAACTGA  
 AACACCAGAAGGTTTGGAGTGACAGCT  
 CCTTTCTTCTCCACATCTGCCCACTGAAGATTTGAGGGAGGGGARATGGARAGGAGAGGTGGACAAAGTACTT  
 GGTGCTAAGAACCTAAGAACGTGT  
 ATGCTTTGCTGAATTAGTCTGATAAGTGAATGTTTATCTATCTTTGTGGAAAACAGATAATGGAGTTGGGGCAGG  
 AAGCCTATGGCCCATCCTCAAAGA  
 CAGACAGAATCACCTGAGGCGTTCAAAAGATATAACCAAATAAACAAGTCATCCMCAATCAAAATACAACATTC  
 AATACTTCAGGTGTGTCAGACTTGG  
 GATGGGACGCTGATATAATAGGGTAGAAAGAAGTAACACGAAGAAGTGGTGGAAATGTAAAATCCAAGTCATA  
 TGGCAGTGATCAATTATTAATCAATTA  
 ATAATATTAATAAATTTCTTATATTTAAGG

Protein Sequence (SEQ ID NO:45):

MLTTLLPILLLSGWAFCSQDASDGLQRLHMLQISYFRDPYHVWYQGNASLGGHLTHVLEGPDTNTTIIQLQLQEPES  
 WARTQSGLQSYLLQFHGLVRLV  
 HQERTLAFPLTIRCFLGCELPPEGSRAHVFFEVAVNGSSFSFRPERALWQADTQVTSGVVTFTLQQLNAYNRTRYEL  
 REFLEDTCVQYVQKHISAENTK  
 GSQTSRSYTSVLGVLVGSFIIAGVAVGIFLCTGGRRRC

SNP Information

Context (SEQ ID NO:65):

CTGTGTGCAGTATGTGCAGAAACATATTTCCGCGGAAAACACGAAAGGGAGCCAAACAAGCCGCTCCTACACTT  
 CGCTGGTCCTGGGCGTCCTGGTGGGC

R

GTTTCATCATTGCTGGTGTGGCTGTAGGCATCTTCCTGTGCACAGGTGGACGGCGATGTTAATTACTCTCCAGCC  
 CCGTCAGAAGGGGCTGGATTGATGG

Celera SNP ID: hCV25620145

SNP Position Transcript: 721

SNP Source: Applera

Population(AAllele,Count): african american(A,32|G,6) caucasian(A,38|G,2) total(A,70|G,8)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:45, 20, (S,AGT) (G,GGT)

SNP Source: Celera;HGBASE;dbSNP

Population(AAllele,Count): no\_pop(G,2|A,9) total(G,2|A,9) ;no\_pop(A,-|G,-) ;no\_pop(A,-|G,-)

SNP Type: MISSENSE MUTATION

Protein Coding: SEQ ID NO:45, 20, (S,AGT) (G,GGT)

Gene Number: 209  
 Celera Gene: hCG37135 - 79000075741507  
 Celera Transcript: hCT28365 - 79000075741527  
 Public Transcript Accession: NM\_006404  
 Celera Protein: hCP48412 - 197000069371280  
 Public Protein Accession: NP\_006395  
 Gene Symbol: PROCR  
 Protein Name: protein C receptor, endothelial (EPCR);CCCA;CCD41;EPCR;bA42O4.2  
 Celera Genomic Axis: GA\_x5YUV32VYHA(29091543..29096965)  
 Chromosome: Chr20  
 OMIM number: 600646  
 OMIM Information: PROTEIN C RECEPTOR;PROCR

Transcript Sequence (SEQ ID NO:26):

CAGGAGGGAGGCCGGCCCCCTAGTAGGAAATGAGACACAGTAGAAATAACACTTTATAAGCCTCTTCCTCCTCC  
 CATCTCCTGGCCTCCTTCCATCCTCC  
 TCTGCCCAGACTCCGCCCCCTCCAGACGGTCCTCACTTCTCTTTCCCTAGACTGCAGCCAGCGGAGCCCGCAGCC  
 GGCCCGAGCCAGGAACCCAGGTCC  
 GGAGCCTCAACTTCAGGATGTTGACAACATTGCTGCCGATACTGCTGCTGTCTGGCTGGGCCTTTTGTAGCCAAG  
 ACGCCTCAGATGGCCTCCAAAGACT  
 TCATATGCTCCAGATCTCCTACTTCCGCGACCCCTATCACGTGTGGTACCAGGGCAACGCGTCGCTGGGGGGACA  
 CCTAACGCACGTGCTGGAAGGCCCA  
 GACACCAACACCACGATCATTCAGCTGCAGCCCTTGCAAGAGCCCGAGAGCTGGGCGCGCACGCAGAGTGGCC  
 TGCAGTCCTACCTGCTCCAGTTCCACG  
 GCCTCGTGCGCCTGGTGCACCAGGAGCGGACCTTGGCCTTTCCTCTGACCATCCGCTGCTTCCTGGGCTGTGAGC  
 TGCCTCCCGAGGGCTCTAGAGCCCA  
 TGTCTTCTTCGAAGTGGCTGTGAATGGGAGCTCCTTTGTGAGTTTCCGGCCGGAGAGAGCCTTGTGGCAGGCAG  
 ACACCCAGGTCACCTCCGGAGTGGTC  
 ACCTTCACCCTGCAGCAGCTCAATGCCTACAACCGCACTCGGTATGAACTGCGGGAATTCCTGGAGGACACCTGT  
 STGCAGTATGTGCAGAAACATATTT  
 CCGCGGAAAACACGAAAGGGAGCCAAACAAGCCGCTCCTACACTTCGCTGGTCCTGGGCGTCCTGGTGGGCRG  
 TTTCATCATTGCTGGTGTGGCTGTAGG  
 CATCTTCCTGTGCACAGGTGGACGGCGATGTTAATTACTCTCCAGCCCCSTCAGAAGGGGCTGGATTGATGGAG  
 GCTGGCAAGGGAAAGTTTCAGCTCAC  
 TGTGAAGCCAGACTCCCCAACTGAAACACCAGAAGGTTTGGAGTGACAGCTCCTTTCTTCTCCACATCTGCCCA  
 CTGAAGATTTGAGGGAGGGGARATG  
 GARAGGAGAGGTGGACAAAGTACTTGGTTTGCTAAGAACCTAAGAACGTGTATGCTTTGCTGAATTAGTCTGAT  
 AAGTGAATGTTTATCTATCTTTGTGG  
 AAAACAGATAATGGAGTTGGGGCAGGAAGCCTATGGCCCATCCTCCAAAGACAGACAGAATCACCTGAGGCGT  
 TCAAAAGATATAACCAAATAAACAAGT  
 CATCCMCAATCAAAATACAACATTCAATACTTCCAGGTGTGTCAGACTTGGGATGGGACGCTGATATAATAGGG  
 TAGAAAGAAGTAACACGAAGAAGTGG  
 TGGAAATGTAAAATCCAAGTCATATGGCAGTGATCAATTATTAATCAATTAATAATATTAATAAATTTCTTATATT  
 TAAGG

Protein Sequence (SEQ ID NO:46):  
MLTTLLPILLLSGWAFCSQDASDGLQRLHMLQISYFRDPYHVWYQGNASLGGHLTHVLEGPDTNTTIIQLQPLQEPES  
WARTQSGLQSYLLQFHGLVRLV  
HQERTLAFPLTIRCFLGCELPPEGSRAHVFFEVAVNGSSFVSFRPERALWQADTQVTSGVVTFTLQQLNAYNRTRYEL  
REFLEDTCVQYVQKHISAENTK  
GSQTSRSYTSVLGVLVGSFIIAGVAVGIFLCTGGRRRC

SNP Information

Context (SEQ ID NO:66):  
CTGTGTGCAGTATGTGCAGAAACATATTTCCGCGGAAAACACGAAAGGGAGCCAAACAAGCCGCTCCTACACTT  
CGCTGGTCCTGGGCGTCCTGGTGGGC  
R  
GTTTCATCATTGCTGGTGTGGCTGTAGGCATCTTCCTGTGCACAGGTGGACGGCGATGTTAATTACTCTCCAGCC  
CCGTCAGAAGGGGCTGGATTGATGG  
Celera SNP ID: hCV25620145  
SNP Position Transcript: 872  
SNP Source: Applera  
Population(AAllele,Count): african american(A,32|G,6) caucasian(A,38|G,2) total(A,70|G,8)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:46, 20, (S,AGT) (G,GGT)

SNP Source: Celera;HGBASE;dbSNP  
Population(AAllele,Count): no\_pop(G,2|A,9) total(G,2|A,9) ;no\_pop(A,-|G,-) ;no\_pop(A,-|G,-)  
SNP Type: MISSENSE MUTATION  
Protein Coding: SEQ ID NO:46, 20, (S,AGT) (G,GGT)

**TABLE 2**

Gene Number: 8  
 Celera Gene: hCG15848 - 146000220474494  
 Gene Symbol: ATF6  
 Protein Name: activating transcription factor 6  
 Celera Genomic Axis: GA\_x5YUV32W2NG(88209..298031)  
 Chromosome: Chr1  
 OMIM number: 605537  
 OMIM Information: ACTIVATING TRANSCRIPTION FACTOR 6;ATF6

Genomic Sequence (SEQ ID NO:67):

ACATCCAACATTTAATGATTTAAAAAGAAAAGCCTGGTAACTAGAGTGGAAGGAAGCTTCCTTAATCAGAGCAG  
 CTGTGAAAAATATATGGGAAAAAATC  
 TTATTTAGTGGTAAACTTTTCAAACATTCACATTAATAATTTAGAAGTAAGTTAAAAGATGTCTAGCCCCTACCTT  
 CTGTATCGTATTAGAGGGCCTAAC  
 CATTGCAGTAAACAAGGGGCTAAACAGTTTAAGACGTAGGAAAAAAATCTGAACTCACTAATAAATCTCGACA  
 TCTCTAAAAGTGGGACAGTATGTATA  
 TGCCTTCTGCTATGATGCAAGGAAGTACCTATGAAGTATTCTTGCAAAAGCAAACAAAAAACCAAAACTAATCA  
 AGCCTCTAGATGTCACTGGCAGTTT  
 CTGGAACAAATTAAGCATCACCATGAAGATGTAATCAGCAGAATCATGAAGTGGGAAATTCTATAGGACCAATG  
 ACTTAGATTCTTCAACAAATGGCAGG  
 GGGGAAGAAATGGAGAGGGGAATTGTCATGTGAACTAAATGCAATGTGACAATCTTGTTGGATCCTGATTTTAAC  
 TAACTGTGGAAGGAAGGGCATTTTTG  
 AGACAGAAAACCTGAACATGGAACAGGTATCAGATGATATTAAGGAATTATTAATTTTGTGGGTATGATAATGG  
 GATTATCAGTATGTTTTTTAAAAATA  
 AGAAAAGGCAGGCTAGGTGCAGTGGCTCATGCCTGTAATCCCAGCATTTTGGGAGGCTGAAACGGATGGATCA  
 CAAGGTCAAGAGATCGAGACCATCCTG  
 GCCAACATGGTGAAACCCCATCTCTACTAAAAAAATACAAAAAGTAGCTGGGCGTGGTGGCGCGCGCCTGTAGT  
 CCCAGCTACTCGGGAGGCTGAGGCAG  
 AAGAATCGCTTGAACCCAGGAGGCAGAGGTTGCAGTGATCACGCCACTGCACTCCAGCCTGGCGACAGAGCAA  
 GACTCCATCTCAAAAAAAAAAAAAAAGG  
 CCTTGTCTCTTAGAGATATACATTGAAGTATTTATGGGTGAAATAATATATCTAGATATGCTTTAAATACTCCAG  
 GAAGGCTGGGCAAGGTGGCTCACG  
 CCTGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGTGGATCACCTTAGGTCAGGAGTTCAAGACTAGCCTGGCC  
 AGCTTGGTGAAACCCTATCTCTACTA  
 AAAATACAAAAATTAGCTGGGCATGGTGGCAGGCACCTGTAATTCCAGCTACTTGGGAGGCTGAGGCAGGAGA  
 ATCGCTTGATTGTGGGAGGCAGAAGTT  
 GCAGTGAGCCGAGCTCACGCCATTGCACTCCAGCCTGGGTGACAGAGAGAGACTCCGTCTCAAAAAAAAAAAAA  
 AAATTCTGGGAAAAACACAGAACAAAA  
 TGAAAAAAGTTGTGGAGAGGAGAGGTAGGATAGATAAAACAAGAATGGCAAATTGTTTAAGCTGAAGCTGGAT  
 GATAGGAACACAGGGATTTATTATATC  
 ATCTCTGTATTTGTTTGAAAATTTTCATACTAAGTTTATTTTACAAATTAAACATTAAATAAAATAGAATTTAAA  
 AGGTTGGTGTTCATAATGTTC  
 TGCTAAATAAAAAATCCTAAAGAACTTTAGACAAATTTTTTTAAAGTCATTTTTATTTTATCACATACTAGTTTCCA  
 TCTACCATAATATGATTTTTTTT

TTTAGAGACAGAATCTTGCTATATTGCCAGGCTAGCCTTGAACACCTGGGTTCAAGCAATCCTCCTGCCTCAGCC  
 TCTCAAGAAGCTGGGATTACAGGT  
 GTGTACCACTGTTCTTGGCACAGACAAAATACTAGAATAACAGAATTCAACAAGGTTTTTGGTACAAGATTGCTG  
 TACAAAAATCCACAGTGTTGCTATA  
 CCCCATAATAGCCAAATAGAAAAAAAAAAGATGTGAAAAACAAAACCTATAAACTATGAAAATGAAAATGAAAA  
 TCTTAGTGACCTCAGGAGAGGGAAGG  
 ATTTCTTAAGAAAGCTACAAGTCAGCTATCATAACAACAGATTGATAAATTTGACTACATTGAAATTAAAATCTAG  
 AGAACAAATCACCGTAAACAAAGA  
 CGACAAGTGTCAACCTGGGAGAAGGTATTTATAACATATACCAAGCGTTAACTTAAAAAAGCCCAACAATGGGC  
 TGGGTGCGGTGGCTCATGCCTGTAAT  
 CCCAGCACTTTGGGAGGCTGAGGCAGAAGGATAGCTTGAGCCTAGGAGTTTGAGACCAGCCCGGGCAACAAAG  
 CGAGACTCTGTCTCTCTCTCTCTTTA  
 AAAATGTATATATTA AAAAGCCCCACAATGAATAAATTTAGAAAAGAAGGGGAGACTTTATGGCCTGCAAGGTG  
 GCCATCCTGCAGGCTGGGAAGCGTGC  
 CTCCAGAAACAGGCGCTTTGAAGGAGGAGGGCTTGGAGTGTAGTTCAGGCTGAACAGGTTGGCTAAACCTACA  
 TTTTCACCAGGTTACAGGAAGAGCTAT  
 GAATATTCATGAAAGTGGTCCTGACACATGCGTATTGAACAAAACTCATGTAACATATGACCCATATTCACTTT  
 GGGATGGAGACTTAACTTTTTTGT  
 GTTGTTGTTGACGAGTCTTGCTCTGTCACCCAGGCTGGACTGCAACGATGTGATCTCAGCTCACTGTAACCT  
 CCACTTCCCGGTTTCAAGCGATACT  
 CCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGAACCTGCCCCACGCCTGGCTAATTTTCGTATTTTATAGTAG  
 AGATGGGGTTTCCCATGTTGGCC  
 AGGCTGGTCTCAAGCTCCTGACCTCGTGATCTTCCACCTGGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGC  
 CACCACACCCAGCCAAGACTTAACA  
 TTAAATGTATTACGAGCCCTATACGTCAAAAGGTCTTTCAGGACATGAAGCCATGCAAATGTGCGACCTCTGTA  
 GACCAGCCAGAACCAGTCCATGGCC  
 AGTGGTCTTCTTACCTGGAGAAAGTTACTGAAATCAACAAAAAGTTGTCCAATGAAAGCTGTAGCTATGGCTGGT  
 GGAACGGGGACTGGGGGTGAGTAAG  
 TCAGCATCTCGTGGAGCTGCACATTATTTAATATTGTTTATCTTTAGGCCAGTACTTGTTTAGCTGCTAGAGAAG  
 AATCTTGTGGCAGTTAGAACACAG  
 TTTATTCTTTAAGTATAGGGGTGTGGGACTAAACCTTGCCTGGAATGGCTTTAGGTCTTTTTTATAATTTGGCAT  
 CTTATTGCCAGAGTCTGTTCTGTC  
 AGTCTTATCATCTCTGTTTTAACATTAATGCTGGCCAGTTGTTGTGTTAACTACAAAAGGGAGGGAGTATAACA  
 AGGTGTGTCTAACCTGTCCCATCGT  
 GGTCTAGAACTCAGTTTTAAGGTTTTCTAGGATCCCCTTGGCCACAAGGGGGCCTATTCTGTTGGTGGCGGACTT  
 AGGAGTTTAGTTTTACTTTATACAG  
 GTGAATGAATAGAGCAGGCTATCTAATGACGTACACATCATTTGTGAAAGAAAACCTCATAGCAAACCTTTATAAA  
 AGGTAAGAGTGGACAGTTCACAGAA  
 GAGGAAATAAGGTTTCATAAGTATGGAAGAATGCACCACTTTGTTAGGGAAATGTAAATTGCAATTTTCACAGC  
 GGTA AAAATGGTAAGAGTCTGACCAT  
 ATCTTGTGTTGGAGAGAATGTGGAACAAAGAGAACTCATCAATGCTAGTGAGGGTGTAACTGATAAAATAGTA  
 CATAATTATTTAAGACATTCTTGAG  
 GCTGGGTGCGGTGGCTCACGCCTGTAATCTCAACAGTTTGGGAGGCAGAGGCGGGCGGACTGCTTGAATTGAG  
 AAGTTTTGAGACCAGCCTGGGCAACAT

GGTGAAATCCTGTCTCTACAAAAAATACAAAAATTAGTCGGGTGTGGTGGTGGGTTCCCGTAGTCTCAGCTATTG  
 TGGGGAGAGGCATGAGGATCGCTTA  
 AACCTGGGAGGTCAAGCTGCAGTGAGCTGAGATGGCTCCACTGCAATCCAGCCTAGGGGACAAAGTGAGCCCC  
 TGTCTCAAAAAAAAAAAAAAAAAAAAAA  
 AAGAGGGAGAGAGAGAGAGAAATTGAGATTCATTCAAGGAAATACCATGGGTAGATCCTTAATATCCAATAAG  
 AAAAATGATTTGGAGAACATTAAAGTT  
 TACATACATTAAAAAAACCCAAATGACACTAAATGCATGTGTCTATAAACATGTTTAAGCACCGTAAAACTAAAA  
 ATAAAAGTGGACTAGAATTACACTT  
 TTATTTTATAAAATAAAATTTTATACCCATAAAATTCATGGTAGAAGTGGCCTACGCCAAGGGAGGAAAATGAA  
 GCAGGGAGGAAAGAGGAATTCAATT  
 CTGTAGTATTTAATTTCTTATGAAAAATCTGGGACAAAAATGAAAAATGTTAAATTTTAAATTCTTGGTGATAGG  
 TCACAAATGTTTGTTTCATAATTA  
 CTTGCACTTTTTGAATAAAAAATAAAATCTTTTGAAATTAAAAAAATTGGCGATTTGTCATGGTGATTTTCTTAA  
 AGACTAAGGTGTCTAGTGGGATAC  
 AGAAGACTCGATGATTTTATTTTATTTTAGGTTTTTTTTTTTTTTTTTTTTTTTGGAGACGCAGTCTCGCAATGTCGACC  
 AGGCTGGTGTGCAGTGGTGCGA  
 TCTCCGCTCGCTGCAACCTCTGCCTCCCGCGTTCATGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTAC  
 AGCTACTTGCCCGCCACCACGCCT  
 GGCTAACTTTTTGTATTTTATGTTGGAGACAGGGTTTCACTATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCGT  
 GATCCGCCCCGCTCGGCCTCCCAA  
 AGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCCTTATTTAGTTTTTATGTCAATTTTGAATTTGTAAT  
 TGGGTGAGAATTAAGTGTTATAA  
 ATTTACCTTAAGTGAAAATTTATTAGTTTTCTAATGACAAAAGTAGTTACATCTGACGTAAGGGGATTTAATTAT  
 CTTTTCTGGGCTTGCGTCCCGGA  
 ACGCTGCTGCATTATGTAATCTGTAGTAATATACCCACGAGCAGTTTGGAGGGTTCTGGGAAGCACATTTGTTTT  
 GCTTACCTACCATGGGAAAGATTCC  
 ACGTGGTCCTAGAAGCATACGTATTCGTGCGTTACAGTATTTTAAATATTCAGTGGCCGCTGAAATTTAAATGG  
 ATGAAAAAGAAATTCTCTTAAGTAG  
 CGTTCTTGAACTGCTTTATCAAACCTCAGTTCATATTTTCAGTTGGAGTTCGTGATGTATGTGAATTTCTTTAAAT  
 AAAACCCGAATCTTTAGTAATAC  
 AAATATTGGGAACCTTCGATTACTGTAGGTCTTAGGGTAGACTCGCTTGGACTTTGACTAAATACTATGCTTTAGA  
 AATTTATCTTTAGGAGATAAATTTG  
 GGGGGTTTATCTTTTTATATTTGCTTTGTTTCAATTTATTTAAATAAAAGTAGTCTTTCTAGAAGTATGTTTTATAG  
 TACCACTTACCTCGGTTACTGTC  
 CCAGTTAAGCAGTTTTACGAATGTGGAAAAATAGTAGTGTTTTCTTTCATACACGCGTATATGAAAAATTTTAA  
 CATTGATAGAGAATTATTCGTAAA  
 AATTAAAGTAAATTTACTGTTAGTCTCTAATTTTGTTCTGAGATAGCCACGCTGTGGCATTAAAGAAGGATAATCAA  
 GGAATAAGCCATAATGTTTCAGAAA  
 GGGGCTTTACTTTTTATTTTAAATATTATTGATAAACTTTGTTTAGTCGAATTGATGTCTGCGTGTCTTCCCCGCC  
 TTCCCATAAAACAGCGGGACAG  
 AAAGTCCTGAAAACCTCCAAAAGGGAAATATTTACTCAGAGGTCTATTCAAATTTGCTTAGTCTCTGACGTTTTCT  
 GTGTCGTACAAATCCCTCGTGCGG  
 CARAAGTAGTTTGTCTTTAMTAGGCCACCGTCTCGTCAGCGTTACGGAGTATTTTGTCCGCCTGCCGCCGCCGTC  
 CCAGATATTAATCACGGAGTTCCAG

GGAGAAGGAACTTGTGAAATGGGGGAGCCGGCTGGGGTTGCCGGCACCATGGAGTCACCTTTTAGCCCGGGAC  
 TCTYTCACAGGCTGGATGAAGATTGGG  
 GTGAGTGGGATCTGAGAATGTACCAGKGTGGCTCGGGTTYGCGTCTAAAGGTTTCTTCCTCCCTACTTCCGCCCA  
 CTCGTGGTGACAGGTGTGGACCAAC  
 CCTGCCTGGGGAGTGAGGCCCTCGTCACTCCTGTTTCGTGYCCCTGGAAGAGTCTGGGAGTATCGTTCGGCTTCT  
 GTCCTGGCCACAGGGAGAGTAGAAT  
 ATTGTCTAACTGGAGTTAACTGTCGTTTTTCACTATTCCCACTGGGGAAGTACCGGGTGATCTTTTTGTGTGGCTA  
 ACACTGGACGTCTTTAGGAGGTTT  
 AGAGCAAAATGGCTGGGACCGCCAAAAAGGCTACAAGCTATTTCTGAGTGTGAAATGTGGGCGGTGAGCACA  
 ATGGAGTGGGTTTTTGTGTGTTTGGTG  
 TATATGAGAGAATGTGGTTATATACAAGAAATTCTAATCTTACAGCTTTTACTATTGGTTTAAGCCTGATGTCYGT  
 CTTGGCAGAAAAGCAAAACACCCA  
 AGCACCTGACCTCGGGTCCCTGCTTAGCTACACCTAATACTTATGGTGTAATCCTGACTGCTCTCCAGATACAGCT  
 CTTAATTTTCTTGTTAATGGATAT  
 TTTGTTTGGTTTTTGGTGGTCAGGAAGGGTGGAAATGGGATGAAGTGGAATGAGAATGAAGAACTTTGTTATCT  
 AAGTATTTAGAGGAGACTTTTATTCT  
 CCTCTTGATTGCAAGATATAATAGTAATATTGTTCTTAGGTTCCAATATGAATTTGGGATTGTGTAGGTTCCCAA  
 TGGGCCCCGCTAGTGAGCCTACACC  
 ACTCATCAGCCTTGGAAGTGGGGGGAACCCAGTTTATCTTTCTCCCTTGGATCAGCATTCTCAATAGCAGGAGTG  
 GAAGGAATCTCAACTTTTTATTGAA  
 CAAAGTCATTTAATTGTTTCAAGGACAGATGTTCTCTAAACTATATTCTATTGGAGATTTATTTAATCCTTGGGT  
 AGATTTTTTAAGAGTTGCTGGAA  
 AACTTAAGATTTTTTATTCCCATTTTCCCAAGCCAAATGCATTAGTGACAGTAACCATTTATTTAGCTCTTTGTGTC  
 AGGGACCCTGTAAACTCTTGAC  
 ATGAAGTATTTTGTTTAATCCTACCAACAGTCCAATGAGGTACTGTTTATTTTAGAGATGAGGAACTGAGGCTA  
 ACAGAATTTGTGCTCAGATTCTTAG  
 AGCTAGTTAGTGGCAGAAGAAGGATCAGACCAGGTTTTTCTACCTCCAAAGTCTGTACCCCTAGCATCTATTCTA  
 TAATGCTTTCCTAACTCATGTCTAC  
 AGATTTTATTCTTCTAAGATCCCTGGTTATTAATAAGACTACTTGAATATATTGGGTAGTTATGATTTATTTTGTG  
 AGCTAGTCAGCTATTATATTTTCT  
 ATGACATTTCTTTTCTTTACCTCTTTTTTTGAGATGGAGTTTCGCTTTTGTTGCCAGGCTAGAGTGTAATGGCGCA  
 ATCTTGGCTCACCGCAACCTCCG  
 CCACTCGGGTTTGAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCACCACCACGCCCG  
 GCTAATTTTGTATTTTGTAGAGAGA  
 CGAGGTTTCTCCATGTTGGTTAGGCTGGTCTGGAGCTCTGGACCTCAGGTGATCCGCCACCTCGGCCTCCCAA  
 GTGCTGGGATTACGGGTGTGAGCCA  
 CTGTGCCCCGCCAGTGACATTTCTTAAAGTATAACTGTGGGACACTTCTGGTAGCATAATTTTATGTCGTCTTCT  
 GATACCCAGGCTGCATTTTCCAGA  
 CTTATCTGATCACTTGACTTAATGTTCTCTRACAGAGCATTGAGTTCTTTGTAGTCAGTTCACCGACTCACTTCTAT  
 CGGGTATCCTGCCCTGTCCCTCC  
 CTGAATGTGCCCATTTCCATTGTTTTCCAGGTTTCCCTTTTCAAAGTGTAAACAGTCTTCTCTTTGAAGACTTTA  
 ATTTAATTCCTTAGGTGTTTCCAG  
 TAAGTMGGATAGATTTTGTGGCAATAGAGTGAAACACTACTGTCTGATAGTGTTTCGTTCTGTGACAGTTGCTG  
 TTTTGGATGTTCAAGTACTTCACTG

ATTTACTTGATTTTGTGTTGTTGTTGTTGTTGTTTCTTTTAGAGACAGGGTCTTGCTCTGTCGCCCAGGCTGGAGTGC  
 GGTGATGCAATCATAGTTCACTGT  
 GACCTGGAACCTCTGGGCTCAAGTGATCCTTCCATCTGAGCCTCCTGAGTCGCTAGGACTACAGGTGCACACCGC  
 TACCATGCCTGGTTAATTTTAAATT  
 TTTGAATTTTTTTGTAGAGACAGGGGTCTTGCTGTGTTGCCAGACTGGTCTCCAAGTGGCCTCAGTCATCCTGCCT  
 CAACCTCCCAGGTGATGGGATTAC  
 AGGCTCAAGCCACTGTGTCTAGCCTTGATTTTTTTTTTAATATTAGACACTATCTAGTTAAAATTTAAAAATCAGATT  
 ATGTAATTAGTATGTCCTGTGAT  
 TTTAACAACCTCTAAATAATTTAAATAAAGTGATTAACCTCAGAACAGCCTGTGAGGGATGTGCTTTCCTTTTTCTTTT  
 TAACCTAGGTAATGAAATTTAGA  
 GCAAGAGTCTTCAAACATTTTCTGTAAAGGTTCAAGATAGTAAATATTTTAGGCTTTGTGATTACATGGTCACTGTT  
 GCAACTACTCAAGTCTGCTGTGAT  
 AGTGCAGGAGCAGCCAGAGACAGCATAAACAAGTGAGCATGGCTGTGTTTCTATAAAGCTTTATAGACACTGAA  
 ATTTGAATTTACATAATTATTTTAA  
 ATAATAAGCCTACTTTTTTAGTATAATTATAGATTTACACAATAATTGAGCAGATAGTACAAAGAGTCCACATACT  
 TCCCCTTCATCCCCAGCTTCTCCT  
 ATTGGTAACATCTTGCATTAGTGTGGTACATTGGTTACAATTGATGAGCCAGTATTGRTACATTATTATTAATACTAA  
 GCCCATCATTTAAAGTTCACTCTG  
 TTGTACATTCTATGGCAGGGATCCCCAGCCTTTTGGCACCAGGGACTGGTTTCATGAAAGACAATTTTCCACAG  
 GAGGTGGGGACAGGGAGGCACAGGA  
 AATGTTTTTCAGGATGATTCAAGTGCATTACATTTATTGTGCACTTTATTTGTATTGTTATTACACTGTAATATATAA  
 TGAAATAATTATATAACTCAACA  
 TAATACAGAATTAGTGGGAGCCCTGAGCTTGTTTTCTGCAACTAGATGGTCTCATCTGGGGGTGATGGGAGAC  
 AGTGATTAGATCATCAGGCGTTAGAT  
 TCTCATAAGGAGCGTGCACTCTAGATCCCTCACATGTGCAGTTCACAGTAGGGTTCACACTCCTATGAGAATCTA  
 ATGCTGCTGCTGATCTGACAGGTCTG  
 TGGAGCTCAGGCGGTAATGCAAGTGATGGGGAGTGGCTGTAAATACAGATGAAGCTTCGCTCACTCACCCACCT  
 GCTGCTCACCTCCTGCTGTGTGGCCT  
 GGTTCCCAACAGGGTTGGGAACCACTGTTCTATGGGTTTTGACAAAGGTATAATGACATGTATCCATAATTACAG  
 TATTATATAGAATAGTTTCACTGCC  
 CTAAAAAAGCCCTATGCTCTACCTATTTTTTCTCCTTCTCCTTCTCCCTACCTCTGGTCAGGTAACCACTGATCTTTTC  
 ACTGTCTCCATGGTTTTGCCTT  
 TTGCAGAATGTCATATAGTTGGAATCATACAGTATGTAGCATTTTTGAATTGGCTTCTTCACTTAGCAATATGCA  
 CTTAAGGTCCCTCCATGTCTTTTC  
 ATGGCTTGATAGCTCATTTCTTCTTAGCACTAAATAGTATTACAAAAATACTATTTACAGTGTTACAAAAAATAGT  
 ATGTATGTTCTATAGCCTATTAGT  
 CTGCCTAGGCTGCCATAAGACAACACCACAGACTGCGTGGTTTAAACAATGGAAATTTATTTTCTCCAGTTCTG  
 GAGGTTAGAGTTCAAGATCAAGATG  
 GCCGCACCTTCAGGTTCTGGTTAGGGCTCTCTTCTGGCTTGCATGCACATGGCTGCCTTCTTCTTGTGTCCTCAC  
 AYGGCACTGAGAGAGAGAATCTTC  
 TTCTTCTTATAAGGCCACAGTCCTATCAGATTAGTGCCCCACTTTAATGACCTCATTTAACTTTAATTACCTCCTAA  
 AGACWGTATCTCTCGATATAGTA  
 ATATTTGGGGTTAGGGCTTCAACATACGAATTTTGGAGGACACAGTTTAGKCCATAAATTCACAGTTTGCTTATCC  
 ACTTACCTATTGAAGCACATCTTG

GTTGCTTCCAACCTTTTGGCAATTATATATAAACTGCTGAAAACATTTGTGTGCAGGTTTTGTGTGGACATGCAT  
 TTTCCAATTGTTTGGGTAAATATC  
 AAGGAGCACAATTGCTGGATTGTGTGGTAAGAACATGTTTTTGTGTTTGTGTAAGAAATTGCCAACTGTCTTC  
 CAAAGTGGCTATACCATTTTGCAT  
 TCCCGCTAGCAATGAATGAGCATTTTTGATGCTCCACATCCTTGCTAACATTTGGTGTGCGCAGTACTTTGGATT  
 TAGTCATTCTAATAGGTGTGTAGT  
 GGTATGTTTTAATTTGCAGTTTCCTGATGACATATGATGTTGAGCATCTTTTCATATGCTTATTTGCCGTCTGTATG  
 TTTACCCTCCTCTCTCCTTCTC  
 TCTTCTTTTTACTCTTGCTTGAATTTGTGCAAAAGCTGCTTTATTCTGCTTTCTTCTTTCTTCCCTCCTCCTCCTCCT  
 TCTCCTTTCTCCTTTTTCTCCT  
 TCTCTTTATTCTTCTGTCAGCTGTACTTAAGATATGTGTATCAAGATCAGAAGCTTTCGTCTGCCACTGTGGTTTTCG  
 TTTATGTCTTCAGTCTCTTSTGT  
 GAAGGTTCTCTTGGCTTTCTAGCACTAAATCAAAATCTGGTGCTCCTGGATGACACAACTTCCCGTAATGCCCCG  
 GTGAACGGTTGCTAATTTTATTCTC  
 AAAACCTTTTCACCATAGGGCCTCTGTTGGTGGGATAGGTGYCTCCATTGTCCTTCACTGCTATTCTCCTTTCTCCTC  
 TCCCTAAAAATATCTAAGTTGAA  
 ATCTCATTTTGTGACAATATCCCATCCTCTACCTCTCCTGATTCTGATCCTGGATCACTCTCTTTCACCTTTTCTTTTCTTTT  
 AGAGATGGGGTCTTGTCTC  
 TTGCCCAGGTTGGAGCTGGAGTATGCGGTGGAGCAATCACAGCTCACTGCATCTTCAGACCCCTGGGGTCAAAG  
 GATCCTCCTGCCTCAGCCTCCCGAGT  
 AGCTGGGACTACAGACACATGCCACCATGCCTGGCTTCTTTTTGGCTTTTTGAAAACATTTTGCCACTAGCTTACT  
 GTTACTTTCTCACCATTACTTCC  
 ATCATGATTCTTGGATGATTTCAAGTATCCTTAGAGAAGATCCTTCAGATACACTGACCTGTCATTTTCTGGACCTC  
 CTCTTTATCACTGATCTTACCTT  
 TACCCAGTCTCAGCCATTTGCTTTCCTGATCATAAGCCTAAGTCTTGTCACTACCAATAACATCTCACTTTCAAACA  
 TCCTGCTCTGACTGCCTTCGCCT  
 GTTTTACCAGCTTACTTCCTTTAGTATTCCAGTTTCAATAGTTGTTCAAGTCTTCTGAGRTTTTGCCATGTTTCCAGT  
 GTCTCTCATTCTCTTCAGTTCA  
 TCACGTCCCTCTTACCCAGTTTAAATACCATGGACAGTCATTGTCACTATTTCTTCTTTCACCTTTTCATATTCATG  
 TCTGTCTCTTTTCTGGTAGGTT  
 TCTCTTGGATAAACCTATTAGTTAAATTCAAGTTTTGCCACTTCATGCCTGTCCTAGTATAGCTGAATCAGCTGAA  
 GAAAACACTTCAACTGTCCTGACT  
 GGTCTCAATTTAAGTGTTGCCAGACCAATCATATTTCCCTGGTCCTCTCACTGTCCATTCTGATGATTGCTTTATA  
 AAGTTTCTTCTATCTTCAAACCT  
 CCAACATCTCTCCCATCCTTTGAATCAGAAGAGAACTTCTACAGTCTCCCACTACCTCATTTACCCACCTATCAACA  
 TGTGCTTTCTCTTTTGTGGT  
 GAGAATGAGTGGCTTCTTGCTCCTAGCTAGAGCCAGTCCTTCCATATGTGCTTTAGATTCTTCTGTTTTGTTCAA  
 GAATATTGCTCAAGCTATTCTTCC  
 TCCTGTTTCTGTCATCAGCATTTCCCTCTCTACTAGATCATCTCTGTCAGTAAATGAACATGTTGTTGTTTCTCCTA  
 GAAGTACTGTTTCTATATCTAG  
 ATAGTACTCTAGCTAGAGTTAAAAAACAACCAACAACTCAAACTCTCTTCAGATGTTACTTCA  
 CCTCTTGCTGTTACTGCATTTCTCT  
 ACTGTATGTTACAGCAGAATTCCTTTTAAAAAATGTCCACGTTCACTGACTCTAGTTTTTTATTCTCCTTGTCTCTT  
 GAACCCAGTCCAGTCAGGTGCC

ACCACTCCACTGAAACTCCTGTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGCTCTGTTGCCAGGCTAGAGT  
 GCAATGGCACGATCTTGGCTCAC  
 TGTAACCTCCACCTCCTGGATTCAAGCGATTCCCGCCTCAGCTTCCTGAGTAGCTGAGATTACAGGCGCCTGCCA  
 CCATGCCCCGGCTACTTTTTGTATTT  
 TTTTTTTTTTGAGACGGAGTCTCTCTCTGTTACCCAGGCTGGAGTGCAGTGGCTCACTCGGCTCACTGCAAGCTCC  
 ACCTCCCAGGTTTACGCCATTCTT  
 CTGCCTCAGCCTTCCGAGTAGCTGGGACTACAGGCACCTGCTACCACACCTGGCTAAATTTTTTTTTGTATTTTA  
 GTAGAGACGGGGTTTCACCGTGTT  
 AGCCGGGATGGTCTTGATCTCCTGACCTCATGATCCGCCTGCCACCTTGGCCTCCCAAAGTGCTGAGATTACAG  
 GCGTCAGCCACTAATTTTTGTATTT  
 TTAGTAGAGATGGGGTTTCACCATGCTGGCCAGGCTGGTCTCAAACCTCCCGACCTCATGTGATCCACCCGCCTCG  
 GCCTTCCAAAGTGCTGGGATTACAG  
 GCGTGAGCCATCATGCCTGGCCTGAAACTGCTGTTTTTAAGGGCACTACTGATTTCCACATTGCTAAACCCAGTG  
 ACTGTGTCTGATCTTATTAACTTA  
 TAAGGAGTAAATGAATAAATAAATGGGCAAGGTGGAGGTGAGGTGGGAAAGTTTGTTGAGTACACTAAAAATA  
 GTAACCTACCATTATTTACTACTGTTGA  
 GGGAAATCTAATAGTGTGGGACTGGTGCAAAGGATATAGAAATAGGCAATAGGTTTTAGGATACTTTGGGGTA  
 ACTGATGCTAACAGTTTCTTATGGATT  
 CTTTTGGAATGTTCTATAAATATACGAGTATATTTATCAACCCACCTTCAGTTATTTTTTTAACCACAGAAAAGTCA  
 AAAGATCAGTACACTGATAACG  
 TGCGTTTTCTTATCTAGATTTACCAACTGTTAACATCGTGCCACATTCACTTTATATCTTAATATTACCCACACTTT  
 CCTCCTTTTTACATCTGAAAG  
 TGGGCTGCAGATAATATGAAACTCCTAAACACTTCACCATGTATTTCTAAGAATGAGGAAATGTTTATTTTTAAA  
 AAGCATCAATTCATTTAATATGC  
 TTTATACTTCAAAGCTAACATCACAGGGTTCTTTTTACCTTCTTGCACCATATGCTTGTATTTCTTCTGCCGCTA  
 TGAGAATCCTGATTCCCAATAA  
 CATCCTTTGTTTTATAGGCTCACTTCAGCAGCTCTGTTGGAAATAAACTGAAAGAGGGGCAAGGGCAGAAGCA  
 GGGAGACGTTTTGGAGGCTGTTATAG  
 CAATCCCCATGAGTTACAATAGTAGCAGTGGAGTAATGAGAAGTAGTTGGATTATGGATATGTACACACACACA  
 CACACACACACACACACATACACA  
 CATATATACATACATTTTAAAATTTTGAAATAGTTTGAAATCTATAGAAAAGTTTTAAGAACAGTGCAAAGAAC  
 TCTTGCATGCCTTTTACACAGACT  
 CCTCAGTTGTAAATTTTACTGCATTTACTCAATTTCTCTTTCTCCCTTCAAGAGAGAGAGATGTGTTGACCCA  
 TCACCTTTAAATACTGTAGTGTA  
 TACTCTCTATCCTCCTGCCTCCCAAAGAAGGCCACTTTCCTCTATTACTACCCTATATTTCTTCCAGTCAGGAGAT  
 AAACATTGATACAACAGTGTCAT  
 CCAATTCAAAGACACCATTCAAAGTTTGCCATCTGTTTCAATAATGATTTGCTTTTCTAGTCTAGGATTTTACTTGC  
 AAACACTCATTGCTTTTACTTAT  
 CATGTCTCATCAGTGTCTTTAATCTGGAACAGTTCTTTAGTCTTTTCTGTCTCTAATGTCCTTGACAGTCTTAAA  
 GAGTAGAAGTCTTTTATTCGACA  
 AGATGATCCTAAACCACTAAGGGGTTCTTGGAATAATGACCAATTCCAGGGCTGGGAAAGTGACAGGGAAAGT  
 GCAAAATAAGCCTGGAAAATCTTGGGC  
 CAGAAATTCTCAGCAAATGAACATGTTAAAAAGAACAAAGGCCCAATATGAATGAACTCCCCGTGGTCAAATCT  
 GGGAAAATTTGGCATCAGAATAAAA

TGAGGGTATACATTATAGTGTAATACTATGCGATGTTCTATGCCATCATATCATCCTGTGATATAGTCACATGA  
 TCACTTGGTCAAGTTGGTATCTTC  
 TAGGTTTCACCACTGTAAAGCCACTGCTTCCCATTGTTTATTGCTATTTTGGGGGAGATAATCTGACATAA  
 CACCAATATCCTTTTCTTATTA  
 CCTCTAACCACGAGCCCTAGCCTCATTAATAACCCCAATCTATATCAGTTACCACTCTAATGGTTGCCAAATGGTA  
 ACTATCTAACTCTATCATTCTGTC  
 TATCATTCTATCATTGTTGATTGACATTGCACTATAAGAACCAACCTACCTACCTACCTTCCTTCCTTTGTTCC  
 GTGATGAACTCATTGATTCCTT  
 TTTTATTTGGTGGTTGGTAAACCGCTACTCTCATTTTATTCTGATGCTCAGATTTTCCCAGATTTGACCACTGAGAA  
 TTCATTCATATTGGGGCCTTTGT  
 TCTTTTGGACATATTCCTGTCATTTGCTGAGAATTTCTGGCACAGCAAGATTTCCAGGCTTATTTTGCATTTCCCT  
 GCTGCAAGCCCTGGAATTGGTTA  
 TTTCTCCAAGGAACCCCTTAGTGGACTTCTGTTTGGTGGAAGATGGTATTTAGTATCCAAGATCTGGTCACCCAG  
 TATTTTCATTGCTACTGGGATGTCA  
 TTGCTTCTAGGACCTCTTAGCAGACAGAACTAGGAAATATATATACATGTGTGTATATATATATATATGTGCATGT  
 GTATATATGTATATATGCATAATA  
 CAAACATCTGCATATGTGTGTACATATATATATACTTCCACAGCTATCTGTATGTGTTATATTTTAAATGAGTTTAT  
 TTAATAATACTTCTAGTTCTAAT  
 CCAACATCTTAGGATTCTTTTAGTCTCCCGACTCTTCCATGCTCTTAAATCCTTTCTCAGACTGTGAGAAATCTGA  
 CTCCAATTATTTTAGTAAAAGGC  
 AAAAGATTCATTCTGTTCTGAAGAGAAACCAGAGCATTCTAATTATTTTAGATATATTCAACCATTTGCTCAGTCAC  
 CTAACATATTTGTTGAGTATGACC  
 ACTCTCCTAACCCTCAGCCTGCGTTTTCCATCTGCTACTCCATTCCCCCTCCCACTGGGCTGGGGCATGCTGGCA  
 TCTTTGCTGGGCAATAGTCTCTCC  
 CACACCATCCTATTTCTTGGATATGAGGCATATTGCTGGAGATGACTGTGCACAAAACCCCATCACCTTTGTTGC  
 CCTGCTTCTTTGGAAGCTGACATC  
 TCTGTGTTGGGAGGGGTAGGGAAGAAAAGGAAGGTTGGGAAGGGAAGGAAGAATGGGAGGAAAGGTAGGA  
 AGCCACATTACAGATATATATTTTGAAGAAC  
 AGAGCTGACAGAATTTTCTGGTAGATTGATTATGGAGAGTGAGGAAGAGAAGTAAATAATTAAGTAATAATGAT  
 CCAAGGTTTTTGGCCTGAGCAACTG  
 GGAAAATCCAATTGTCATTAAGTATGAGGACAAAGCAGTAGGAGGTTTGGTGGGGTTGATGAGGAGCTCA  
 TTTTGGATATGTTAGGTTTGAATGCC  
 TTTTAAACATCTAAGAGGAGATGTCATACTCACTGCACATGCATCTTGAAGTGTGGGTTGGAGATGTATATTTGG  
 AAATCCTTGGTATATAGATTTTATT  
 TAAAGCCGTGAGCTCAGCTGAGATTACCAAGGAAGTAAGATAGAYAGAAACGTTTATGCTGAGCCCAGGGACA  
 AATTTGTGTTAGGGAGATGAGGAAGAA  
 CCAGCAAGGTAGGTCAACCACTGAGGTGGGAGGAAAACYGGGTAAGTGTGATATTCTGGAAGTCAAGTGAAGA  
 AAATGCTTCAGTTAAAATGAAAGGATC  
 AACTATGTTAATGCTGTTGATAGGTCAAGTAAGATGAGGACTATGAGTTTACCATTGAATTTAGCAATATGGAAG  
 TCACTGGGGACATTAATAATTATGT  
 GTATGTGGAGCGGGGAGCTTGATTGGAGCAGGTTTAGAACGAATGGGAGGAAAGAAATGAGAGAGCCAGATA  
 GAGAACTCTTTAAAGGAGTTTCAGCCTA  
 TATATAAGCCGCAGTCCTCTGCTTATATGTTTCTTGGCTCTGCACCTGAGTAGTCTTTTCTGAGAGTCATGTGAGT  
 AACCTTTTGTATTACTACGTATA

CATTCTTTTGCTAAATACTATTTTGTCTTAGGCGGTGAAATGTTACTGTAATTGGTCAGGTTAATAACAGACTGAC  
 AGTGCAGTTCCTACACTTGTGCAC  
 TGAGGAAATCATCAGAGCAGTTTCAGCAGCATAAAATAGGTAGAGAGTACAAAGGAAGACCTAGTTTTGAAAA  
 ACTCTTCAAAGTGCTTAAAGTTTTGAA  
 AGTGTGAGAGGAAATAGGAACTTCATAGATTCTATGGTGAGGAGCCTTCTTGGGCCTTGGTCAACACCTTAACCT  
 TTCTTTACAAAAAGTTTTCAGTATA  
 GTGTATTGAAGAATGTTTTCTGTATCCACATATTTCTTGATTTGTAATGAAACCTCATTAGGAAATTGGAATATAG  
 TTAGATTTGCTGAAATGAATTTGA  
 TGGTTTAGATGAATGAATCCTTTATTGTCTAGATAGAAAATTGGTTGCACCCTTAACCTAAGGCATGGATAACAA  
 ATACGCAGTGTGTATATACTTTCCA  
 CAGCCCTCTCCTATTGCTTATGAACCATGTCTACTTTCCACAGTTGTCTCCTATTACTTATGAACCACATTAACCTCT  
 CATGCTGAACCCAGATGTAGCA  
 TCAGAATTCTTGTTAGTAATGTTCCATAAAGCCACTACCAATTAGTCTTACATGACATATGACATAAAACCTTTTTT  
 GATCCCTGGTGTAAGAATTCAGT  
 TCACAAGATAAAAATTATGCAGAAAACAACACATAAATTGGAAATGAGCATTGAATTGCTTATTCACATTACTGG  
 AGTTTATCAATAATAGTAATGGCTA  
 ATGAGAAACCTTAATAGAGGTGACATAGGGACACAGTGCTTGTTTCTTTGTCAAATAATTGAATTGACAGTTCAT  
 TTCTATTCTTTTCTTTGTCCAAAG  
 ATTCTGCTCTCTTTGCTGAACTYGGTTATTTACAGACACTGATGAGCTGCAATTGGAAGCAGCAAATGAGACGT  
 ATGTAAGTATTTACTAAGGTTGAAT  
 AATGTGATATTTAGTCTTTAACCCTAATTATTGCAAGAAAGTTTCAGGACAACAGGTTGAGGCTAGTAATTTAAGT  
 TACATACTTAATGGTAAATAGCTT  
 CTACCATTAAGTATGAAAGYTAGTAGCCATTTTGTAATCAGTAAATGTAGGACAGTAGAAATTGCATTGTATTGC  
 AATTCTAGGCCACCTTTGGCCCTA  
 GGGATGGGTGACTTTGGAAAAGTCACTAATCTGCCTACCTGGTTTTTACAGATACTTTGCTACATAAGGGAATTG  
 GCAGGTGACTTTGGAAAAGTCACTA  
 ACCTGCCTACTTCTTGGTTTTTACAGATTCTTTGCTTTTAAAAAGGAAGTTGGTAAGTTCTGCTTGACTAAGTTAGT  
 CCTTTTAAAAAGTTGTGTATTCA  
 TGTAGAAAGCTCTGAAAAGTAGGTGTATCCTTACCTTTGATTGGTTAGAGGTCTGTGTGAAAACCTATTATAACAC  
 TGAGTCCCCGAGAATATCTTAGATT  
 TTCTTTTCTTGATTAGTACATATAAACTCTTTGCCCTATAGAACTTTGATTTTGTCTTATTCTTTAATGGATGAGGT  
 GCTGTTTCACTTAAGAAGTAAT  
 TCAAGTGAGCAGAGTAATTTGTCAATCAGGTTGAATATCACATCATTTTAGGCTCTTATTAACCTGAGTACAAAAT  
 ATGTAAAGATTGCTTAGTGAGTTT  
 CAGACATTTCTTAAACAATGATTTAGCAGTTGTTTTTACCATTAATCGACAGAAGAGAATGAAGGACATGATATA  
 TTTTAAAAAATCAGTACTAGGGATT  
 TTAATGGTTTCATTTACTCCATGAGATACTTTAGAATGTGGCCCATGATTAACTGCCTTTTAAATATTGGCTATTT  
 GAATAAGCAACTGATACTCATAT  
 TTCATAACATCTAGAATCCTAAAAGAAAATGTTTCTTTTAACTATGTCAAGAACTGTGTGCTCAATTGGAAAGA  
 TTAAGGGAAATATATTCCTGATT  
 TACAACATCAATTCTATTATTTTATGCAAATCACTCAATCTTTCAGTTCCTTAGTTTTCTAGTATGACATGAAGGG  
 ATTGAATCTCTGAGACCCCTCAT  
 GTCATTCTTCTTCCAAACGGCAATAGAATTTCCCATGCTAAATCACTGAAGAGTTAGAGTATATAAGTGCCAGTAT  
 GTTTCATTTTGCATGAACTGGGCC

TGTTAAATTTCTGATGTGATAAAATCTTTATGCTGTAGAAGGTATTTGACTTATTTGGCTGAAGATTGAGTCAGAA  
 GTTTCCTCAAGTAGATCATGCCCT  
 TATTTAAAAGTGTTGCTGGCACTCAGGCAAGCAACATTGGTAACTTTACTTAGATTAAGAAAGTTTAAGGATTTTT  
 TGGTTGCTAAAACAAGGCTGCTTA  
 ATTTTTTTTTTAAGACAGAGTCGCACTCTGTTGCCAGGCTGGAGTGCAGTGGTTTGATCTCGGCTCACTGCAAC  
 CTCTGCCTCCCAGGTTCAAGCGAT  
 TCTCATGTCTCAGCCTCCTGAATAGCTGGGACTACAGGTGCATGCTATCACGCCTGGCTAATTTTTGTATTTTTAG  
 TAGAGACGGGGTTTCACTACATTG  
 GCCAGGCTAGTCTCAAACCTGAGCTCAAGTGGTCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCA  
 TGARCCACCACACCTGGCTAAGGCT  
 GCTTAATTTTTTGATGTGAAGATCCAAATAAGAGAAAAAAGTGAAGGAGATTGATTTTTTAAAAAATGGTTTTT  
 TATGATTTAAATACTTTTTTTTCAT  
 GTGCACTTTCACATGTTCTCTTTATGTACCTTGACTAATATTTACTCAGTTATGGGCATCATAAGAAAATAGTTTC  
 CCTTSTGTGTTTCTGTTACATTC  
 TCAATAAAAAATGAACCATATAGTAGTCATTTTGAAATGATAATATTCTTCTAGAAATTGTTAACAAATGGTCGTTA  
 CCAATTACTAGATACATTTTTACT  
 TTTTTGCCTGAGGTGTATTTTTTATTTTTATTTTTATTTTTTTTTTGAGACGGAGTCTTGCTCTGTCGCCCAGGCT  
 GGAGTGCAGTGGCGCAATCTCG  
 GCTCACTGTAAGCTCCACCTCCCGGGTTTACACCATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTCCAGGC  
 ACCCGCCACCACGCCCGGCTAATTT  
 TTTGTATTTTTAGTAGAGATGGGGTTTACCATTGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCACCT  
 GCCTCGGCTTCCCAAAGTGCTGGG  
 ATTACAGGCGTGAGCCACTGCACCCGGCCGAGGTGTATTTTATTGTTGCTGTTTTTGTTGTTGTTGTTGTTTTGTT  
 TTGTTTCTTTGACAGGGTCTGGCT  
 CTGTCTCCAGGATGGAGTGCAGTGGCATGATCTTGGCTCATTGCAGCCTCCATCTTTGGACTCAAGCGATCTT  
 CCCACCCAGCCTCCAGAATAGCTG  
 GGACCGTAGGTGCACGCCACTATGCCAGCTAGTGTTTGTATTTTTTTTCTTAGGTAGAGATGGGTTTTGCCATGT  
 TGCCCAGGCTGGTCTTGAACCCCT  
 GGGCTCAAGCAATCCTATTGTTGCTGTTTTATCATTTGTGACATTAAGTTAATATTTACTGCCTTTTGACTTTATTG  
 CATTATTTATGTATTAAGTTTAC  
 TTTTCTTTTTGTCTCTGTGACTCCAGACACTTAATTTAAAATAGTACTGTTTGGTACTAACTTGATCTCAGACTTAC  
 ATGAGGACTCTGAAATTATGTGA  
 CACTTCTTTATTCATGTATATAGTATAAGGTGGTTATCTGGATGACCCAGAATATGCTTATATTTTGGTTAGAAAA  
 ACCACCTAAATGAAAAGAACTTAA  
 TGCAGGCTTTAAGAAGTTTGCTGGACCCTGTTTTAGAAAATTAGTTTTATTTTATAGGATTAAAAAGTGGTGTTTT  
 TTTTGATTCATAAAAGTTTTAGAA  
 ATTTAAGAAGTATAGAACTGTAGAAAAAAGGAAAATTACTCAGTGATTCATAAAAGTTTTAGAAATTTAGGAAG  
 TATAGAACAGTAGAAAAAAAAGAAAA  
 TTAAGTGGTAACTATCCTCTTTATGATTCTTGCCTGTCTTTTGGATATATCTATATTCTTTCTATTGTATATAGA  
 GATTATATTATGAATAGCATT  
 ACTTTGTCTTGTTTTCTTAATGGCAATGGGAAAATGTATTTTGCTGGATTCCCTTAAGATAAAAGACACATTTTT  
 GACTCCTGTAATGGTCATTCTCA  
 TAGAAAATAATGAAGCCAATAATGTTTTAAATACTGAATCCCTTACCTCTTTCCTAGTTGTTGCTGCTGTGCTTC  
 AAGTAACTGAATTGATTTTATCT

CTAGGCTATCCAAGATGTATTCTCATACAAGACTGACTTAGAGTATTCTTTTTACTAGGTTGGCTAATAGCAGTTA  
 AATTATTTGTGCCAAATTGTGTCT  
 CACAGTTTGATTTTAAGATGTGTATGTTTTATTCTTTAATATTCTTTTGCTGATTGAAACCTACAGGAAAACAAT  
 TTTGATAATCTTGATTTTGATT  
 GGATTTGRTGCCTTGGGAGTCAGACATTTGGGACATCAACAACCAAATCTGTACAGGTAATTATGTGTTTCACTG  
 GTAAAAGTTTTAAAAGATCAATTT  
 TATTTGTAGTTTGGTCATACTCAAATTAGGTGGGGTTATTGGGCTATTCTGGTAGGTTAAAAGATTCCTGGTG  
 TTATGCATTCTACTTCTGGTATGGC  
 AATTATGTAAACAAATTTCTTGCCAAAAATGGRTATGAGGAGTTTCATAGTTATGTTTCAGTAGTGTAATGAATG  
 AACAGTTCATATCAGAAATAGAAAT  
 AGTGATTTTGAATTATGAATAATGTCTCTGTATAAGATTAACCTGAGAACCAGACAGGGCTTAGGGAATSAAAG  
 GGAAGAGCCTGGGTTACCAGAACCTA  
 RGAACCTGGTGGAGGAGTCCCATGCATTTAATACATGTACTTCTAAGAAGGGGCAGCACCTGGCTGATACTAAT  
 ATCTTTAAGGGGGCCTGATGAGGTCTG  
 GGGTTGAGTATACTGAAAGAAGCCAGAACCTGGAGCAAATTGCCAGTGCTGGGATACAGCTGATGGTAGAATG  
 AAGAAAAATCCCTTCTCCTCCCTTCTA  
 GCTTTTTCTAGTATCCCCTATTGACAGGACCTAAAAGGATGCCAGCTGGGAATAATTTATATCACCGATTATAGA  
 AAGTGGACTTAGTGCTGAGAGACAG  
 TGAATAAATAACCAGCACATTCTCCTATCTAAATTCCTCTCTACCTCTGCACTACCTATTCCTCTTCATTTTATGTAT  
 TTACTTATACTGTTTCATGCCT  
 GTTTCCTTCATAGAATGTGGGATACACAAAGTCAGTGTTCTGGTTTTATCGTTTGACAACTAAATCTTAGTTGC  
 TTAAAGCAACAACCATTTTATTAG  
 ATTGCTAGAATCTATGGGTCACGAATTTTGACTGGGTGCTTACGAATTTTGACCGGGTGCTCCAAGGATGGCTTG  
 TCTCTGCTCCATGATGTCCAGACCA  
 CAACTGGGGAGATTTGAAGATTGTAGTAGCTTGACATTTGGGGGCTAGAATCATCTTGAGGTTTCTTCACTCATA  
 TGTGTGGCACCTGGGGTGGAATGAC  
 TTGAAGACTAGGACTACTAAATGGAGTGCCTGCATAGGGCCTCTTTGTGTGGCGTGGTGTTCTCAGTCATGGTG  
 GTCTCAGGGTAGTCAGACTTCAAAGG  
 CTCCAAGTATGAGTGTTCCAGCACACAAAGCAGAGGCTGCATTTCCAGCTTCAGAATTCACATAGCATTACTTCT  
 ATCTTCTCTTGATTGAAGGAGTCAA  
 GCCCACCTAAATTCCTGGGGAGGGGAAATCGACTCTACTTCTTGGTGGGAAAGTGGCAAGATCACATTGTAGAA  
 GATCATGAGGGAAGATCATGAAGGAT  
 GGTAGATATCGTATGGCTAATTTTGGAACACAGTCTGCCATCCCCACCCTATTCTATCTTTGACACCAAGGGAT  
 ACTGCCTGGGCCATAGTTGCTCAG  
 TACATTTTTATAGGATAAATTAATAGTTTAAAAGATATGAACATTTTTTATGAGTTTGAATTGTATTGTCAAAAAT  
 GCTTTCCAAAATATTTCTGTTTC  
 TTATTTTTTTAAATTTTTTGGGAAGCTATAATTTTTGCATATGGTAAAATATTCAAAGGTACAAAAATCTGTATA  
 GTAGAAAAATGTCTCCTGCTTTT  
 ATTCCTCAGCCAACAAGTTCTATTTTCCAGAGACAACCACTGTTACCAGTTTCTTCTGTTTACCCTTCCAGAGAATT  
 AGTGCAGATGCAAGTGTACACAT  
 ATATATACACACACACAGACATACATATATACCCCCTCCCCACATATATATATAATTTGTAACACAAGTGGAAA  
 CATACTAGATATATTGTTTTCTGC  
 ATCTTTCATTTTCACTCGTTATATTTTGRTGACCTTAGCTTCCATATCTGATTGAATTTTAAAAGTAGATTTTACCTT  
 CTTTCTTGTTCATTTTATTATA

AGTTTTTATTGTCCAGTGGGTAAAACCTTTCCTCCATGTTTTCCAGTTAAAGATATTAAGGCAGAACCYCAGCCAC  
 TTTCTCCAGCCTCCTCAAGTTATT  
 CAGTCTCRTCTCCTCGGTCAGTGGACTCTTATTCTTCAACTCAGCATGTTCTGTGAGTAGCCAGTCTTTTACAATG  
 ATTTTGSTTATAATATGCTATCT  
 GGAAAATATGTATATGGAGGAGGCAGTTAACATTTATTGAATACTACTATGAGTTATATTGTCTCATTTAWTGAG  
 GGAAGTGAAGTCAAGAAAGAGCCT  
 TGTGGGAGCTAGAAAGTTGTTAGTGCCTTTGGTTTTGCTCTCTTTTCCCCTCTGTTTCATGCCTCCATAATTTTCATTG  
 TTTTCATTGTTTCATTGCATTTT  
 AACAGTATCCAGTATTTGTTTATTGGCTTACTGTCTTTCTCCACCCCTTTCTTTCTCCTCCCTATTAGGATATACAT  
 TCTTTATATATGGACATAGAAG  
 AATGTAAGATGGTCAGTTTTTATATCTCTTTAGTCACAGCATCTAGAATGATTGCCTTAGTTTGGGAATGCCTAGT  
 GCTACCTTTTGAAGGGGCTCACGT  
 TTTTATAGAAGTCTGGTAGCTCTTTTAAGAACAATTCTGCTCTGTGTCCAGACAAAAGCCCTGAGGCCAGGAAAA  
 GGAAAGTACCTGATTTAGGTACTTTT  
 TAGAATGGGGAAAGGGAAAGTATAAGAATAAAAAAATCAAATAATATTTTAGTATATTATATGTGCTATAAATT  
 GTGGTATCTTTGTTCTGCAGGCTCT  
 ACATTTAGTTACTAAATCCTGGTGGTTATATGCACATATTAATTCAAGGATTTATCTTTTCTGCCCCTGCCAGGCCA  
 TCACTATCTCACCTGGATTAAGT  
 CAATAGGATTTTAACTGGTCAGCCTTACAACCTCTCAAACCTTTCTTAGCAGTGGAATCTTCAAAGAGAAATCTTT  
 CAGTGTCAACAAGCTTTACTGTATC  
 AGTCATTTGCTTAAACTTTTTAGTAACTTCCCAGTCCTTACAAGGATGAAATTTAAATTTCCAGAGTACTAGGCT  
 TTCCATGATCTGGCCCCCATTTAC  
 YTTTCCAGTCTCTTTTCTTGCTATCCCCTCCTACTTCCACATTCCCCTTAATTTGAGCCCTGTGAGAGTTCCTCTCTG  
 AGAGCTCCTATGCACTCTACTC  
 ATTCAACCGCTGTGCCAGGCACTGTTCCAGATGCTGGCAGTGAACAACAGCAGTGAGATTGCTGCTGTTACAGTT  
 CTAGGTTTTTGCACATCCTCTTTCT  
 TCTGCCTAGAATAGCCCCATTTTTTTCTTGCTTGGAAAACCTCATCCTTCAAGTCCCAGTTCAAAAACCTACTTCCTT  
 TATGAAGCTTTTCTCTGATTGCT  
 GCTATTGGTAGCTCTCTGAGGTCCCACAATTTCTCAAATGATTTAATTTTTTAAAAAATTAATAATTACAAAAGG  
 TAGCATACTGTATACCCTGTCCTG  
 TACCTTGTTTTTTTAGCTTAGCAGGATAACTCTTAATTTTCAAGAGTATATAATTTTCCCATTATGCTACATAATGTTTC  
 ATTGCATGGCCTTATCATGCTT  
 AATTTAGTGATTTCTTATTGATGAGTATTTGTGTTGTTTTTAGTCTTTTCCAGACCCAATTAACCTGTACATAAGT  
 CTTTTCCAGACTCAAATAACCTA  
 TAGTAAATAACCTGGTACATAAGTCACTTTATATTTTACCAGGATATTTTGGGATATATTTCAAGAGGTGGGAT  
 TGTTGACACCAAGGTAAATACACT  
 TGAATTTTGCTAGATACTGCCAAATTTCTCTCTATAGGGGTTGTAAAATTTTACATCCTCATCAGCAGTATGTAAG  
 AGGGCCTGTTTTCCCATCATCACT  
 CCACAATAGGTTTTGTTTTATATCTCTTTTATTATAAGTATGGTTAAATAATATCTATTCATATGTTTTATATGTTTA  
 AGGCCTATTTGCATTTCTTTT  
 AAAAAATGAAGTGTCAATTTCAAATTTTTGTTCTTTTTTTTTTTTGGAGATGAAGTCTCTCTCTTGTCTACCAGGCTG  
 GGGTGCAATGGCATGATCTCGG  
 CTCAGTCAACCTCCGCCTCCTGGCTTCAAGCAATTCTCCTGCCTCCTCAGCCTCCCGAGTACCTGGGATTACAGG  
 CACCTGCCACCACACCCGGCTAAT

TTTTGTATTTTATAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTTCTGACCTCAGGTTATCCA  
 CCTGCCTCGGCCCCCAAAGTGCT  
 TGGGATTACAGGTGTGAGACACTGCGCCCGGCCGTTCATTTTTAATTGAATTGTATTGGTCTTTTTCTCTTTAAT  
 ATTTTAGGTGCTTTATTATATATT  
 AGGATTATTAGTCCTTAGTTTGTGAAATAAGCTGCAAATATTTTTTCCCAATTTTTCTTTGTCTCTTTGCTTTATT  
 TCTGGTGTTTATTTTTCCCCAG  
 GTGTACTGTTTTCTATAGCCAAATTATTGATCTTTTCTTTAATTGCTTTTTTGTTCACCTCAAAGGAAGGTTTTACCA  
 CTGAGAGGTTATAAAATAATTC  
 ACTCATATTTTTACTCTTTTTTTTTGCTTATGTTAATTTTTGTTTGAACAATCTCAAATTCACAGAAGTGTTGTGA  
 GTACGTTGCAAATAACTTGTTT  
 TTTCTTATGCATTTAGAGTAAGTTACCAGTCCGATGCCCTATCACTCCAATACTTTAGTGTCTACTTCCTACAA  
 GCACAGATATTTCTTAGCCATAA  
 TACAAGCACAAAATCGGGAAATAAACATTAATATGTTACTATAATCTAACCTTCAGCTCTCGTTTAAATTGTGCCA  
 GACATCCCAGAGGTGTCCTTTAGA  
 CTATCAAAAGGACCCAGTTCAGAATCAAGCATTGCATGTGGTATCTTGTCTTTTAATCTTGTTTCAGTATGAAAC  
 AGTTCTTCATTAATCCTTGACTTT  
 CTTACCTTGTCACTTTTGGGGATTAGGAGTTAGTTATTTGAAGAGCATTTCCTAGTCGGGTTTGTCTGATGTTT  
 CTTTAGGATCATATTCAGGTTTTA  
 TATCTGTGGCAGGAATGTCACAGAAGTGCTGTCTGTGTCCAACCTTTGATGACCCAATTTTGATTTTTCTAATAC  
 TGATTATTTTCACTTTGATCCCTT  
 CATAGATGTTATCTGCCAACCTATTCCATTGTGAAGTTGCTTTTTCTCTCTATGCCATACATATTTGTGAGAAAGA  
 TACTTTAAGACTATGTWAAAATA  
 TCCTGTTTCATCATCACACTTTATATTTGTTCACTTATTTGTTAATATCTGTATGGATTCATCGTTTCTTATTTTATTCA  
 ACAGATTATAATCCGTTACTA  
 TAATTATTTATTCTGATTCTTACATTGTTCTGGATTGGCCAGTGGGGGTCCCTTCACACCAGCTCTTATGTTCTTT  
 TAACATGTCTCCTTTGGGCTCTG  
 ATGTCCTTACACTAGGCTGTCTTCCCATGCGGATGCCCTCCTCACTCTGCTGTGACTCCTCATGCCAGGCCGCCTG  
 CCATATCAGTGCCCTCTCTACCTT  
 TCTTGGACCCTGACATTCCACCAGGCTACTTCTCTGCGTGGATGGTGGTCTTGCTTTACCCCATCTAGTGGCTTAG  
 GATAAAATTATTCAGGAAGGAATA  
 TTTTCTAAGTGACTATGAATAAGTATTTATCATGAGATATTATTTTAGTGTTAGATCTCGGATCCATTTGAAATT  
 ACCCTTGTGTATTATATGAGATA  
 CAAATCCAATTTTATGTTTTTCCAAATGACTATCAAATGATTATTCTAGTACCATTTATCAAAAAGTTCATCTGAT  
 TTCAATTATATGCATGTATTTA  
 AATTTATGTAAATGGCATTATAGTAGATATCTTTATATTATTTACTTTTAAAAATCAGCACAGTGTTTTTAAGATAA  
 ATTCCTATTGCTCTGTGTACATC  
 AAATCTGTTGCATCTAACTCTGTACAACATTCAAGTATGCATTCACCTACATTTTACCTACACCATAATGAATATTT  
 TAAAAAATGTTCCCTTGCGGACC  
 CATGTGAGAAATCCTTTAGAATATATGCCTAGAATTGCTGGGTAGTAGAATGTGCAGGTATTTAATTTGACAAGT  
 GCTTCCATATTTCTCTCCAGGATCA  
 TCATAGTACTCTGTGGGCTGGCAGCAGTGCCTGAGGATCCCTGTGACTTCTCTCCTGATACTTGGCATTGTCTGCC  
 TTTTAATTTTACTCCATGGCAC  
 AATGTGTTTTTTGTTGTTTGCTTTATTTGTTTTACTTTGTATTCTTTGGTTATTGTTTTCAACATTTATTACATT  
 CTTATGAGCTTTTGGACTT

TTTTCTTTTATACATTGTTTATTCCTGTCCTTTGCCAGTTTTCTACTCAGTTGTTGTCTTCTTGTTTATCAGCAGGA  
 ATTCCTCATATACTTGAAATA  
 GTAAACTCTTGTTGGATTAGACATTCAAATTGCCTATATCTTTCCCTTCTATTATCTATTAACCTTGTAGTCTT  
 TATTAAGTATAAATAGTTAATT  
 TTCAATTACTCAGGCTTATCTTTTTTTCTTTTCCTTATGGCTTATTTAAGAAGTCCTTCTATGCTTTTAGAGCAC  
 AGAGGTATTCTGCATTTTCTTT  
 TATTAACCATATGGTTTACCTTTCACATTTAGGTTTCTTAATACTCTTAGACACCACTTTAATATGTGGTATTATGT  
 AGGGATCTAGCACTAGGTAGGGA  
 TCTAATTTTAATTTTCTCTATAAAATGAATCAGTTTTCTAACTGTATTGTCTTTATAATCACTTGTAATCACTCTTTT  
 CAACAATATTGTATATCTCTC  
 TGTTTACTGAAATCTTTTATGCCTTAGTTCAGTTGTGTAGTTTTCTTCATATGGGTCCTGCACATTCCTTAGAAAAT  
 TTCTAGGTATTTTGTAATTGTTA  
 TTGCTATTGTGAGTGAAATATTTATTTATTTGTTTTATTTAAATCTTTTTTATTTTAACTTTTTTTGGGTACATGGT  
 AGGCATATATATTTGTGGGGTA  
 CATGAGATGTTTTGACACAGGCATGCAATGTAAAATAATTAGATCATGGAAAATGGGGTATCCATCCCCTCAAGC  
 ATTTATCCTTTGTGTTATAAACAAT  
 CCAGTTACATACTTTTTGTTATTTTAAAATGTACAATAAGCTCTTATTGAGTATAGTCACCCTGCTGTGCTATTAA  
 ATAATATGTCTTATTCCTTCTCC  
 TTTTTTTTTTTTTTTTTTTTTTTTTTACCCATTAACAATCCCCATCTCACTACCAGCTTCTCAGTACCCTCCCCAGCCTC  
 TGGTAACCATCCTTCTACCC  
 TATGTTTCATGATTTCAATTGTTTTGGTTTTTATGCCACAAATAAGTGAGAACATGTGATGTTTGTCTTTCTGTGC  
 CTGGCTTGAAATATTTATTTTTT  
 ATTATGTTTTCTAAGTGTATATTAATAACAATATTTTCAAGTCAAATTTTTTTTTTCAGTGGATTGTTACATGTGCTTCA  
 AGTATATAGCATATTACCTTGC  
 TAAAAGGCACGCTGTCTCTTTTGTTTTAATGTTTTTAAATTTCTTCTTTTAGTGTAATTATACTAGAATTACAGAA  
 AAAGCAGGTATCTATTTTTTATC  
 CTGACTTTAATTGTAATGCAYTCAGCATTGGACTTGGCTGTATAATTACTGATAGCTTTTTAATGTTAAGAAAGTA  
 TTCCTCGGATCCTTGTTTTCTAGG  
 ATTTTATTGAGGAATACTTGTCGAATATTATCAAATACCTTTTAAGCATCAAAGGCCGTGAAGATAATAGTCTCCT  
 CCCGTGTTACCTTGTAATCTTTGG  
 TTTCTTGGAATGTATTTTAATACTCTACATATTGCTCAGCTTACATGCTGAGTAGGTGTTAAGGATGAAGGA  
 AGATGCAATTGTCACTTGTCCTCT  
 CACCAGTGAGACTGAGAGCATTTCTTTCTCCCATATCCTCTCAAGGTCATTTAAATTTTTTTTTATGATAAAAAAAG  
 TTTAAATTTGCCTTTGACTACCA  
 TTTGCGTTTCCTTTTCAGTGAAATGGCTGTTCTGCTTTTGTCTATTTTTCGCAAATATTTTCTACCAATATATATC  
 TTTTAACTTTGTTTATGGTCTA  
 GTTTTTATTGTTAAATAGTTTACTTATTATTTGTTCAAATCTATTTCTTCTTCTTTGTGGCTTCTGTGTGTTCTGACT  
 TAAGAGGGTCTGCAACACCT  
 GAGAGTAAACAAGTATTCTCTTAAATTTTATTTTATATTATAATCTTATAGATTATTTTGAAGATGTTATGAAGTA  
 GAGATTATGTCTATTGCGTGATC  
 ATGCATGCCATCTTTTCTGTCTGAATTGAATTGCTACTTCTATGTTGTTTCTTTATTCTAACAAAATGAGATAGATT  
 CACTGAATATTAAGACTTTGG  
 GAAAATGCGAATGTTATCTTTATGTATCTCAATCATTATGAATTCACCTACTAAATTAAAGAAAATAGCACTGTTG  
 GCTGGGTGGGGTGGTTCACGCCTG

[illegible]

ATGGTTTAGTAAACAGTATATTGACCTGGTCATGAGGAGACCTGAACTCTGCTTTCAACTAGCTAAGAATTTGTG  
TGTTTCRTATTTTATCCACCAGATG  
GCCTCATACACTTGTGATTCTATGAGAGGTTTATGTCATTATTGAAAGAAGTTCAAGTCTGAGGATTTGGTACTAT  
TCAGATATGGGAGTCATTTCCGTA  
TGGAGATCTCTTAGGTAGATTAGTCATCATTGTACTCATACTGATATCTGTATACACTGGGTACAAGAAAACTTT  
GCAGCATTCTTATCTCTGGATTTA  
GTTTGTTGGCACCAGATATGACTAGGCGATAATGGGGTATGGGTTGTAAATAAACACCAATGTTACATTTGTACT  
GTATCTCCAATTGTGAATAGGTAGC  
TCGTTTTTTGGAGCATAGTTTTAATGAGGGTAAATTTGATAAATTTAATTTCCAGATACTCTTTTAGCTTTGTTCTA  
TTACATGGTCATAGACAATATGA  
CATTGATAAAAAGAAGACCTAGATGAGAGAATGGATGAATCAGTGAAAATTCATCCATCAGTCCCACCAAAAAT  
AAACAGACAAACAAACAAACAAAAAC  
CAGTTTGAAATGCCTGCTCTCTCAGTACAAAGTCAGAACTGCTGCCTTTTCCATCTGAAAAGTAGCATTCTATAG  
AATAGGTATTCTTGTGATTTATAA  
ATGTGTATTAAGGGTTCATGAACTCCTTGAAACAAAATGCAGGATTTAATGTATGTTTCATATGTGGATTATTCTAG  
GGATAGTCTCCAAAACCTTTCATTA  
GGATCTCATAAGGATCTGTGCCTTCCCAAAGGCTAAGAATTTAACTAATGAAGAACCTCATTGTAGTTCCCAT  
AGCTATTAATAATACGTCTATACAAA  
AAGTCTGACATTAATATGACTAGCGTATAAAGCAGGGGTATTCCTAAGAATTTGCCCTTAGCAGATCATTGACTA  
TGTTTGAGTGATGGAAGATCTTACC  
TATAATAAGAAGACAACTGTTCTTTTCTGAGCAGTTTATCAGTTCCAGAAACCTGAATTACACATTATTTGAAAC  
TTCAATACTTTCTGTCTATCTGAA  
GACTATAGTTATTTTTTTGTGAGCATGTGCTAATATGAGGAACTGCACCTTGCCTGTTTTTTTCTTTTATTTGAGA  
CAGAGTCTCRCTGTGTTGCCAGG  
CTGGAGTGCAGTGGTGTGATCTTGGCTCACTGCAACTTCCACCTCCTGGGTTCAGTGATTCTCCTGCTTCAGCCT  
CCAGAGTAGCTGGGACTACAGGCA  
CGTGCCACCATGCCAGCTAATTTTTATATTTTAGTAGAGACAGGGTTTCACCATGTCGGCCAGGATGGTCTTG  
ATCTCTTGACCTCGTRATCGCCTGC  
CTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGCGCCTGGTGCCTTGTCTGTTTTTATAGTACAAT  
CTAAATAGGTGGTATTATTATAATA  
ACATACAGATTTAAAGTATGTTTTTCATTTATTTGAGAATGGGGAGGAAGAAGGGAAGTATGGGCAAGACATTA  
TACAGTTAATCATGTTAATACTAGTA  
GTAAATATGTTACCATTGAAAAAGAAGAGGTAATTTCCACAGGTTTAAGACTGAGATTCTCTAGCTTGGAATGTA  
AAACTCTTCACAATTTGGCTCTAGT  
TTAGTTTTTTGTTTTTTTGTATTGTTGTTTGTGTTTGAATAGAGACAGGATCTCACTATATTGCTCAGTCTGGT  
CTCGAATTCTTGGGCTCAAGGG  
ATCTACCCACCTTGATCTCCCAAAGCTCTGGGATTAAGGCGTGAGCCACCATGCCTGGCCTCCTAGTTTAGTTTTC  
GAACTTTATAATATTACTTTTT  
TATAAATCCCTTCTTCTAGCTCACTAATCTGTCATTGTCCTTTGAATTCTCTTGTGTTTTCTAGTTTTTTGGCTCACC  
ATTTTCTACCAACCAGCAACA  
ACTTATGCCTACTTCTACTTCTCAGTCCTCCATGTCCTTTAAGACTGTTTTGAGTAGTGTAGATGAGAGATAATGA  
AAGCCTGGAATATCAATGAAAGTG  
GAGAACCAGGATGAAGAGTAGAGATGTTAAAAGTGTAAGTCTTAGAGATTGGTAATTGGATTGTGTAGAAGG  
AGGCTGAACGATAGAGAGGAAGGTGTC

TAGGATGACTTCCGTTTTCTATCTTGGGGCTTTTTGGGTAGATGATGGCRAAAATTATTGAAATGGGAGACAGA  
 GGTAAAATATGTTGCCCAGATCACT  
 CTGCTAATAAGAAGTGGAGCCAGAACTTGAGCCCAGGTGGTCTGGCTCGGAGTCTATGTGAATAATCACTATAC  
 TGTATCGCCTTAAAAAAAAAAATGACA  
 GATGTATTCCTTGCTCTCAAGAATTCTACACTTTAATTGGGGGATTAAATTCACATAAACAGTTAAATAACAGTCA  
 CTGTGATAACTAATAAATTAATAA  
 CTCATGCAGTGCCCTTATGGAAGCTATTATTATATAGAAAAGGAGATTAATAATAATAATAATTGCATTTAAGTTC  
 TTAATGTGCGCAGACACTCTTCT  
 AAGTGCTTAAGTGCATTATCTTATTTCAAATAATTTACACATTATTTGAAACTTCAGTATGTTTCTATGATTTAT  
 GGTGATGGGTTGTGATATCTCCA  
 TTTTTAGATGAGAACGCTAAGTACAGATTATTTGCCCTAAGTCACACACCGATGCTGTTTCAGACAAAGTAGGC  
 ATTCAATAAATGTTTGATACATAGT  
 TAAAAATCTTTGAAGAAAGTTTAAAAGAAAGAACTATCAAAATCAGTGAGTTACACATAGTCAAAAAAATATATG  
 GCATTGAATCCAGCCATGTTCCGTT  
 AAGGGGGATTGTCATGTAATGCCAGAGAGGATTTTGAGCATGCTAAAATGGATAAACAGCCTTTGGAATGTCCT  
 TTTAACTGTTTCCACACTAGGAACA  
 GTAGTAATGTTTCTGCTGCTATTTAAGGCCCAACAATTCTTTGCATTGTAGTTTGTGGCCTGGTCAGCATTT  
 ACCAAGTCACCATGAGAGTTTTTT  
 GTTCTTTCACTTGAACATATTGTCATAGAAATTTAGTTGTTTCCCCTCATTCTGTGTTGAACCCAAACATGCCTAGA  
 ATATTTAGAGTATCCCTGATTTA  
 GCTGCTGGGTAATTTGAGGATAGTTGGGTACTCTGGGATGAAAAGGCTTGATACAATATGGGAAAAAAGGAA  
 GCAGCAGTGGCTTGGTAGRACTCAAGC  
 CACATTTGGCCTTGCCTTGGTATTCTTCCATCTCCTTTGTGTGATGCGGTCGTCTCAGCAAGGAGACAGTGGTAAT  
 TACACAATTGGTAGTACAGATTTG  
 GATTATTTTTCTTAATTAGTTTAAGAAATATTACAAAGCCCTCTGTACATGTTATCTGCTAGGGCTTTAAATCTGTA  
 ACTGTAGCTAGAAATAGGATTCT  
 TTCTTATAAGTTGCCTAATTCATTATTTACACCTCTTCTGTTCTTTAAAATAGTCCCAAAGTTCCTGAACAGTGT  
 TTAATTTTGGGGAGGGATCTTT  
 TAATCATAATTTTAACTCAGAGGATGTAGAACATGGCTTCCTATAATACTCTTCTCAAAGTATTGGTTTAAGTCA  
 GTGGTTCTCAGTCTGTATATTGTG  
 GACCCTGGGATGTGGGTGATGGGGGTAGTAGCTTTATGGCATTTTGCAAAGCATCTGCAAATCCATACTTGAAG  
 CAAGCATTTTCTGTAGACCAAATAAT  
 TTATTTCATATATCTTTGTACAGCTCTTCAAATTTTATAGCTATACTTCTGACTAATAAACTCCAAGTTCATTTTAA  
 ATCTTCACTAAATTCTTAGCAG  
 CCTTTTCCATCATGTATTTTCTCAGTACACAGAAGCTGAAAGGACAACCTTGATGGTTTTGGGTGACCAGGATAT  
 TATTTTATATTAAAAGAGAGCCCT  
 TATATTCAGGAGCCTTTGGGGAGAGACTATGGGTTTTCTAGGTGTAGAATTATATCATCTGTGTGAGAGATGGTT  
 TGAATTCCTCTCTTCATATTTGGAT  
 GCCTTTTATTTCTTTCTTTGCCTGATTGCTCTGGCTATGACTTGAGCACTTGGGAATTCCTGTTTTTTGGTTTTTTT  
 TTTACATTATGATAAATATAT  
 ATTTTCATAGTGCCAGATTTACTGTAATTTCTTATTGTTATTAAACCTAACACTTGGTCATATGCTATTTACTTTAT  
 GTGTTTCTAATTTTTATTTGCG  
 TATTTGGCCTAAGCATRTTTATATCTGCTTCTGAGTAATTTCTGCATTTTATATGTTTTTACTTCATAAAAGATGGA  
 GGCCACCTCAAAACAAATTAGGC

ATCAAAGGAACATACCTCAAAATAATAAGAACCATCTATAGCAAGTCCACAGCCAACATTGTAATGAGTGGGCA  
 AGAGCTGGAAGTACATTCCCTTTGAG  
 AACCAGAACAAGACAAGGATGTCCCTCTCACCACTCATATTCAACATAGTGCTGGAAGTTGTAGCCATAGCAGTC  
 AGGCAAGAGAAAGAAATAAAAGGCA  
 TCCAAATAGGAAGAGAGGAAGTCAAACCTACCTCTCTTCACAGACGATATGATTCTATACCTGGAAAACCCATAGT  
 CTCAGCCCAAGACTCCTAGATCTGA  
 TAAACAACCTTCAGCAAAGTTTTAGGATACAAAATCAATGTACAAAAGTCAGCAGCATTCTATACCCAGTAATG  
 TCCAAGCTGAGAGCCAAATCAAGAA  
 TGCAGTCTCATTACGATAGCCAAAAAAGAATATAAAGTACCTGGGAATACAGCAAGCTGGGGAGGTGAAAG  
 ATCTCTACAATGAGTTACAAAACACTG  
 CTGAAAGAAATCAGAGATCACACAAACAAATGGAAAAACATTCCATGCTCATGGATAGGAAGAATCGATAGTGT  
 TAAATGGCCATACTGCTCAAAGCAA  
 TTTACAGATTTAATGCTATTCCTATCAAACCTATCAATGACATTTTTTCACACAATTAGGAAAACTATTCTAAAGTTC  
 ACATGGAAGTGAAGAAAGAGCCTA  
 AATAGCCAAAGCASCCTAAGCAAAAAGAACAAGCCAGAGGCATCAAGCTACCTGATTTCAAACGTACTACA  
 AGACTACAGTAACCAAAACAGCATGA  
 TACTGGTATAAAAACAGGCACATAGACCGATGGATTAGATTAGAGAACCCAGAAATAAAGCTGCACCCCTGCAA  
 CCATCTGGTCTTTAACAAAAACAAC  
 AGTGGGGAAAGGACTGCCTATTCAATAAATGGTGGTAAGATAACTGGCTAGCCATATGCAGAAGATTGAAACT  
 GGATCCTTTCTTTTCAGTGTATACAAA  
 ACCCAACTCAAGATGGATGAAAGACTTAAATGTACAGGCTCAAATTATAAAAACCCTAGAAGAAAACCTAGGAA  
 ATATGGTTCTGGACATAGGCCTTGTC  
 AAAGATTTTCATGATGAAGACTCCAAAAGCAATCGCAACAAAAACAAAAATTGACAAATGGGATCTAATTAACT  
 AAAGAGGCTGGGCGTGGTGGTTCACA  
 CCTGTAATCCCAGCACTTTGGGAGGCTGAGGGGTGCATCACTTGAGATCAGGAATTCAAGACCAACCTGGCCAA  
 TATGGTGAAACCCCATCTCTACTAAA  
 AATACATAAATTAGCCAGGCATGGTGGTACATGCTTGTAATCCCAGTTACTTGGGAGGCTGAGGCAGGAGAATT  
 GCTTGAACCCAGGAGGCGGCAGGTGC  
 AGTGAGCTGAGATTGTGCCACTGCACTCCAGCCTGGGTGACAGAGCTCAAAAATCAATAAATAAAAATCAACTA  
 AAGAGCTTCTGCACAGCAAAAGAAAC  
 TATCAACAGAGGAATAGGACAGCCTACAGAATGGGAGAGAATATTTGCAAACAGTGTATCCAACAAAGGTCTAA  
 TCTCCAGAATCTACAAGAACTTAAC  
 AAGCAAAAAACAACCCCATTAAGAAGTGGGCAAAGGATATGAACAGACACTTTCAAAGAAGACATAGACATG  
 GCCAATAAGCATATGAAAAAATGCTCA  
 ACATCACTAATCATTAGAGAAATGTAAATCAAAACAACAGTGAGATACCATCTCACATCAGTCAGAATGGCTACT  
 ATCAAAAAGTTGAAAAATATCCGAT  
 GCTGGCAAGATTGCAGAGAAAAGTGAATGCATATACATTGCTGGTGGGAATGTAAATTGGTTCAGCCACTGAAC  
 GTAGTTTGTAGATTTCTCAAATAACT  
 TAAAGTAGAACTACCATTTGACCCAGCAGTCCCATTACTGGGTATATACCCAAGGGAATATAAAGACACACGCAC  
 ACATATGTTTCATCACAGARCTATTC  
 ACAATAGCAAAGACATGGAATCAACCTAGATGCCCATCAGTGGTGAAGTGGATAAAGAAAATGTGGTACATATA  
 CACCATTCAATACTACACAGCCATAA  
 AAAAGAATGAAATCATGTTCTTTGCAGCAACATGGATGGAGCTGAAGGCCATTATCCTAAGATAATTAATGCAG  
 GAACAGAAAACCAATACTKCATGTT

CTGTCTTATAAGTGGGAGTTATACATTGAGTACACATAGACATAAAAAGGGGACAATAGACAACAGGGGCTACT  
 TTTGAGGGTGGAGGATGCAAGGAGGG  
 TGAGAATCTAAAAATTACCTGTTGGGTACTGTGCTCATTACCTGAGTGATGAAATAATCTGTATGCCAAACACCT  
 GTGACATGCAATTTATCCATGTAAC  
 AAACCTGCCTGTSTACCCCCTGAACCTAAAATAAAAGTTGGGGGAAAGAAAACTGAATCCAAGGGAAAACGAG  
 AAGGTAAAAATGTTTAATGATCATTT  
 ATGCCGAAGTATTAATACATTGTATCTACTCCAAATCCATGAAATTATTTTTAGATCTTCTTATCTGGGATAAAATG  
 CTGTTTATTGAAATTCTGCTACT  
 TTCAGTGTGTTTTTTGAGGCCATCTGAAACAAGGGCTTTGCTTTTATAAGTTATCTTTTTTAGCCAAGACAGATGT  
 TCTTGACAAACAGGTGTTCTGAAG  
 CTTTATTTTGATTTTTAGTTTTCCCGATCCTTTGAAATATTGAGCTGAGTTTTACATGGGTCAACCTGTCTGAAAG  
 CATAATGCTTTTTTTTTCTTAAA  
 CAAATTATAAATTTACATGATATTTTAGTTGCCAGTTATTTTAGTACCTTCATTTTGTTAGATAGGCTCAAGCAGCT  
 GTGGATGTTAGGCAAGGCTTTCA  
 GCACCTTATAAAAAATCTGCCAGATCTGAAAGGAAACACTTAGCAGGCAAGTAAAATTTTAAATAGATTAGTGTTA  
 TGAGTTGTATAAGTCTTTAGGTAA  
 TGGTACAGGTTTAGTACCCCTTAAATGAAATGCTTGGAACCAGAAGTGTTTCAGATTTTGGGTACTTTTGGATTTT  
 AGAATATTTGCATATACATAATGA  
 GATATCTTGGGGACAGAACCCAAGTCTAACATGAAATTCATTTATGTTTCATGTATATCTCATACACATAGCCTGA  
 AGGTAATTTTATATAATATTTTAA  
 ATAATTCTCTGCATGAAACCAAGTTTGCATACAGTGAACCATCAGAAAGCACAGTTGTCACTATCTCCTGTGGAC  
 AATCTGTAGTTGTTTGGCACCACCA  
 TCATTCCTGGCTCTGAATTCATATGCTACTGACAAGCAATCATTTTCTTACACTTATTCACAAATTCACACCTAAGA  
 ACTTAACAAAAATTATGACATAC  
 CATTAATACAGTGAAAAAATAATGTGTGCAGAGTAACAAGGCAGGACAGTAGAATCACCAGAATACCAGTGTC  
 GCTACTACACAACAGCAACAACAAAA  
 AATGGCAGCTTTCAGTCTCTGCCTTTGATGCTGTTTTATTAAAGCCTACTGTACCCTCTATTTTATGTCTTCAGG  
 TGAGAAGAAACATCAGAAGCAGT  
 TGAGAGACTGGAAGTGGGTCCTCTAGGGATGAGGAGGCATTCTGCTGGATGACTTCTAAAAATGTTTTCTATGG  
 AGTCATCTGCCTCATTAACAAGTTTT  
 TGTYTTAGAAGTCTCTTTTTGATTTTGTAAGTACTGACATGATTTCTTTTTCTGTTATGACTGTATGCTGCTTTAGTCCT  
 TCACTAGGCCCATCACACATTT  
 TTACCATGTCATCTATAGGCACCTTTTTCTGCAGCGTTAACAGCATCATCTTCATTGTCATTATTATAGATTGAGTAT  
 CCCTTATCCAAAATACTTAGGAC  
 CAGAAGTGTGGAGTTTTCCAATTGTGAAGTTGTGTCAGCACTCAAAAAGTTTTGGATTTTGGAGCATTTTGGATT  
 TTGGGTTTTTGGATTAAGGATGTTC  
 AACCTCTAATAATTCTCGTCAGATTAGCATAAATTTAAAATAATGATACCTCTGAATACTCAGAAAAGAATTGAA  
 GGCCATTTTTGGGGGAAGAGGGAAA  
 GGGCTGGTATCATCTACAGACTATTTTGAATCACAGATTTATGTTTACCGTGATTGTTAGCTAAGATTTAATCTCA  
 AATGGTTTTAAGTTTTTTGGTATT  
 TGTAGTCTTTTTTTTTTTTTTTTTTTTTYGCCTTTTAGGGATCTGCCTGCCTCAGCCTCCTGAGTAGCTAAGACT  
 CGAGGCACATACCACCATACCT  
 GGCTAATTTTTAAATTTTTTTTTGTAGAGACAGGCTATCACTRTGTTGCCTTGGCTGGTCTCAAAGTCTGGCCTCC  
 CAAAGCACTGGGATTACAGGTGT

AAGCCACAGTGCATGGCCTGTAGCTCCTTCTTCTTAATATTGAGTTACTAAGTTTATGTTTCAGATATTTTTGGTA  
 TTGTATGCATTTGTAAATCCTAT  
 TAATTTAGAATACATGACTCTCCATATTCAGTGTAAGAGATTTGTTATTTATAATAAGACAAATCCTTTTAGCRCTA  
 ATAAATTTTCAGAGGTTAATTCTT  
 AGAAATCTAGGGATTACAGCAAATTTGATAGCTGTTTAGTTGAGAAAGATAAATAAGAAAGTATATTAAAGTAT  
 ATTGTTAATGATAACTCAAAGAGATG  
 AATTTAGTTGGTCACATTAAAAACATAACAAAGTCACCATAAGCGTGTGCTGCTATTGATCCAATGTTTTAATTGA  
 TGGTGTCCAATCCCTTGGTGTTGA  
 TCTTACCCATTACGTCTGCAGAAAAGACCTAATTATAAGCTGCTTCCCCTGCCACAGTAGTTGTGTACCCTTTGAT  
 TCCTTTTCTTTTTTCTAACGCAGG  
 CCAGACGGTTTTGCTGTCTCAGCCTACTGTGGTACAACCTCAAGCACCTGGAGTTCTGCCCTCTGCTCAGCCAGTC  
 CTTGCTGTTGCTGGGGGAGTCACA  
 CAGCTCCCTAATCACGTGGTGAATGTGGTACCAGCCCCTTCAGCGAATAGCCCAGTGAATGGAAAACCTTCCGTG  
 ACTAAACCTGTCCTACAAAGTACCA  
 TGAGAAATGTCGGTTCAGATGTAAGTTTTGAACTTAGTGCTTCTCTTAATGCCTGATTTAAAAACCAACAAAAA  
 AACACCTTTTGCTTTTGTTGCCTTG  
 TTTTGTTTTTTTTAGAGTTTCTTATATTGCATCAAATAAAAGAGTGTGCTGAATTGCACATCTAGTCTTTGGAGGA  
 AAATTAGGGTAAGGCTTACTACG  
 TTAGAATCTAAGTGAAAAGGTTTTTGAAAGTGAAGTGGTTTGATGAAAAGCAAAGTATGCATACCTCCCTCTTCT  
 TTAATCCTTCTTTAGAACCAACCCT  
 CTTATTATTCTCTTCAAGCCCTTACACCCTCAACCCAGTCATCCACTAATATTTTACAGATTCAACGTTTCCTGATTG  
 AAATGTTTTAGCGATCTTCAGA  
 TAAGAGTCTATAGATTACTTTTTAGGTAGGTTTTATTTAGCTTTCTTCCTTGCCGTTTTGAAGTATTTTCACCACT  
 TCATAATGATGACTAGCTACTC  
 TTAGGAATTCTGTCTCTTAGTGATTATTTGGTACTTTCCCATTTTTGCAAACCTTTAAATTTTCTAGCATGTTGTAG  
 GGATTTCTTATAATTTATAAATC  
 ACATAGTTTCCTAAGACCTGTCACTCTCTAAAGCCAGTTACTGCACTTGCCAACTAACTCTTGCTTTCTTTCAAAT  
 GATTGGAAATAACTTATGTATTA  
 AATTGAACATCTTGGTCCTAAAATTTAATGAATGTTAAAGCTATATAGTCAAACAGTGGATTCTCTAAAGTTACGA  
 TTGGGAAAGCAAAGTGTGTATGTT  
 TTCTTTAAAAATAAAAAATGGATACTACTAGTTGTTAATATTACTACTTCATCCTTTCTTTAGGTAGAATGCATTTGAA  
 AAGTAATTTTAAATCCTGTGGG  
 TTATGCAAGGTTGTTTATCAGACTACTGTCTGATCTGTCTGCCAGAATTTATTAACCAGCATGTCAGTGACCCATT  
 TAATTATAAAATTGCTCCACAGTT  
 GCATAAACAGTAATCTCAGCCAGCAATATTGTAAAAATTCATAACAAAGGAATATTGTCTGGTGAAAGTAGAAC  
 ATCCAGTTTCTAGAATAAGAGTCCCA  
 TCGTTTGTATAGGTATTGAAAACAATGCATGTCAAGACATGGAATAGGAAAATTAGTGTTGCTATGAGACTAGA  
 GACCATGGACAGTACAGGATTAGGCT  
 TGACTGTTATGTTGGTTTGTGTTGAGGACTAGAAGAGGTAAGCACAAATGCTTATGGCATAATAGGCTCTAATACA  
 TAGCGTGTGTTTCAGTTGAAGATGA  
 AACAAATGTTGCTGGGATGGGAAGGTGGTGAGCACAGGATAGAGGCAACTGGATTGGCAAATGTTGATAATGA  
 TGGAAGCTGGGTGTTTATTTTTCTATC  
 CTTTTGGTGCATGGAAAAATTTCTACAATATGTTAAAAAAATGACAATAACAAAAATCCAAAAGCTTTCTGTGAT  
 ATTTCAATTAAAAAGTTGTGTGAC

CAATAAGAAAAGAGTATGATTTTCATGGGAGACCAACCAGGTAAAGAATGATGAAAAAGGATTTTTAAAAATAA  
 TGCTTTGAAATTTCTTTTTTTTTGTT  
 TTATTATACTTTAAGTTTTAGGATACATGTGCACAATGTGCAGGTAGTTACATATGTATACATGTGACATGCTGG  
 TGCGCTGCACCCACTAACTCGTCA  
 TCTAGCATTAGGTATGTCTCCAGTGCTATCCCTCCCCCTCCCCCACCCACAACAGTCCCCAGAGTGTGATGT  
 TCCCCTTCCTGTGTCCATGTGTTT  
 TCATTGTTCAATTCCCACCTATGAGTGAGAATATGCAGTGTTTGGTTTTTTGTTCTTGCGATAGTTTACTGAGAAT  
 GATGATTTCTTTTAATACTCTT  
 TAAATTATGAACTTGAATGTTTCTTATATAATACCCCCAAATGAACTTATAATAGAAAATATCCAAAGTTGCCTT  
 TATGTAATATGTTAGGACATTTAA  
 AATATTTTTATGAAATATTTTAGACATGAAAAAAGATAAAGACTAATTAGAACTTTAAAAATACATAGGTCAAC  
 AGTATATCACTGTTGTAAAAATA  
 TTAACAATATTAATCTAAGTCATCCCAAACATGGAAAAGTCCAGAAAAGCATAGCTAGGATGGGGAGAGTATCT  
 AGAAACCATGGGGTAGAGGAGTGGGT  
 TAAAGGTTGGAATGATTCAGTAACTGAAGAAGAGAGGATGGAAAGTAGACCTGGGAGCCATCTTCCAGGAAA  
 GGAAGGTTGTTATGTGGAAAATAGAAT  
 TGTTCTGCTTGTCCAGAAGAGTTATCTGGAACAGGTGCAGGAAAAATATAGAAAGGTTTCTTCTAACATAAGG  
 TTTTTGATACAATCATTGTATTTA  
 AAAACCAGAAGAGCAACCCCCCTGCCTTTTTTTTTCTACTTTTAGGCTTACTATCTGAATAATGAGCATCTTAAT  
 GTTGACAATCATCACTGTATGCC  
 ATTGACCCATTTTATAATCTCTAACATATTCCAATTAATTATATGTTGGTCTATTAGTCAGACATCCTTTATTAGTT  
 CTCTGGTGTGAAAGACTGTGTC  
 TGTTCACTTTTGTACTCTTAAGGGGGATCTGTAAATTTAAAGTATTTTAAATGAAATATTTTAGACATGAAAAA  
 GATAAAGACTAATTAGAAGTTAAA  
 AAATATATAGGTCAAGAGTATGTCAGTCTAAAAATATGAGAAAATGTAGAACCAATAAATACTGCTAGAAACA  
 AAAGCTAATACTTATTGGTTCTATGT  
 TTTCTCATCTCACACATCCCCAGTTCTTCAACGTTAATTGAGACTTTGTCAGCAAAGTCTCAATGTATGAATGTCT  
 ACATTTATTGTATGAATGTATAT  
 TCAGATGAGGTAAATTTTTTAACAATGATATACCGTTGACCTACATATTTTTAAAGTTCTAATTAGTCTTTATCTTT  
 TTTTATGTCTAAAATATTTTAT  
 TAAAAATACTTTAAATTTACAGAGCCCCCTTAAGAGTACAAAAGCCAAACAGACACAGTCTTTTTGAAGTTTAGA  
 ATTGACAGGATTATAAAGAAAGATT  
 CAAACACTTTATGTTTAAATGCAAAGAGACCTAAGGACAATTAGAATAATTTTCTAGTCAACCTGCATTGAGGTTTT  
 TGTGCTTATTTTTACTTGGCCAGC  
 ACATCGAGACGTGAGCCAAGCCTTTCAATATGAGTTTTTTTGGTAGAAATTTTACTGAATGTTTATTTCAAGTAA  
 AGAAGCATTAGTCTGTCCTAGCAG  
 CAGGCCAACCCACTCAGGAACCCAAAGTAAATCTGTTCCAGAAGGAGTTGCAGAGAAGAACTTTGATCAGCTTG  
 TATAGCCTTCTGTGTTAGTAACTTGG  
 TTGACCTCAGTATTACATGTGTCTGTAAAAACACAAGTCCTATTAAATTAGGATAAATTAATAAATTCCTAATTT  
 AAAATGATCAAAAAAATATTTTT  
 TATCATATACCTTTTGTGCCTTTAAACCAAAAGTRTGGTGAAAGAATTGTAATCCACCAAAAAAAAAAAAAATCAT  
 ATCAGTTTTTTTGTGGCTTTGTAAT  
 GCTGACATGGCTTCTTTTTCTACCTTTTTTAGTTCTTTTGTGGCTGGTTTCACTCTCCAGAGGCCTGCTACAGCC  
 ATAAGTAGCGTTTTTAGCATACC

ATATTCTGCAAACCTGCATTAAATGCCACATTTTCTGTTATGAAACAGTTTGCCTGGTTCAAATTATATCTTACTCTT  
 CCTGTCTTTTGCTGGTAAATCTG  
 TTTTGTTTAACTCTTGGAGACCTCACCTAAGTTTTCAAGAAATATGGTTTGTTTTGGCTGTGATCCTGATGGAAA  
 AAATGAACATGATCTCTCATTTC  
 GTGTAGAAGTTCTCTGTGCAGTTTCTTAGTAAATGGCAATCTTGGAACACAAAGTTGACTTTATAGGACAGCAT  
 GCTTATCTTTCTTTAAAGGGCTATT  
 TAGTGTTGCCACACATCAAACCAAGGGCTTAAAAACAAGCAGTCTGTTTAAATCTGCATTACACTCACTACTTCCA  
 CCTACCCATCATAATCGCATAAAC  
 TTTATGAAACATTACTGCCCTTGTTTTAAGGTTATAGGAATTACATTGTTTTAGTTATTGGAACAAAAGACCAGAA  
 AAACAAGTGGTAAAATTCTACATC  
 TTCCACAGTGGGCCACATTGGAGCAAATACTCTTAGTGGAATCAGTAGGTGGGAATAATGGAAAGAAAAGTAT  
 CCAGAACTGATATAGAACTATAAATA  
 ACCTCAGATGTGAGGTTGATGAAGGGATACAACTACTGTTAGGAGTCTCCTTCCTTCCGGCTCAGTCACCCAC  
 TGAATATCTTCCAAACTCTGTATCA  
 CCTTTACCCTGTTTCTAAATCCTATAGTTAGGAACTTGATCCCTCCATCCCAAATGACTTCCCTTTGGGAGTC  
 CCTGTCTTGGGTTCTAGAGTCA  
 TCATGAGAAGGTAGGTGCTGGGGGGAATCCTTTGCTACTGGTTTGCAGGTATTAGAAATGGTGCGAAGGGGAT  
 TCTCCTGTCATACCAAATTAGAATTT  
 GTATTTCTTTCTCTAGAATTGGTGTACAGTTAATTTAGGGATTTCAGGTATTTTAGTATTGTTCTATGGGAGTG  
 ATGTGACATGGGCCTAGATGGATG  
 CCCTGCACACAGTGTTCTCTTGCAGTGGAGGAACATGAGCTTGGAGAATCTACCATGTGTATAGCAAGCAG  
 AAAGCAAGCCTCCTCATTGTCTAGAG  
 GGAGATTTTATCTCATCCTTCAAGGTTGCTTGTTGGAACTGAACCTTGAGAATGGCCTGGGAAAAGAGTGGTC  
 CGGACCTTGCATTCTTGCATACCCA  
 ATAAGACGTGTAGGAGCAAACTCATAAGAGACTCCCTGGAGGTGTGTTTCTCAAGAAAAATATGTCTATAATA  
 ATCTATTTGCAATGCATAGCAACTCA  
 GAAAAACAGAGACTGCCTATGTTCTATTTCAAGAAAAAGCAAACTTCAATAGACGTTGAAGTTGCTTATCTGT  
 AACATCCAAGAATAGGACGCTGGTT  
 AGGTCAAAATGTGAGCTGAATCTCCTTAAATTATGTGGCTAGCTTAGATCTCAGAAGATAAATAATGAATAGGCA  
 CTTCTTAATAACTTCTAACAACAAA  
 CAAGTTGAGATCATGATATTTCTTTAGGCTGCAAAGTAGAGGAATCTATCAGAACAAATCCAATAGGAAGGATT  
 GGAGAGCTGGAGTTATCCAGGGTTT  
 AGTAAAAAATTACTTCTGTGCTGACTGCATTGTGGCTTTGCTGCTACCAGGGACTCTGCAAATAAAAGAAAATA  
 TGCTTTGTTTAAATATTTGCAGTAT  
 CTCTTTAAATGTAGGTTTGGAGCTTCTAAAAATTTATCTTGAATAGCAGTTCAGTTTCATTTGAGGTTACAGTAGC  
 TTTGGAAGGGGAAGTGAACAGCTT  
 TTTTTGTTGTTGTTTATTTTCTTCTGATGCATTGCATTAGGATAAATTTATGTTGTGGCAACCTCAGCATTCTT  
 TATTCTGTACTTTTTGTACTTT  
 TTAGTTTGCATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACAAAGTCTTGCTCTGTTGCCAGGCTGGAG  
 TACAATGGCACGCTCTTGGCT  
 CACTGCAACCTCCACCTTCCAGGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGCTATGT  
 GCCACCACACCCAGCTAATTTTG  
 TATTTTAGTAGAGATGGGGTTTACCAGTTTGGCCAGGCTGTTCTCAAACCTCCTGGCCTCAAGTGATCCACCTGC  
 CTCGGCCTCCCAAAGTGCTGGGAT

TATAGGCATGAGCCACTGCGCCCAGCAGTTTGGCATCTTTATAAGAATAATAAATAATTATTTTTAAGGATTCTGA  
 TTTTATCTTAATAAACAATGAAA  
 TACTTCTCTCTCCTTTCTCTAATTTCCCTTCTCAAAACCTTGTGGGGACTCATTCTAGGGAGACAGTCCTTTTATACC  
 CCAGAGACTTTCTTGCAAATCA  
 GTTTCCTTGTCTGGTAGTCATTCTTACTAGTTTTGTGTGCATTCTTCTGGAGACAGTTCTTTAATATGTAAGTACTT  
 ATAGTTTTATTTATTGACCTATT  
 TCATAGAAAAGGTAGCATACTGTACATGTCATCCTGCAACTTGCTTTTCTCACTTATACCTTAGAGATCTGTTTCA  
 ATCAGTACTAAAAACTTTTCTCAT  
 TCCTTTTTTAAATTTTTATTTATTTATTTATTTTTTTAAAGACAGTCTGTTGCTCTGTTGCCTAGGCTGGTGTGCAGT  
 GGCATGATCATAGCTCACTATT  
 GCCTTAAATATCTGGGCTTAAGCAATCCTCCTGCCTCAGGCTCCTGAGTAGCTAGATGTGTGCCACCTGGCTAAA  
 TTTTGATTTTTTTTTTTGTTTTGT  
 AGAGATGGGGGTCTTGTTGTATTGCACAGGGTGGTCTTGAGCTCCTGGCCTCAAATGATCCTCCTGCAGTGGCCT  
 CCCAAAGTGCTGAGATTACAGGCAT  
 AAGCCACCATGCCTGGTCTTTTTCTTATTCTTTTAATAGCTGCATTTTAATGGCTACATAGTAACCTACTTCATAGG  
 GGTGCCATAATTATTGTAATAATT  
 TTGCAGTCCTTTTCATTATAAACATTTCTTCATTGAATAACTTGGTACTTATTTCATTTTGTTTCATATACAAATATAT  
 CTATGGGAGAAATTTTATTTTA  
 TTTTATTATACTTTAAGTTCTAGGGTACATGTGCACAACGTGCAGGTTTGTTACATATGTATATATGTGCCATGTT  
 GGTGTGTTGCACCCATTAACCTG  
 TGATTTACATTAGGTATATCTCCTAATGCTATCCCTCCACCTCCCCCAACCCACGACAGGCCCTGGTGTGTGAT  
 GTTCCCCTTCTGTGTCCAAGTGT  
 TCTCATTGTTCAATTCACCTATGAGTGAGAACATGTGTGTTGCGTTTTTTGTCCTTGTGGTAGTTTGCTGAGAA  
 TGATGGTTTCCAGCTTCATCCATG  
 TCCCTACAAAGGACATGAACTTATCCTTTTTTATGGCTGCATAGTATTCCATGGTGTATATGTGCCACATTTTTTTA  
 ATCCAGTCTGTCAATTCATGGACA  
 TTTGGGTTGGTTCCAAGTCTTTGCTATTGTGAATAGTGCTGCAATAAACATAAGTGTGCATGTGTCTTTATAGCAG  
 CATGATTTATAATCCTTTGGGTAT  
 ATACCCAGTAATGGGATGGTTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATTGCCACACTGTCTTC  
 CACAATGGTTGAAC TAGTTTACAGT  
 CCTACTAACAGTGTAAGAGTGTTCTGTTTCACTACTTTGTAATGATTGCCTGTTTCTGACTTTGTAATGATTGCC  
 ATTCTAACTGGTGTGAGATGGTA  
 TCTCATTGTGGTTTTGATTTGCATTTCTCTGATGGCCAGTGATGATGAGCATTTTTTTCATGTGTCTGTTGGCTGCAT  
 AAATGTCTTCTTTTGAGAAGTGT  
 CTGTTTCATATCTTTGCCCCACTTTTTGATGGGGTTGTTTTTTCTTGTAATTTGAGTTCTTTGTAGATTCTGGATAT  
 TAGCCCTTTGTCAGATGAGTAG  
 ATTGCAAAAATTTTGTGGGAGAAATTTTGAATTGGATTTGCTGGGCCAGAAGGTATGTACTTGTTTCAAGTTTTG  
 ATAGGTTTTACCAAATGGTCTCCA  
 TTGAGGTAGTTCTAATTGAGAAGTCAGTCACAGGGCTTAATGGAGTCCTTGGGTGGGTTTCAAGTGTTCATAAT  
 CAGCATCCTATATTAGTCAGTTCTC  
 ACACTTCTATAAAGAAATACCTGAACTGAGTGATTTATAAAGAAAAGAGGTTTAATTGTCTCATGGTTCTGCAG  
 GCTGTACAGGAAACATGGCTGGGGA  
 AGCCTCAGGAACTTACAATCATGACTGTGGGTGAGGGGGAAGCTGGCACATCTTACATGGCTAGAGCAGGAG  
 AAAGAGAGAAAAGAGGGAAGTGCTACA





ATAAGTGTACGTAAGATTCTTAACAAAAATCAAATTAAGTTTTTTAATGTATTTGAATATTTAAAGTAGGATTGCC  
ACAAATTAAATTTGCCAATGATTT  
TAACTATGTGTGTTTTTAAGAATGTATCTATAAATATGTTTACTGCAATACATATGCAATAATGTATGTTATGTAA  
TTTTTATTTAATTGAAAGATTGT  
ATTTTTATTTTTGGTCAGTTACAGTGAAGAGGTTAATCAGTAATGGCTTCCACTTTATCTATCAGTTTAGAGGGCA  
AGATTCAGCCAAATGAGAATAAGA  
ACTGTTAGAAAGATGTGCAGCAAGTCAATGATTACCATGTGATTCATCCATAGGTTATCAATACTTGATCTGACTT  
GAGTTTACATGAGATATTCGAATG  
TATTTCTTAGCATAAAAAATCCCATTTTAATTTTTTTTTTTTTTTTTTTTTTTTGGAGATAGGCTCTTACTCTGTCACCC  
AGGCTGGAGAACAGTGATGCG  
ATCATGGCTCACTGCAGCCTCAACCTCCTGAGCTCAAGTGATCCTCCTGCCTCAGCCTCTAGGGTAGCTAGGACT  
ACAATTGTGCACCACGATGCCCAGC  
TAATTTTTTAATTTTAGCAGAGACGAGGTCTTCCTATGTTTCCCAGGCTGTTCTCAAACCTCCTGGCCTCAAGCGTTC  
CTCCTGCCTCGACCTACCAAAGT  
GTTGGGATTACAGGTGTGAGCCACCACTACTGGCTCCTATCTAGTATTATATCTAATTTAATTTATGGTCATTGAG  
CTCATTTCTTAGCAACATCACAGA  
AACAAGGCTTTAGTCAATTTTATTAATAAAAAAGAGACTATAATAACTTACAGTGTTTATCAATCAATAAGGTTTTT  
AGTTTTAAAAACCTTATTGATTGA  
TAATTAAATTGATAATTAAATTAACAAGCTGGGCCAGCACTTTGAGAGGCTGAAGCAGGAGGATTGTTTGAG  
CCCAGGAGTTCAAGACCAGCCTGGGC  
AACATAGAGAGATCCTATCTCTACAAAAAAATAAACTAGCCAGGCACGGTAGCATGCACCTGTAGTCCCAGCTAC  
TGRGGAGCCTGAGGCTGGAGGATCT  
GTTGAACCTGTGAGGTTGAGGCTGCAGTAAGCTGCAATTGTGCCACTGCACTCCAGCCTGGGTGACAGAGTGAG  
ATTCTTTCTCAAAAAATTAATTA  
TTAACTAAAAACAAAATGTGGACGTACTTGAACTTTGACTCTTACATTGATTCAACCTTCAAAGGAACTTGTCC  
ATTTGTGAAAGTTTTCTTTTATA  
TAAGTCTTTCTTACAGTCAAATAAGTGTTTACTAAGTGAAATACTACATTTTATAGTGCAGTGGCACTCAAAGGAA  
ACTATTTTGCAGCTGTAGAGTTTT  
TTCTTTTGTCTTTTTACTTTTTCCAAGGCAAGATGCCTGGTGGGAGAGTTGTCTGGAGGCAGGATGATCAGAAG  
CAGGACCTGGCCAGCTGTTCTGCTTA  
TTTCAGTCTGTATGAAATCATCTGATATGCCATAAGAGAGATGTGTCAATTGGCTAATGGGTCAGAGTATTGAT  
TTTATGTTTTTCAAGATTGCTGACAGA  
GAGATTAGGCTTCCAGATACAAGCAGTTCCACAAGTCTCTGTTTTTCAAGACAGGCACATAATTAAAGTAACCACAC  
AGTTCTCTGAGGTTAGAGTTCCATG  
CTGTGATGCAAGCCTCTCTATTTGGTATTTAAAAGCAGTTTGAAGCTGTTGTGTAAGTATCACTGRACTTGGAGTC  
AACTGACCTAAGTTCTAGCCTTGG  
CCCTTCTACTGTCTTAACAGTGGATTAAATGATGAGGCTTATTTTAAAGAATCTTGAAAACCTATAGTAATTTCAAC  
AAGAGGAGATTAATTGAATATATT  
ATGAAACATCAAGATAATTGAAAATCTGTGCTGCCATTATTGCATGTCAAAGGTCATTTTGATACACTGTATGTGC  
AAAAGCAGTTTACCAAACAGAATT  
ACATACAGTGTAATCCTTTTTTCTTATATAGGAAAATACCCACTTCCYCCCACTTAAAAAAATACCAAACACCACA  
ATGTGTGTATGTTTAGCCTACAA  
ATCATATTTTTTTTGCCTAGGAAGTATTGAATAATTTAAATATTTTATATTCAGATAGCTAAAGATGGCTTGCTGA  
GATAAAACATTGGAGCCCTGTGTT

TGGTGATAGCCGTGGGTTTCATTTATTTTTACTGTCTTTATCCTCTGTATTTATTGATAGTATTGAATATTTATTTTC  
 AGAACTATTATAGTGTATCTGT  
 TTCTATGGGAAGAAATGAATATTTGACCACAGATGGTTTTTCATTTTTTTAGTATGGGGCACGTGGATTGACAGCC  
 TTTGAAACTTGAATGCTGTGTAATA  
 TTCAGGTGCTTTTATTGTTTATAGAAATGCCTAATCCTTGTTTGAAATTTATTAAATGGTATGCCATCTGTAAGCCC  
 CTCTAAATTCCACAGATTAATTA  
 GCCTTTAAAGTAAGGTTGCCTTTTGTCTAACTTTTAGTATAAACTCTTCCATCTGGGAACTTTCACTTTTATGTCCC  
 TGTTCACTGTTCTGTGTACTTCA  
 GCACTCTAGTATTAATTATGGACACTGGCATTCTAAGTAACATGAAGAACAATAAAAGTAGACTGAACA  
 AAGTGGCAATGTCCGGAGAATAGTT  
 TGCATTTGGTGAATAATCAATCCATCATGGCTTTGCCATTGCTCTTTGTATGTGCGTGCGTGTGTGTGTGTGTGAG  
 TGTGTGTGCATGTGCGTGTTTAGC  
 ATTACTTAAGAGTGAGATAGTTATTCTTCAGGGAAATTTTTACAGATGATGGTTTCACAAAAATCAGCCTTTTTGT  
 AGTTTAAATAAAAAAACACCCTTT  
 TTGACATTTTCATGTAACATTTCTTGTTGGGTTCTGATTTAAAGGTTGTTGAAAATATTTGGTGTGGCAGATCACT  
 TGGTTGGCTTTGGCATACTTCTCT  
 TTCTAGGAAATGTAGTATCAGGGTGGGAATTGTGTTGTTGTTTTAAAGAAATAGATTACAGGCTGGCACGGTG  
 GCTCACACTTGTAATCCCAGCACTTT  
 GGGAGGCCGAGGCGGGCGGATCAGGAGGTCAGGAGATTGAGACCATCCTGGCTAACACGGTGAAACCCTGTC  
 TCTACCAAAAATACACAAACAAAATTAG  
 CCAGGCGTGGTGGTGGGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGAGGGAGAATGGTGTGAACCTGG  
 GAGGCGGAGCTTGCAGCGGGCCGAGATC  
 TCCCCACCGTACTCCAGCCTGGGCAACAGAGCAAGACTCCATCTCAAAAAAAAAAAAAAAAAAAGGAAATAGATTA  
 CATTCTTTAACTTTATTTACTGGTCT  
 CAGTAAGAGTTATTGAACACTCTAGAGTGTTTACTTTTTAAAGTCCATTGCAAAATATATAGAAATTAGTAATTCA  
 ACCATCTGTGTACTCATACTAATC  
 AATGTAATATAGGTTGGCGCTTTTTATGCACTTAATTTTGCCACCTCTCTCACTTCCCCAAAGAAATTAGWCTAAT  
 GTATAACAATAGGCTTCATACTT  
 AATTTAGAAAACATAACAAAAATAATTTGCTCAGTAAGTTTCTGTTGTAGTAGATATTGTTGGTATTAAACAAAA  
 AAGTAGGATCAGATGTGGTTTGGC  
 TTTCAAGGTATAAAATCTAGGTGGGGAAAGAGAGAGAGAGTTAACATTCTTGAAACAATTGAGAGCATTAGAGT  
 GTTCCAGTATATCCTCCAAATTCTAA  
 GGKTATAAGATGTTTCAGAGAAAATAAAAAATCAGGGTGGGTAGATAATTGAGTTCAAATGGACCTTGGAATT  
 AGGTTAGTGGAGGAGATAGCATTCCA  
 AATGGAAATAAACAATGCGAATAAGGAAACATGTTAGACATGTATGGAGATTATGAAGTAGCCTTAAATGATAA  
 GATTGGAGGAGTGTGGCTTGAATAGC  
 TGTAGATTTAGGGCAGAAGGTCTGGTTAGGCCGTGGGAGGTCACAATATGGAAGACCTTAAAGAAAAGATGG  
 GTAGCAGGGATACATGACTAGCTTGTT  
 TTTGCAATTGAGTAGTAAGATAAAAAATATTTTCTTAGGAGAAGTTATCTTAGAAATGGTGGACAGGATGAGATG  
 GAACACAATTGAAATAGGAGTCTGTT  
 GCAGTAGTATAGGGATGAGTTAATGAGTGCCTGTATTAGGAGAGTGGAAGAGGGAATAGAGAGAAAGTAATT  
 ACTGAGACACATTTTCACTGGAATAAAT  
 TTGGTCATTAGTATATACTTTTAAACATATTCAATGAGTGAATGAGGGATTGAGAAGGACTGAATATAGTGAATAA  
 AGTAAAATAATAGTTACTATTGATA

GAAAATGGATAGATTTTTAAACAATCTAGTTAGAGGGGAGATAGTTCAGTTTGCACATAATGAATTTGAAGTGGT  
 AATAACTGTAAGAAGTGGCCTTATAA  
 TTAATTGAATATATGGTTCTGGGGCATAGCCATTTAAGGGCCAGAGATACAGATTTAAGATGTATGTATATATGC  
 ATATATATTTAAAGATACTTTCTGA  
 GGGAGTTTCTAAGGGGAAAATTGGAAGATTAATAAACGAGCCTTARAAAACATAAGAAGAAATTATTTAACATT  
 TGGGTTTTAGAAGGCCTTATTTAATC  
 ATTTGATTTTAACTTATTTAGTTTGAGCAGTGTGTTTTCAAACCTCTTTTGAAACAACCTAGTAACCTATGCTTTGAA  
 AAAATTGTTGGTTAAAATAATTG  
 TAAAAATGAGTATATTCTGATCTATTTTGATCTGATTTTAAAACCAGGGTATTCTGATTCATTATAACTTTTTCTAG  
 ATTGCTGTGCTAAGGAGACAGCA  
 ACGTATGATAAAAAATCGAGAATCCGCTTGTCAGTCTCGCAAGAAGAAGAAAGAATATATGCTAGGGTTAGAG  
 GCGAGATTAAAGGCTGCCCTCTCASAA  
 AACGAGCAACTGAAGAAAGAAAATGGAACACTGAAGCGGCAGCTGGATGAAGTTGTGTCAGAGGTAAGTGTTA  
 GTAATASGGCTGAGTCGAGATGGGCTA  
 AAGTATYCTCTGGATTAATAAATAGAGAACTTTTACATTTTAAATTAAATTATGTAAAATAATAGTGCTAGGAAA  
 AGGAGGGTATAATAAGTCCTAAAA  
 CATTTGCAGTAAGGATAAAGCATTTGAAGGGGACCCTTTGAATTGTACTCCCCAATCTACATAGAAAAGTTGTCT  
 CCTTTCTAGGTTTCKCCTCATTC  
 ATTGCCTTAGCACCTCTTTTCTGGCTAAGTGAATCACACTGAGGCTTCTAAAGTGCCCTGCCATATTTTCCCTTTT  
 AAAAAAGAAAACAATATAAAAGT  
 GTTTTTATTTTTTCTCAGTCCACCTAACAACTAATGTGATTTTTGCATTCTGTTACTATGCTGACTCCAAGAGGTTTT  
 AATCTCCTCAGTTCTATTCGGG  
 AAATTTGCTGTTTTCAGAATTTCTCACTGCAGTGCTGAACCATATAGATGCCTACTGAATGGGGCTTCTTACACTC  
 TTTTCTCTTGCCCTGAGTACAT  
 GGAATACTCCTATATTTTACTGAAAACAATGATTTAAAAGGTTAGGGAGGCAAACCTATAGAAGTGACTCAAAA  
 TTCTTTGGGGATCCGTAAAAATATAA  
 AAAATGTTGAATTCTGGYCGGGCGTGGTGGCTCACGCCTGTAATTCCAGCACTTTGGGACGCTGAGGCAGGCGG  
 ATCACAAGGTCAGGAGATCAAGACCA  
 TCCTGGCCAACATAGTGAAACCCCATCTCTACTAAAAATACAAAAGTTAGCTGGGTGTAGTGACACCTGTAATC  
 CCAGCTACTCGGGAGGCTGAGGCAG  
 CAGAATCGCTTGAACTGGGAGGCGGAGATTGCGGTGATCCGAGATGATGCCACTGCACTCCAGCTTGGCGAC  
 AAAGTGACACTCCGTCTCAAAAAAAYA  
 TATATACATATATATATTGAATTCTTTCCTCTTGGGTGAAGAGGCAACTGCAATGGAGTGGCTGCTGTGTTTTTT  
 CTCTTCTGCTTCTCTCCTTTTCTC  
 TCCCTGCTGGAAAAGGGAACCTGGAAGTCTCTGCTCTAGTCTTGAAAAGTGTTATTTCTCACCTTTGATTTTTTTT  
 TTTTTTTAAATAAGACAAGGTA  
 TTTGTAAGACAAGAGATTTACGTAACCGGTAAAGAGGAAAATTCTAATCTTAAGCGTTGCTTTTTCTGAAATCGA  
 TGTTAGTTTAATTGTATTTAATGTG  
 GTCATTTCTTTAGAACCAGAGGCTTAAAGTCCCTAGTCCAAAGCGAAGAGTTGTCTGTGTGATGATAGTATTGG  
 CATTTATAATACTGAACTATGGACC  
 TATGAGGTAAGTGAATAGATATTTATTTGGACACTAATGCTAAAACTTAAATTCTCATTATTCTTTAGCTTTTGT  
 CATAAACTAGAGAAAAAATGATT  
 TGTTTTTCAAGATGTGTAAAGAGTAATGAGACAGGAGCCATGCCTTCAAAGATCTCATGATCCAGCTGGGAGGA  
 GAGACAAGTAGAAGGATGACTATGGT

ATAATGTGATGAACATAATAAACAGTGTGAATGAGTCCTAACATAATATAGAGTACGAGTTCCTTGAGGTGA  
 AGGTTCTTGTCGTATTGATGGGTACT  
 CGATACATTTTTTTGTTGAGTGACTGAGTTATAAGGGAGATACCTACGAAGGAAGCTGAACATTGAACTGTAATAT  
 CAGTTTGCCATGTGCAAAGTATGAG  
 AAAGGCATGCCAGGCAAAGGGAACAGCATGTTTAAAGGCACAGTTATGAAATGGCATGACATGTTTCATAAAAA  
 AGTGAAAAGTTCAGTGTATCTTGTATA  
 TGTGGGGAGTGTGCTGAGGCTAGATATATGGTTAGAAGTATAGGTGCTGCTGTGTACAATGAGGCATCCCTAG  
 CTTTATAGGTGGAGGGTAACATAATCA  
 GCTTTCATATTAGGAAGAGGACTTTGTGTGTTGTATTGAGTATAGACTGCAGTGAGGAGAACTATTGGCAGA  
 GAAAGCATTAGAAGACTATTGAGGTG  
 TTTCAGGTAATAAATGAGAGTCTGAACCTTACACTATGGTTATAAAAAGGATGAGCAATCTTAGGGTACCTCTCTG  
 ATGTAGMTTTTTGTGGGATTTGTAA  
 CTGGTATAATGTGGGACATGAGGAAGAAAAGAGGGAAGTTAAGGATAATTCAAAAATTTCTACCTTGAGGGAC  
 CGAATGGTGGTGCTGTCATCGGAGGAA  
 GCAAACATAGACAGATGGATAGTTTTATGAAGAAAGACAATAGTTTTTATTTTAAGTATAGGAAGTTTGTGGCAC  
 TGACTIONATATTGAGAATTAACATT  
 ATAGATAGGTAGGTAGACTGACAGACAGATAATGTTAGTGGTGTAGATGAGATTATCCAAGGAGAACAAAGTAG  
 AGGAAGGAAAGTAACAAGAATCTGGGG  
 GACACCAGTATGAATAGTAGATAGCCTTGGGAAAAGAACTCACAGCATATTAGGGGCAGAACTGAAAGAAAG  
 TAGCTTTTTGAAGCCAAAGTAAAAAGA  
 GCTTCAAGCAGGTCTTGGGCAGTGGTATCACATGCTGCTGAGAGCTTGGGAGAYGYAGAAAGCTGAATTTGGCA  
 ATTATCAGGTCAGTTGTAATCTTGGA  
 AAGCAGATTTAGTAGGGGTAGGCATGAAGTCTGTATTAATAATTGGTTGAGAAATTGAAAGGAGTTAAATTGAA  
 TAATCTCTTTTTTAAAAAAAACATT  
 TTTTATGAAGATTTTCAAACATATAGGAAATTTGAGAAAATAATATGATCAACCCCTTGCTCTCATGAAATTTATA  
 ATCTAATATAGGTGACATATACTA  
 ATCAAAATGATCACACAAATAAATGCAAAATTGTAACAATGTTAGGTGCCATGAAGGAAAGGTACATGATATCA  
 TGAGGCCACATAATAGGGTCACTTGA  
 CCTAGTTAGGGAAGTAAGGCAAGGTTTCTCTGAGAAAGAGGTGATTGTGTTAACATCTAAAAGAAGATTAGGAA  
 TTATTTAGGTAAGATACAGGAAGAAC  
 ATTTTCATATAGAATATATAAACATCTATTCCCACATTTGTGTTGTGTCTTTGTGTTTATATGTATTTAGTTTAAAAAT  
 TTTGGTTTTACATGAAACAATT  
 ATATACAAATTGTGCTTTCTTTAAAATATGTGAATACATTATTCTCTTTAATACTACATAGTTTTTTGTAGTGTA  
 ATGTGCCATAATTTAGCAATTCT  
 CTTTAGGTCTTGGGAACCCCTTTGTACCCTTAAAAGTTACCAAGGATCCCAAATAATTTTTGATTTTGTGGTTTATA  
 TATACTGATATTTATTATATGAG  
 AAATTA AAACTTTAAAAATAAAAAATTTTTAATTCATTTAAAAATAATAAACCGTAAACCATTACTTTACATGTTAA  
 CATAAATAACATGTTAATGAAGA  
 ATAATTATATCTCCCAAACCTCCAAAAATTTAGTGAGATACAGAAGAGTGGTACTACTTTATAGTTTTGTAAAAAT  
 ATTTTTTTTGGGGGGAAGCTCAA  
 TACATCAGTTTTAGTTTTCTCTATATACACATTTATTTGCTACTAAAAATTGAAATAGCATATTGAAATTTCTTTT  
 GTTGCAATCCAAAATTTTGA  
 ACCAGAAAACCAAATTAATTATTATTATGCTGTGGCCAGACTACTCAGTATTTTCTTTACATTCTTCTCTCCAAGTT  
 ACTTTCCTTCAAATTCCTAATGT

TGGGTTTTATCTCACTGACTTCAGGCTTCTAAAACACATGGGGATACTTGAGACCTATTGTTGCTTCTTTAGGTCA  
 AATCAATAGTTTCATAAAAATCTT  
 TGATGTCCGACTTAACAGGATCTGGCAGGATTATCTGTTTTTTCATGTAATCTGTTATGATATATTGTTGGGTTG  
 AAATAAATGAAGAAAATTTAGACT  
 CACATCGGTATGCAGTTGGAAAAGGGAGAACTTTGCAGACTTTGAAAGACTTTTAGGATCCCCAGARTTCCTTG  
 GACTGCTTAAAACCCCAGTGCACACT  
 TGCCCAACTGCTGCCGTAATTTATTTAACCATCTTTTGATAGGCAGCTAGGGTATTTCCAGTTACTTGTTGATATG  
 TGTACATAAAATGTATTTCCATAT  
 ATTTATGTGTCTGTATGTGTATGATAGAATCTAGACTAGAGCTAGAGGTTACGATTATCTGAAATTGTGATAGAG  
 TTTGCCAAATTGCCTTCCAAAAAGA  
 CTGTATCCGTTTACATTCCCACTAGCAGGAGAGAGGAGTATCCATTTTCTCGCATTTCTAACACTGGATATTGTA  
 TTTTAAAATTCTTGCCTATCTTG  
 TGGATCAAAAATGTTATTGTTAGCCTGTCACATTGGCCTGTGTGCTCTTCTCAAACACTCTAAGCCTGCTAGCAC  
 TTCATGGTCTTCACTTTTATTCCC  
 ACTGCCGGCAGTGCCCTTCCCTCAGAGATTCTTTCTGTCAGGTGTTAAGCTCAAATACCATGTACCCCTGAGCTTC  
 AAATAAAAGTTTTAAAAAAGTGAT  
 TAAAAAACATCAAGTTTTTTTGGTCAAAATTAAAAAAGCTCAAATGCCACCCCTGTGAGGCTTTC  
 CTGGACTACCTTTTCTAAATGTCTA  
 TCCCAGCTCCTGCCATTGCATTCCCTATCGCCTTTCTTTCTTCATTTTCTACATAGCACCTCTCCCCTTCAAGTATG  
 CTATGTAACCTAYTTATTTTCT  
 TTCTCTCCCTGGAGGTATTTTTATCTCTTTTGTTCAGTGCTTAGAACAGTGGCTGGCACATAGTAGTTACTCAA  
 TAAATACTTGTAGTATGAATGAA  
 TGTCTCCCTCACCCAATTAGTAGGAATTATATATGTTCATAACTATTTCTCAAAAACTGTGTGTTCTTTTAGTTC  
 CTCATTTCTTATCTGGTATACGT  
 GGTACTTATTTATTTAGTCAGACATACAGCAAATATTTATAGAGCACTTAATGTTTGCCAGGTACTGAGCAAGGA  
 ATCAAGGATGCAAATATGAATAAAA  
 GTACAGTCTAGTAAGAGAATCAGATATGTAAACAACTGTTGTATGAGAAAGAATTATATGGATGTCATATAAAC  
 ATATAGGCATTAAAAAAATGAAAAAC  
 TTTTGTTTACACACTTCTGACACCAGATATGTGGATTTTCTACACTGAGCAATTCTCCAATTCTCTGGACACCAAC  
 TGGATGTCCTACAATTTGACTCA  
 ATTCTTTTTTAAATTTTAATTGTAAAATATTTTTTATTGAGACAGGGTCTTGCTCTGTTGCCAGGCTGGAGTGCA  
 GTGGTGTGATCGTAGCTCACTGCA  
 GCCTCCATCTTCCAGGCTTAAGCAATCCCCGATCCTCAGTCTCCTAAGTAGCTGGGACTACAGGTGCACCTCACC  
 ACACCTGGCAAAGTTGCTCAGGCT  
 GGTCTTGAACCTCCTAAGTTCAAATGATTCTCCTTCTTGGCCTTCCGAAGTGCTGGGATTACAGGTGTGAGCCATT  
 ATATCTGGCCAATTTAAGTCATT  
 CTGACACTACCTGGAGCTAGCACAGGCTCCACAGGTAAAGGGCTGAGTATCCAAAGACTGCCCCTGACTTCAG  
 TCACCAGTTGCAAGTCCCAGGTTGTG  
 ACCTGTAACCTTCTGACTAAGTGGCTATAAGTTGAGGGTCCCACAATTCCCTCTTGAGGTTTGATAATTTTCTAGA  
 ATGGCTCACAGAACTCAGGAAAAC  
 ACTCTATTTACTATTACTGGTTTTATTATAAAGGATATGTGGGGCAARCTAAATGGGAGAGACGCATAGGGTAAG  
 GTATAGGGGAAGAGTGCAGAGCTTT  
 CATGCCCTCTCCAGGTGCACCACCCTCCTAGTACCTCCTTGTGTTTGCTAACCTGGAAGCTCTCCAAATTTCATTGT  
 TAGGGGTTTTTATGGAGCTTCAT

TAGGTAGGTGTGATTGATTAAACCATTTGGCCATTGGTGATTGCACTTAAACTCCAGCCCCCTCTCCCCTACTTGGA  
 GGTGAGGGATGGGACTGAAAATC  
 CCAACCCACTAATTATGTGGTTGGTTTCTCTGGCAATCAGCCTCCATCATCCAAGAGTCACCTCATTAGCATAAAC  
 TCAGGTCCAGCTGGAAGGGGCTTA  
 TTACGGATAACAAAAGATGCTCTTCTTACCCCCTATCACTCAGGAAATTCCAAGGGTTTTAGGAGCTCTGTGACA  
 GGAACCAGAATAAAAACCAAATATA  
 TTATTTCTTATTATATCGCAAATTCACAGGCATATTCTTGAGGGAACATAGAAGTAACCAGTAACGCTGCTAGTG  
 GCAATCATGGGAGGCTTTATGCAGT  
 AGTACTCGATGAAGGAATAAAGAGGATTAGATTATGCTACAGGAAGTGACAGTGGGGTTAGGGGGGACATCTC  
 ATTGTGGAAGCCCTCATGAATCAATAG  
 AATTTTAGGAAAGAAAAGACTGGAACAAAAAAATTTTTTGAGTGCCCCCACCACAGACAAAAAATTTTTAAA  
 AATGAATTAATGATCTTCACAGCTG  
 CTGAAGTAAAGGTGGGGATGGATCTCATTTCCATGGTAGATTGATTGATTATGAAAGGAAGAGATGATGGTCCC  
 TATTTTATCAAGAAGTGGTTTGTGTT  
 CAAAAAGGAGTTAGTAGAAAGAGGGGATTAAAAATCGGAGAAAGCTTAAATATTTGAAAATCTTGAGGATCAA  
 TCAGTTAAGATCTGTATCTTTCTCTGT  
 GTGATGCTTTAGGAACAGCCAGCGAAGAAGTAGTGGAATATTAGGAACTATTCAAGTTAAAAAACATGTAA  
 AAGGCTTAACTTGAAGAAAAGCATCA  
 AGGAGAAAAGCTGGAAGGAATGAGGAAGTGAGGAAAGTATCTTAACAAGTCAGAATATCTTTGACACCAACA  
 AGGAAATATGGAGGAACTGGCTGTGA  
 AATTTATAAAAACCAAGAACTATTTCTTGTAGACAAAATTTGGCATGAACAGTGAAATATTAAAGCCCAGAGAAC  
 AACAGTGATTTGAAAGACTGAAATT  
 TGTCCCTGAATCAAATGTTGGGAATTGAGATGTTTCATCATTTGTGTTATTCAAAGAATTTTTTGATTGGCCCCA  
 TTTTTTTTTTTTTTTTTTGAGAT  
 GGAGTCTCAGTCTGTTCCCCAGGCTGGAGTGCAGTGGTGYGATCTTGGCTCACTGCAACCTCTGCCTCCCAGGTT  
 CAAGCGATTCTCCTGCCTCAGCCTC  
 CCGAGTAGCTGGGATTACAGGCACCTGCCACTATGCCAGCTAATTTTTGTATTTTAGTAGAGACAGGGTTTC  
 ACCATGTTGGTCAGGCTGGTCTCGA  
 ACTCCTGACCTTGTGATCTGCCACCTCAGCCTCCCAAATGCTGGGATTACAGGTGTGAGCCACTGCGCCTGGC  
 CTGTTTTCAGTCTTTTTACTGCAAG  
 GCTATGATCTTCTCTGGTCCGGGAGTCTGCCAGAGGAGCAGGAACTTCCCAGTTGAACCTTATTAGTAACCACCT  
 GGAAGATATTTATTGAACAAGTGTC  
 ACATCTGTAGTCTTGTGTATCTAGAAACCTTAAGCATGCATTATTATGTTTTAATTGGAAAATGCATAGGCTTTAA  
 AAAAAAGATTGTGAGCTCTATCT  
 GAACTTTTGTTAAGTACCACAAAATTAAGTTTGTCAGACAAATAGACATTTGAATCAGAGATTTCCAAAGTATTTT  
 AAATTTAGGAACCTAACTCTTTT  
 AGARCAGATCACATTACCGAGTAAAGTGATTAATGTGGTTTCAGATGTTTCAGCGTAGAACTGAATGTTTGTGCCT  
 TATTGGTATGGGCCAAATGAAAAAA  
 AAAAAATAAACACACAAGTTTGCCAAATATTGGTATGCCTTAAACATTTCTGAAGGCATTTGGTGATGAAATA  
 TAAATATATCTAATGTATAAAGTAA  
 GAGAGTTTTGTCAAAGGTAATATGCAGCAAAGCTGACTAGTTTGAAGAGGTTTGGGTACTGAATAGACGTCAC  
 ACAGGAAAAAAAAAATAAACTAAATGG  
 ACTTATAAGGGTTAATTTTTCTGACCTTATCGATATTTTGTGGATAGTAAAGAAGTCATCTAGCTATTTTTGGCAT  
 AGGATTCAAGGAGCTGGCCCACTC

TGGGTGGTCTATAAATAGTAACTCTAGGTAGCAGGGACTGGCTAAACAGGACAAATGTGAATTGGTTAAAAAG  
 AAAATCCTCTTCTTATTACTCATTGGG  
 GTTAATTCTATGATTATTAGTCATATTTTATTGTTGAGAACCTGAAATTTAACCCAGGTTTTTTTTCTTTTTTTTAAA  
 TTATCACTACAGACTCTGCCTT  
 TTCTACTATGCTGTAAAGTGTTACTTTATTTTATTTTTCTACCTTAAGGTTAAGGAACCTTAACCTTTTGTTTTGG  
 CAATATGGTAGGCTAAAGGTAC  
 AGGAAACTCATTGCAGTTTAGCACACTAAAAAATGATGGATTGAAAAAATAAATCAAAACTCTTATTTCTGGCT  
 CCACTGGAGACAAAGTAAGGGCTTC  
 TCAGAGGCCAGAAATAATAAAAGATAAGATGAGTCTATAATGATAAGCTAGCATTGGAAGTGGTGCTTGTTCTG  
 GATATAGTTCCTGATCCTTGCGGCGT  
 AGTGTCTGGGGATGGGAAAGAAAGCTGTGGACCCAAGTGAGGTGGGAAGTCACAACTGAGGCCTCCCTTTTCC  
 TCTACCCAAATAAAGCCAGAACTTCCA  
 AAGAATGACACTCTTGGTTGCCAAAAGAGGGGAACAAAAGTTGCCCTACCACTCAGGGAGACTGTTATTAAGTA  
 CCTATCTTGGTTTTGGCTTTAGTTAG  
 AGTGGAACAAATGAAAAGTCTTTCCTGATAATTCTTTACTCAAAGCCCCTACCTAATCAGGTTAGAGCCTAAATT  
 TATACTACTAGTGTGATCCCCCA  
 AAATCAAGACAAAAATGCAGTTTAGGTGGGCTTATGTTGGTGGTGCCTGTAGGCACCTGAGGGAATTGAATGCA  
 AATTCTTCTCAGAAGAGACACTTTAT  
 AAAAGAAAGTGTCCATCCTAGAAAACTGGAGACCCATCTCATATGCAGAGACACACACAGGCTCAAATAAAG  
 GGATGGAGGAAGATCTACCAAGCAAA  
 TGGAAAACAAAAGGCAGGGGTTGCAATCCTAGTCTCTGATAAAACAGACTTTAAACCAGCAGAGATCAAAGA  
 GACAAAGAAGGCCATTACATAATGGTA  
 AGGGGATCAATTCAACAAGAAGAGCTAACTATCGTAAATATATATGCATCCAATGCGGGAGCACCCAGATTCAT  
 AAAGCAAGTCCTTAGAGACCTACAAA  
 GAGACTTAGACTCCACACAATCATAATGGGAGACTTTTTTTTTTTTTTTTTTTTTTTTGGTGAGACGGAGTCTAG  
 CTTTGTTGCCAGGCTGGAGTGC  
 AGTGGCGCGATCTCAGCTCACTGCAAGCTCCGCCTCCCGGGTTCATGCATAATGGGAGACTTTAACACCCCACTG  
 TCAACATTAGACAGATCAACGAGAC  
 AGAAAGTTAACAGGGATATCCAGGAATTGAACTCAGCTCTGCACCAAGCAGACCTAATATACATCTACAGAACTC  
 TCCACCCCAAATCAACAGAATATAC  
 ATTCTTCTCAGCACCACATCGCACTTATTCCAAAATTGACCACATAGTTGGAAGTAAAGTAGACCAATATCCCTGA  
 TGAACATCGATGAAAAAATCCTCA  
 ATAAAATACTGGCAAACCTGAATCTAGCAACATATCAAAAAGCTTATCCACCATGATCAAGTGGGCTTCATCCCTG  
 GGATGCAAGGCTGGTTCAACATACA  
 CAAATCAATAAATGTAATCCAGCATATAAACAGAACCAAAGACAAAAACCACATRRTTATCTCAATAGATGCAGA  
 AAAGGCCTTTGACAAAATTCAACAG  
 CCCTTCATGCTAAAACTCTCAATAAACTAGGTATTGATGGAACATATCTCAAAATAAAAGAGCTATTTATGACA  
 AACCCACAGCCAATATCATACTGAA  
 TGGGCRAAACTGGAAGCATTCCCTTTGAAAACCTGGCACAAGACAGGGATGCCCTCTCTCACCCTCCTATTCAA  
 CATAGTGTGGAAGTTCTGGCCAGG  
 GCAATCAGGCAGGAGAAAGAAATAAAGGGTATTCAATTAGGAAAAGAGGAAGTCAAATTGTTTGCAGATGACA  
 TGATTGTATATTTAGAAAACCCCATTA  
 TCTTAGCCCAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAATCAATGTGCAAAAAT  
 CACAAGCATTCTTATACACCAATAG

CAGACAAACAGAGAGCCAAATCATGAGTGAACCTCCATTCAACAATTGCTTCAAAGAGAATAAAATACTTAGGAA  
 TCCAACCTTAAAGGGATGTGAAGGACC  
 TCTTCAAGGACAACCTACAAACCACTGCTCAATGAAATAAAAGAGGACACAAACAAATGGAAGAACATTCCATGC  
 TCATGGATAGGAAGAATCAATATCST  
 GRAAATGGCCATACTGCCCAAGGTAATTTATAGATTCAATGCCATCTCCATCAAGCTACCAATGACTTTCTTCACA  
 GAATTGGAAAAAACTACTTTGAAG  
 TTCATATGGAACCAAAAAAGAGCCTGCATCGCCAAGTCAATCCTAAGCCAAAAGAACAAGCTGGAGGCATCAT  
 GCTACCTGACTTCAAACCTATACTACA  
 AGGCTACAGTAACCAAAACAGCATGGTACTGGTACCAAAACAGAGATATAGACCAATGGAACAGAACAGAGCC  
 CTCAGAAATAATACCACACGTCTACAA  
 CCATCTGATCTTTGACAAACCTGACAAAACCAAGCAATGGGGAAAGGATTCCCTAAGTAATAAATGGTGCTGGG  
 AAAACTGGCTAGCCATATGTAGAAAG  
 CTGAAACTGGATCCGTTCTTACACCTTATACAAAAATTAATTCAAGATGGATCAAAGACTTAAATGTTAGACCTA  
 AAACCATAAAAACCTACAAGAAA  
 ACCTAGGCAATACCATTGAGGACATAGGCAAGGGCAACGACTTCATGTCTAAAACACCAAAAGCAATGGCAACA  
 AAAGCCAAAATTGACAAATGGGATCT  
 AATTAACTAAAGAGCTTCTGCACAGCAAAAGAACTACCGTCAGAGTGAACAGGCAACCTACAGAATGGGAG  
 AAAATTTTTGCAATCTACTCGTCTGAC  
 AAAGGGCTGATATCCAGAATCTACAAAGAACTCAAACAAATTTACAAGAAAAAAACAACCCCATCAAAAAGTGG  
 GTGAAGGATATGAACAGACACTTCTC  
 AAAAGAAGACATTTATGCAGCCAAAAGACACATGAAAAAATGCTCATCATCACTGGCCATCAGAGAAATGCAAA  
 TCAAAACCACAATGAGATACCATCTC  
 ACACCAGTTAAATGGCAGTCATTACAAAGTCAGGAAGCAACAGGTGCTGTAGAGGATGTGGAGAAATAGGAA  
 CACTTTTACACTGTTGGTGGGACTGTA  
 AATTAGTTCAAACATTGTGGAAGACAGTGTGGCAATTCCTCAAGGATCTAGAACTAGAAATGCCATTTGACCCAG  
 CCATCCCATTACTGGGTATATACCC  
 AAAGGATTATAAATCATGCTGCTATAAAGACACATGCACACATATGTTTATTGTGGCACTATTCACAATAGCAAA  
 GACTTGGAACCAACCCAAATGTCCA  
 ACAATGATAGACTGGATTAAGAAAATGTGTCACATATACACCATGGAATACTATGCAGCCATAAAAAATGATGA  
 GTTCATGTCCTTTGTAGGGACATGGA  
 TGAATCTGGAAACCATCATTCTCAGCAAACTATCACAAGGACAAAAAACCAACACCCACATGTTCTCACTCATAG  
 GTGGGAACTGAACAGTGAGAGCAGT  
 TGGATACAGGAGGGGGGAACATCACACACTGGGGCCTGTGGTGGGGTCGGGGGAGGGGGGAGGGGATAGCATT  
 CAGAGATACACCTAATGTAAATGGCGAGT  
 TAATGGGTGCAGCACACCAATGTGGCACATATGTATATATGTAACAAACCTGCACGTTGTGCACATGTACCCTAG  
 AACTTAAAGTATAATAAAAAAAAAA  
 AAAGAAAGAAAAATTTCCAGAGATACAGAGCCAGTTGAACATGAGCACATGTCTGCTTCTTCCCTCAAAAGTCA  
 TTAACACACCAGGAAACCATACCA  
 CCATGAATAAGAGTTAGAAAAAATAAACTACAGAATCACACTCTGAAAGACTATAGATATTTGAATTATCAGATA  
 GAGAATATAAAATATATTAAGAC  
 AGAATTAAAGTCTGGTAAAGGAACAAAAGACTACCAGAATTGGCCAGAACATTTTGAAAAAAATCTGGAAATA  
 AAAATTTAAGTTAAATTATAAAAC  
 AATGACCAGGTAAAGACAGCAAAATAGGTATACTTGAATAGGAACAGCTCCAGTCTACAGCTCCCAGCGTGAG  
 CAATGCAGAAGATGGGTGATTTCTGC

ATTTCCAAGTGAAGGTTTCATCTCACTGGGGAGTGCCAGACAGTGGGTGCAGGACAGTGGGTGCAGCG  
 CACCGTGCATGAGCTGAAGCAGGGTGA  
 GGCATTGCCTCACCCAGGAAGTGCAAGGGGTCAGGTAGTTCCCTTTCCTAGTCAAAGAAAGGGGTGACAGACG  
 GCACCTGGAAAATCGGGTCACTACCAC  
 CCTAATACTGTGCTCTTCCAACGGGGCTTAACAAACAGCACACGAGGAGATTATATCCAGCACCTGACTTGGAGG  
 GTCCTACACCCACAGAGCCTCGCTCA  
 TTGCTAGCACAGCAGTCTGAGATCAAAGTCAAGGCGGCAGCGAGGCTGGGGGAGGGGCACCTGCCATTGCCC  
 AGGCTTGAGCAGGTAAACAAAGCAGCT  
 GGGAAAGCTTGAAGTGGGTGGAGCCCACCACAGCTCAAGGAGGCCTTCCTGCCTCTGTAGGCTCCACCTCTAGGG  
 GCAGGGCACAGACAAACAAAAGACAG  
 CAATAACCTCTGCAGACTTAAATGTCCCTGTCTGACAGCTTTGAAGAGAGTAGTGGTTCTCCCAGCATGCAGCTT  
 GAGATCTGAGAACGGGCAGGCTGCC  
 TCCTCAAGTGGGTCCCTAACCCCCGAGCAGCCTAACTGGGAGGCACCCCCCAGTAGGGGCGGACTGACACCTCA  
 CACAGCTGGGTACTCCTCTGAGACAA  
 AACTTCCAGAGGAACGATCAGGCAGCAGCATTTCAGTTACCAATATCCGCTGTTCTGCAGCCACCGCTGCTGA  
 TACCCAGGCAAACAGGGTCTGGAGT  
 GGACCTCCAGTAACTCCAACAGACCTGAAGCTGAGGGTCTGACTGTTAGAAGGAAAATAACAAACAGAAA  
 GGACATCCACACCAAAAACCCATCTGT  
 ACGTCACCATCATCAAAGACCAAAGATAGATAAAACCACAAAGATGGGGAAAAAACAGAGCAGAAAAACCAGA  
 AACTCTAAAAATCAGAGCGCCTCTCCT  
 CCTCCAAAGGAACGCAGCTCCTCACCCGCAACAGAACAAAGCTGGACGGAGAATGACTTTGACGAGTTGAGAG  
 AAGAAGGCTTCAGAAGATCAAAGTACT  
 CCGAGCTAAAGGAGGAAGTTCGAACCAATGGCAAAAAAGTTAAAAACTTTGAAAAAAAATTAGATGAATGGA  
 TAACTAGAATAACCAATGCAGAGAAGT  
 CCTTAAAGGACCTGATGTAGCTGAAAACCATGGCACGAGAACTATGTGACGAATGCACAAGCCTCAGTAACCAA  
 TGCGATCAACTGGAAGAAAGGGTATC  
 AGCGATGGAAGACGAAGTGAATGAAATGAAGCATGAAGAGAAGTTTAGAGAAAAAAGAATAAAAAGAAACGA  
 ACAAAGCCTCCAAGAAATATGGGACTAT  
 GTGAAAAGACCAAATCTACGTCTAATTGGTGCGCCTGAAAGTGATGGGGAGAATGGAACCAAGTTGGAAAACA  
 CTCTGCAGGATATTATCCAGGAGAACT  
 TCCCCAATCTAGCAAGGCAGGCCAACATTCAAATTCAGGAAATACAGAGAACGCCACAAAGATACTCCTCGAGA  
 AGAGCAACTCCAAGCCACATAATTGT  
 CAGATTCACCAAAGTTGAAATGAAGGAAAAAGTGTTAAGGGCAGCCAGAGAGAAAGGTTGGGTTACCCACAAA  
 GGGAAAGTCCATCAGACTAACAGCTGAT  
 CTCTAGGCAGAACTCTGCAAGCCAGAAGAGAGTGGGGGCCAATATTCAACATTCCTAAAGAAAAGAATTTTCA  
 ACCCAGAATTTTCATATCCAGCCAAAC  
 TAAGCTTCATAAGTGAAGGAGAAATAAAATACTTCACAGACAAGCAAATGCTGAGAGATTTTGTACCACCAGG  
 CCTGCCCTACAAGAGCTCCGGAAGGA  
 AGCACTAAACATGGAAAGGATTTATAATCCCAAAGGATTATAAATCATGCTGCTATAAAGACACATGCACACGTA  
 TGTTTATTGCGGCACTATTACAAT  
 AGCAAAGATTTGGAACCAACCCAAATGTCCAACAATGATAGACTGGATTAAGAAAATGTGACACATATACACCA  
 TGGAATACTATGCAGCCATAAAAAAG  
 ATGAATTCATGTCCTTTGTAGGGACATGGATGAACTGGAACTATCATTCTCAGCAAATATTGCAAGGATAAA  
 AAACCAAACACCGCATGTTCTCACT

CATAGGTGGGAATTGAACAATGAGAACACATGGACACAGAAAGGGGAACATCACACACTGGAGACTGTTGTGG  
 TGTGGGGGGGAGGGGGGAGGGATAGCAT  
 TAGGAGATAGACCTAATGCTAAATGACGAGTTAATGGGTGCAGCACACCAACATGGCACATGTATACATATGTA  
 ACAACCTGCACGTTGTGCACATGTA  
 CCCTAAACCTAAAGTATAATAATAATAACAAAACAAAAAATAGATATACTTGAAGAAAGAAATAGTGAA  
 TGGGACGATGGATCTGAAGAAATTAT  
 ATAGAATGTAACACAGGGAGACAGAGAGATAGAAAATGTGAGAGATATTGGACAAGAGGTTATGCCACCATAA  
 CTAAGGAATAAAGTAACATTCCCAGTC  
 ATGAGACATATTGACATCTGCAGCTCCTGATGGGATGCACTGAGAAGGGCACAGTATCACTTCTATGATGTTCTT  
 GTCCCAAAAAATGTATAACCTGAAT  
 TGTCTTTAATTTTATTTTATTTTATTTTATTTTAGAGACAGGATCTTGCTTTGTTTTCCAGGCTGGAGTGCAGCAGCA  
 TGATCATGACTCACTGCAGTCTT  
 GAACTCCTAGGCTCAAGCCTCCTGAGTAGTTAGTACTCTAGATACACACCACCGTACCTGGCTCATTTAAAAAAG  
 TTTTTTGAAGTGATGACATCTTAT  
 TATGTTGCTCAGGCTGGTCTCGAATTTCTGGGCTCAAGTGATGCTCCTGCCTCAGCCTCCTTAAGTGTTGGGATTG  
 CAGATGTCAGCCAGCATGCCCAGC  
 CTTACTCTGGAATCTTAAAAATATTGCTAACTGAGCCCTACCCAAGACCTATGAAGTCTGATAGGGTAGGTCAAA  
 GGCATTTGTTTGTAACCAAACATT  
 TTTCCAGGAGATTCTGAAGTATGTCTAGGGTTGAAAACCTGTTGTGTGTAGTTTTGTATTTCTTTCTTCCTATACAT  
 TGTTTTGTATAGTTTTGTATTTT  
 TTTCTTCCTATACATTAGCCACTATGGAAATTGTAAAAGCACTTATTCATTAGATTATAAGATAATGTACGTAAGG  
 GAAGTTTTTATTAAAATGTCTCTA  
 GACAGACACCTCTGTATAGATAGGTTTGTCTCATATACTTGAGGACATTTATGGACCTCTGTTTTGGACAGAGTA  
 GGGGCCACAGCATAATCTGTGAAGG  
 GAACTTCTTCCTGGCACATAGGATTGCAGATTATAGATTCCTTTGTTATGTTTTCCCCTTTCTATTTTAAAGGTCAC  
 TATAGTCATCCTCCTTTGCTGCA  
 AATCATAAATGAACATACTAAGGAGGTTTTAAATATGTTTTAGTTAGGTCTTAAAAAAATTAACAAATGTTTGGA  
 ATTGACAGTGCTTTTACCCTGTTTT  
 TGTTCTGGGTTAATGCTGCGCCCTCACTAGGTCTGGAATCTTTATAGTTTGAGACATCAGACTTGGATGAGGAAG  
 GGTAAATCATGCTACTAGTGGGCCA  
 GTTACTCTGTGGTTTTCTTATGAGACTGTGCTTTAAGGAAAAGKTAACCACGGAGCATATGGCATTGAGTTTGAG  
 TCATGTGGAACCTGGTTAGAAGTAAA  
 ATGTTTCTATTCCAGCATTCTCTGAGTTTGTGAGCTTTTCATTATTTTATAGGAAGAATTCTAGCTTCTTATCCTGAR  
 ATGAGCTTGTCTCCAAATCTGG  
 TTCCAATTTTTTGTGTTGTTGTTGAGCACATTTGCCCATATGATATCACTATTGGCTGTAGGTTAAAAGTTACCGT  
 GTAACCTTAAACGTCAGAATTCAA  
 ACCACCAAGTATACACGGTTTTGTTGAGCCAGATTTTTAAATGAAATCTAATTTGAGTTTTTGGCTAAAAGTACTT  
 TCCACATTTATCAAAGTGACACAT  
 GAACTGAAGAAATACAATGTATCTGAGTACCAGGTGTTAYTGAATCTGGAATTGCAGGAGGGTAACAGAGCCAA  
 AGGCAGTCCTTTCTCAAATAGTCTT  
 AACGTATTCATSTTTGTCAAAAATGTGTCTACACAAAGTAATTGGTTCTTTTTCTCTGTAGATTTTCATCCTTTGTT  
 GAAAAAGATGTCTTAGAACCTAT  
 CAAGTGCAGAGCTGGGTTTTGAGGGCATGGAAAATTCTGTTGATGCTTTTCTCTAGTTTTGGCTTTCATGTGAGT  
 TGTACATAGGGAGGCTTATAAGGAG

CATCCTAGTGGCTTTACCTTGTATCACTTGGCTACTTTTTCCCCACACTTCATTGTTGTAATGAAGGGAAAGGCAG  
 CTGAAACCAGAAACCTACATCCCT  
 TGCCAACTTAAGAGATAAAAGGAAATTTATCTCTTCTGCTTTCTACCTAGTGAGAAGGGGTCGGTGAGAACTAG  
 GCCAGGTGAGAAAACCTACTATCAGA  
 GTTTAAGTTGAGTACTTCGATATGTTCTTCAGACCATAACATCTTAGTTAACATAAAACAAGGACTAGTGGATAGTT  
 TGTGTGAAATAGCATTTCACCA  
 CCATTTTCAGTACTTAATATAGCTGGGTGGAATCATGCTGTGTAAGTGCTGCAGAACATAAATAAAGGAAAT  
 AAACATCTTAATTTTGTAACAAGTG  
 AAGGTCTAGGCTAATATTAAGGGTAGTATAGGATACTTTTGAAGCTTTTGAATGTACAGAAGTACATATGTAATT  
 ATTATATATTGCTTAGACTTCTAGG  
 CAGTTTGCCATTCATGTCTTTTTGTCTCCAGATTGAGCTGTTTTGTCACATTAGTGAGCATTGAAACATAATTTT  
 AAGTGTCACTTCTACAAATGGTTC  
 TTAATTGATTTATGACTAATGTCAGTATGAACTGCAGGTGTGTGCCGTGCCTTTGGACTTTTTGGTGTCTGTGTGT  
 TCTGAACATAAATTGGCCTTAGTA  
 TGTTATGCACACTGTCAGTTTATCACAATCCATGGCCTCTGAGGAGCCAGAAGTACGGAATGTACAGACGATCTA  
 TGTACCTTATATTATTTCTGTCTC  
 AGCCCTTTCCCAGAATTCCCAATAAAGCAGATTTCTCAATTACTTGCTTTTATTTTCCTTTCTCCCATTTTTTTTTT  
 TTGGTAAATTAACAATGAGAG  
 ATAAATGGAAATAAGGCAGTTGGAAGAAAAGAGGATGAAAACAGGTGAAGATTCTTAGTTGTGGAATAAGCTA  
 GGTGGCTATGGGTGATACCATGTGT  
 GCTTGGGTAGTAAAGGTAGCTAGTTTTCTGCGTGTTGTATTAAGCCATTATTTATCAAATACTAATTACGTTTT  
 ATGCACTAAATACTAGTCCAAGTA  
 TTCTATTTGATATATGGATTAATTCATTTGTTCTCCTAGTGCCATGAGGCAGGTATTATTATTCCTGATTTACAGA  
 TGAGGAAACCGAAGACACAGCTA  
 ATATGTGGCAGAGATCAGAATCACAATCTGCCTCTGGAGCCTTGGCTTAATCACTGACAGAGCAACAGCCACTCA  
 GGCAGAGAAAAAGGAGTTGGCATT  
 ACAGTTAGAGGAGTTAGGACTYTAGTAACTCTGCCC GTTCCCTAAAGCTTAGGATTGGCTGTGATAGGGGAAA  
 GTAGCTTATGAATTGACTGGATTGAT  
 GTAAGTATAACAGAAAAGAGAGATTGTACTCAAATTTACAGCCATGTTTTTTGACATTTAGAATAATCAGAGGAT  
 ACTTTTCTGTAGTGAGAAATATACA  
 AAGAGTTGTTACACACACACCTTTTATATTAAACACTGACTCAGTGATGTCAGCAATCTAATTACTATAGAGTAT  
 CAGTTATCTAATACTGTGGAACAA  
 ACCACCCCAAAACCTAAACGCTTAAGACAACAGCAGTTATTATTTGGTAAAGAAATCTGCAAATTGGGCAGAGT  
 TCAGGACAGACAGTTCTTCTTTGCC  
 TCCGTGGARTGTTAGCTGGGGTGGCTTAGGTTGGAGGACCCACTTTCAGATTGGCTCATTACATGGCCAGTAA  
 GTTGGTGACGGCTTTCAGCTGGGACC  
 TGAGCCTGGGATGTGGGCCAGGGTTCTTGGTTTTCTTCATTTGGACCTCTCCACAGCTGCATGGGTTTCCTCACA  
 GTTTAGTGKCTGGGTTCCAAGAAT  
 AAGCATCATCTATTAGAATAGTGCTTGCTTTATAATTGCTTATTGAATGAATGAAAGAATCTCAGGTTGTTTTCTT  
 GCTCATGTTGGCCCATTTGGCAA  
 TAGCCACTGTAAAGTGCAGGTTTCTTCCCAGAGCACGGTAATCCCTGGCTCTGAGAACTGCTATTTAGAGGGTTC  
 TTTATCTTGACTTGAAGGTCTATT  
 TCTTCATTTAACCACATTTATGTTTCTAATAAATTACCCCCAGTTACTACTAGGCTGCAAAAATCATCTCTGAACCA  
 TTCAATAAACAGAGAAAAGAAAA

CCTTTTAAAATTTGGAACTACCTTGAAACTCCCCATTTTTCTCCTTTTAGTCTCAGTTGCATTGAAAGTTAA  
 CATAATGTTCTGTTCTGGATCT  
 GGAAATGGTGTAATAGTGACTTCACCACCTGCCATTCTTGAGTTGTGGGTTAGTTTCCCTAAAATGTAGATCTTG  
 ACAAGAGGCAGATGGTGCATGAAGT  
 TTATTTATTTGTTTATTTTTTGAAAAAGGAATACCTCTTATGAACCAGGAACTAGAACACCCCTTAGCTTACATTCT  
 AACAAGAAGAGAGACAATAAACA  
 TAAGTAAAATGAGAAAATTATAAATTCTTTGGACAAAGGGATAAAGTATAGTAGGGTAAATGGTTTTACATGTT  
 GAAGGGTAGTTGTTGGGACTGTAGTT  
 TTATTTACTTATTTTAATTTTTTTTGAGGCAAGGTCTTGCCCTGTTCCCCAGGCTGAAGTACAGCAATGCAATCAC  
 GGCTCACTCAGCCTTGACCTCCT  
 GAGCTCAAGCGATTGTTCCACCTCAGCCTCCTGAGGAGCTGGGACCACAGGCACACTCCACCACACCCAGCTAAT  
 TATTGTATTATGATTATTATTATCG  
 TAGAGATGAGGTTTCACTGTGTTGCCAGGCTGGTCTCAAAGTCTGGGCTGAAGCAGTCTGCCACCTTGGCTT  
 CACAAAATTCCGGGATTACAGGTGT  
 GAGCCACTGTTCCCTGCCAGGACTGTAGTTTTAAATGGTGATCACTGTAAGGCAGGTGTTTGAGCCAAGGCTTA  
 AAGGAGAGGAGAGGGTGTAGCCATATG  
 GATATCTGAGGAAGAACAGCATTCTTGGCAGAGGAAACAGCCATTGCAGGGATCCTGGGGTGGAAATGTGTCT  
 TGCATGTTCAAGGAAAAACAAGGTGGC  
 TTTGCTAAAGAGAGAGTAGTAGGAGATGATGTCAGAGAGGTGAGATACAGGGAGAGGTCAAWTGAGGTAAG  
 ACTTCTGTCGTTTTAAAGATTTGGGCTTT  
 TACCATGATAGAAATGGGAAGCCATCAGAAGGTTTTGAGCAGAGGTGACATAATAATCTGACCTAGATTTTAAA  
 AGGATCACTTTAGTTACTATGTTGCC  
 ACTAGCCTGAACCTGGTAGGCAGGAAGACAAGTGTCTGCTCTGTGTCATCTTTGGGCTTTATAAACCATGTTCC  
 ATCTGTTGTCTCATTTGAATCTCAG  
 GAAACCTTGTGAAATATTTGTTATTTTTTCATGTTTTACATGTGAGGAATTGGAGGCTTAGAGAAGTTGACCAAG  
 TATCCAGACATACTCAAGTGAGCA  
 GTCTCAGAGTAACATCTTGAACCCAGATCTCCTGTCTCCAAATTTTGCTCATTCCACTATTTACATTAAACCT  
 GAAGATCTCCCTGCTCTTTTCAT  
 CTTAACRTAATTTAGCTTTTTGTGTGTGGCTCTCCAAGYCACAGAGCAGCAGCAGCTTTTGAGGATAAGAAGCA  
 GTTTCTTTTAAGGATAAGAAGCAGT  
 ATGGTTCTTACCTGGTAAGATGGCTTGAATTCAAATTGACCTACTTGTACCATAAGAAAGCAGGTTTGCAAGGA  
 GTAGAATAGACTTGGAGATATGCCA  
 TTTTTTGTTGAGCTACTTGGTATGAAGACTGAGTATCTGAAACCACACTGTAAGGATTAAGTGGCTTCATAAATAA  
 TCTAAAACAGTTGTTGGCTTTTAG  
 AATTTAATAAATGAAAAAGTAAGGTAACAATTCTTATAATACTTTTTCCAGACTTTCGGGGAAATTGAAATCATTT  
 GTGTTTTATTAATCCATGCTTTTC  
 AGGACCAATGGCCATAATGTGAGCAAGGTTTATTGGTGTAATGATTATTTTGTGGCCTAAACTACTGTGACTC  
 TTTGGACTAGCATGTTTCTGATCAT  
 AGAGACAGCTAGAGACTTCTTCTTTTTCTTAATATTTGTAGCTTGTTTTAAACGTACTGAGTAGAACCTGATTTA  
 AGCAATTATACAAAGTTTGGTTAG  
 ATTGCTTTATTAATATTATGAACTAGAAACAGTCACCAAATGCGCATTCACTATACTAATCCTTACTATGATTGAGT  
 CATTTCTTTGAGTTCTTGAAAT  
 CTAATTTTTCTAAATTAACAAAAGGGGACATTTGAATGATGCAAAAGCACTCATGGCAAGAATCTAAAAGTTCA  
 TATAAATAATAATTTCTCATACTC

TTTTATACATTTGACCATTACAAGATCTGTCAAATTCCTTTCAATATTTGTGATGTCACAGTCGTACTAAAACAATC  
 TTGAATTCTGGCAGCGAAATGAA  
 GTTCTTACCTGTCTTTCAGATGTTAATAAAAAATAAGGGAAATACAAACCTGTAGTTATTCAGATGGTTAAAATTAG  
 TTTCTGTGGGTCATTTCATAGAGT  
 TCAAGTATTTAAATGTTTTTTTAGCTGAATTGGACAATCAACTTGTCTGTAATTTAGCTTAAGAGAAATTCCATTTT  
 AAGACAAAAAAACTACAGGGAAT  
 GTAGGCCAAAAAATAATAACAACCAAAAAACAAACAAGTGAAAAACGTTAAAAGATTCTTAGCAGAATTTATTAA  
 TATATTACCTCTTCAAAATATATTCT  
 AGCCAATGGATAGATAATAAAATTGAGACCCACAGACTATTTGGATCCTAAATTTTGTTCATCATGACTCCCTTG  
 TTGAAAATACTTTATTGACTCTGT  
 GTAAYGTAAATAAGATAATTTTGTGAAGCTCATTGAATGCGTCTGAGCTCCTTAACTCAAAAAAGGCTAAAAGT  
 TACTGAAATAGCTATGGTTAAGGGT  
 ATATTGTCATYACAATGTGATAATCATTACCAAAGTGGGCACTKGGAAAAATAGTGACCACTTATGAGCCAAACC  
 AATCAACATTA AAACTGATTTCCAG  
 TAAGTACAGCTGCCAAGTGATTATAGAGCAGCCTGGATACAGATGGGATAGACTACAAGGATAGTTAAGCTCAA  
 AGCATTTGCCTGTTAACAGCCATGAA  
 TTGTGAGGTGAAATGCCAGTTTACATAGTGTTAGTAAAATCATTAAAAGAACAATGATTAGTAGAAGAATTCTTT  
 ACAAACCTTAAAGTATCAATAGTTC  
 TGAATTAATTCAATTTATAAATTATCCATAAGTCATACAAATAGCCACTTTGTACTTTTTCATGTCCCTCCAAACTC  
 CCCCAGTCCCTATTTCCAAACTG  
 TCCACTTCTGTTGAATCTATTGTAAAAATGTGTACCTTTTAGGGAACTGAGATAGGAAAGATGTGAAAAACAGGA  
 TTAGTTTAGAATAATGGTTCTCAA  
 CATTGGTAAAGTCACCTAGACTAGTACAACCACTGCGGAAAACAGTGTGGAGATTCCTTAAAGAACTAAAAGTA  
 GAACTACCATTTGATCCAGCAATTCC  
 ACTACTGGGTTTCTACCCAGAGGAAAAGAGGTCGTAAACAAAAAAGATATTTGCTCATGCATATTTATAGCAGC  
 AAAATTTGCAATTGCAAAAATGTGG  
 AACCAACCCAAATGCCCATCAATCAACAAGTGGATAAAGAACTGTGGTATATATATATGATGGAATACTACTCA  
 GTCATAAAAAGGAATGAATTAATAG  
 CATTCGCAGCAACCTGGATGGGATTGGAGACTATTCTAAGTGAAGTAACTCAGGAATGGAAAACCAAACAGCGT  
 TATATTCTCACTCATAAGTGGGAGCT  
 AAGCTTGGGATATAGGTGCACCAAAATATCACAGATCACCACTAAACAACGTACTCATCTCACCAAATGCCACCT  
 GTTCCCCAAAAACCTATGGAAATAA  
 AAAATTTTAAAAAAGATAATTATTAATCCAGTTTTTGGTGAATCAGTGAGTGACAGTGGTTATAGTAGTGGTGG  
 TTTAAATCAAGGAATAAATGTTTTTC  
 AAAGTGAAAGGAGCACATCCTACCACACAGTTCAAAAACAATCACAAATATGGCAGGCTTGCTGAGAGCTTTCA  
 TACATCATTTATTGTCATGCACTTGT  
 ATGATTATTGTATACTGTACACATTTTTATTTTACAATGATTTGTATTTTCCAATGCRTTTATTGCAGTTCAGTGTG  
 GTAGATGGCTGGAGCCTATCCTG  
 GTAGCTCGGCGCACAAGTCAAGAATTAACCCTGGACAGGGTGCCATCCCTTTGAAGGGTGCACTCACGCATGTT  
 CYGCACACTCATTGAGACTGGAGCCA  
 TTTAGACGTGCCAGTTAACTTAACATGCACATCTTTGAGATGTGGGAGGAACTGGAGTACCCACAGGAACTC  
 ATGCAGACTTGGGGAGAGTATGCAAA  
 CTCCACACAGACAGTGACCTTGGCTGGGAATTGACTTTTTTTTTGTCATCAGTGTTATCACAAATGACATTGCAT  
 GAACCAGCATTGTTTGAAGACTTG

CTGTATTCTGATTTTTTTTATCTTTCATCTTTGAGTCATTAGCACCAAATATAGTTCTTGGAACATAATAGGTATTTA  
 CTACATGAATAAATTAATGGAT  
 GAATCATTCACTCTGGGAAATGTTGTCAACATAGTGCAGTGATGGAATTCCAATCCTGAATCCAAATCCTGACTTT  
 CTTAGTTATTTAGAGCATAAACTT  
 GGGCAAACCTGTTTAAACTCTTTGAGCTTCAACTTCCTTTGAGCCTTCATTTCTTAGTACATAGTAGGTACTCAA  
 AGTGGTAGCTACTGTAAAATATA  
 CTGTATTTGCATCTTAGAAGAGCATTAAACATTCTTGTGATATCTTTATGGTCAAGATGGGAAATTTTGAYCTAGG  
 TGGTATTTTCAAGTTAATTCATAAGT  
 AGATAAACCTACGTTTTTACAGATTTAGAAGGAGCTTTTTATTTTAATAACAACCTCAGTACAGGATATTGATGGCA  
 TGCTGATACTTTCTTGGAAGAATT  
 TCTTTTGCTAATGAGGAGCCTAAAAGTTCTTGATTGATTAGTAGGTAAGTGTTTTTAAAGGAATTGATATAAATTA  
 TTCAGCTGGATCCTAAAACCTCAGT  
 TACACAAATTCAGAGTTGAGGTGGGATAGGAAGAGAAGCAAAAGTAATGGTTCAAAATCGATTTGTGTGAAAG  
 GACTTAGGGTTTTAATTACACTGGGCT  
 CAATATAAATAGAGTGTAATATAACTGCCAGGAAAGCCTATGCGATATTAGACTGCATCAGTAGAACAACAGTG  
 TCTAGAACAAATGGAGGAATTATCCT  
 TTTCCATTCTGTAGTGGTCATACTACCTAAGAAGTACTGTGGCCAATTCTGGACATTACCTCTGAGAAGCATGTTK  
 ATAGGACAGTGATCAAATATGCTA  
 AAGGAACTTCAGACTATGCTGTAAAGAGGAGCATGTGAAAGAATCAAGAACTATAATCTTGAAAAAGAAAGC  
 CTGGGGTAAGGGGGATGTTAGATGTCT  
 TTGGACATTAGAAGAAGGCTTACTTAGACTTGTTCTGGATAGCCACAGAATCCAGAACTAGTATCAAAAGGTGG  
 AAGTTTTTTTGTTTTTTGTTTTTAA  
 AAAGGAAGCACATTTCTAGTTAAAACAAGGGGAACTCTTAATAATTAGGGTGGTAGAGATACTGGAGGGATC  
 AGGGAGTGAGTACTTCCACCAAACAGG  
 AACATTAGTATATGGATATTTCACTAATAACAATAGCGGGGTCAGTAGGGGTGAATTAATTGAGATTCTCTAAA  
 CTTAGGGGAAAGTTCTTAACCTTTGA  
 TTTTATTCATAGCTTACTTTCATAATTAGAATTTTTGGCACCATGCTTGAAAGAAAAAGTAAACACTTTTTTTTTAA  
 AGTTAGTTTCCTAAAGAAGAAAT  
 AAAATTTCTTAGTTGTCATCACACTCTTCATTTGACTTCCCTCCCAAATTGCTATTGTTTTCTGCTTCTCTCATGTTTT  
 CTKAAATACTCATCAGGGTGA  
 CTTTTCTTAAGTTTAGACAGACATTCCACTAGGTCTCTTACCTCATTGTGGCAAGGTTTATATTGATGTGTGACTG  
 CTCAGGTAAACCTTTAGTCTGAA  
 TGATACAGCTATGACAAAATATAGCCTCAGTTGTAGCAAAATTAAATGTTACACAATGGCTTTAATTGGTGGTAC  
 ACCTTTGAAGAGTTTGCAGTATGCC  
 CAAGTGCTAAAGCATACCATGTGAGAATCACTGGGTAGCAATGATTTCTGGGGGTGAAGAATTGAGGAGAATG  
 GAAAATATATTAAGGGATCAGAAAAAT  
 ATTCTAGAGAGGTGTGTCTGCTCCCGTTTTCCCCAACACTGATTTCCCCAAATTTGGAGACAACTACTATTAT  
 CTCAAAGAACCGCTTTTCCTTTAG  
 AACCACAGGCTCCTTAACCTTTTTTCATTAGCCTTACTTTCCTCTTGCAATTTATCCCTCCAAATAGAATCTTTT  
 GTGCTTCTCTAGGGAGATTCCC  
 TTTGTGTTGAATTTTATGAGTAGTGTAATTGTGTGGGTGTCATCATAAAACAAAGTGCRAACACACACACAAAG  
 TCTTATTGATGAGCCAAGTTCCTAA  
 GCACTATTAATGAGTTAAAAGTGTTAATTAAGGCTGGGCGCAGTGGCTTACATCTGTAATCCCAGCACTTTGGGA  
 GGCCAAGGTGGGAGGATAGCTTTAG

CTCAGGAGTTTGAGACTAGTCTGGGCAACATAGTGAGACCTCATTTCTACTAAAATTAATAAAAAATAAAAAATTAG  
 CCAGATGTGGTGGCACATGCCTATA  
 GTCTCAGCTACTCAGAGGGGCTGAGGTGGGAGGATTGCTTGAGTCCAGGAGGTCAAGGCTTCAGTGAGCTGTG  
 ATTGTGCCACTGCACTCCAGCCTGGGT  
 GACAGAGTGAGACCCTGTCTCAAAAAAAAAAAAAAAAAAAGTTAAGTAGGCAGTATTTGCTTATATGTATATA  
 TTTTGTTGTTTTCTCCAAATGCTTAT  
 GCCTCTGAAGAGGATATACATTGATGAGTGAAGATATTAGGGAGGGTAGGGAGAAAGAATAGAAGTTTTTACA  
 TATAAATACAAATAGAAGTATTTAAAA  
 AAWTTTTTTTTCTTAGCTTTATCTGATAGCCTGCCTTCTGGAATTGTGAAAGGCTAACAATGAATTTGACTTTTATT  
 CCTACTTCATTTGAAGTATTCTC  
 TCATTGAAAACACCAGTTTGCACCATCCAGGAGGCCTTCTTATTTTCACTTATTATTGGAGTTTAAAATTTGTTTTTA  
 GGTGTGTCGTATTTATTTCTACT  
 TCATATATACATGTAATTATCAAGTTTTTAAARCATATAGAACACTACTATTGTTTAAACAGTTTATAAATAAGGTA  
 CTTTTATACTTAGAAAAATTGTT  
 TTTGATRTGTGCAGGTTACCATGCTAAGGATTTGAACACCTTCAGTACCCTATTTGTACTTGCKTCTATGCCTAGC  
 ATTTTTGTTACGTGCATGGATCTG  
 CAAAGGAATTGTGGTTTGTTCACAGCTGCATGTAGCAGGCATATGTGTGACATATTTTTATTTTCACTATTATTT  
 CCTGTTTTTTATTTTCACTGTT  
 GGAACAGGATTCCAGGAGAATGAACCCTAGTGTGAGCCCTGCAAATCAAAGGAGGCACCTTCTAGGATTTTCTG  
 CTAAGAGGCACAGGACACATCAGAT  
 GGTATTATCCAGAAAAACAGCTACAGGTAAGATGGCATGCATCTATCTTTTGGCCTAAGTGATTTAATGAAGCTG  
 AAACAAGAGAATTATCTGTAGTATA  
 ACATTGCATCTAAATTGTTCTGTATATTTTGAGTAGTAGAACACATAAATTACTGAAACCATTAATACATTTTACT  
 TTTAATATTGTAAAATTAAATTT  
 ACTTTGAAACCTTTTGTAGTTTTTAAATTCCTATAGAGATGTAATCTTAAATCTCAGTTTGTGTGCATTAATATAC  
 ATATTGAAATATATACATATTAT  
 ATATAATGTCTTTAATAATTATTTTATCTATCATGTGTTTATACATATAATACATCTAATATTTGTTACTGTCTCACC  
 TGTGGATGGATATGATATGGTG  
 GAATAACATGATTGAACCTATTTGCAGTCATTAGGTAAAATCATTTATACTCATTTATTGTGATCTTTGTTGAATCC  
 CTCCCTTGTTGCAATTTAATAAT  
 GACTTGATAAATTTCCATATTGCTTGTCAATAGCAAAAGGCTTTAATTGTATAAAAATGCAATTTCTATATTATGTT  
 AACTAGAAGAAGGTATAGTGGAC  
 ATTTTTATAGTTGTATACTTTTTGATAAACATTTCCCTGTGTTATTTAGAATTTGTATTAGAGCCAAATGGAGCCAG  
 GTCCTTGACTAACAAGGCAAAGG  
 GAGGTCACAGTATGGTAACTATTGCACAGTATAATGCTTGAAAATTATTACAGATTTGTTTACTTTTATCTCTTTA  
 AGGATTTGTAGGTTGAGGGCAATC  
 TTTCTTCCTGATTTTTACTGTGTTACATCCATACTGTTGAAAGATGATCACCACAATCCTGCAGGCTTGTCTCCACT  
 TAGCCTTCCTGAAAAAAGTGTTA  
 TTTTTGTGGAGTACAGTATGCCAAGTCAGTGGAGGACCCAAAGGATTTTATGGCAGTAGCTGCAACGCAACATA  
 GCAATGATAGATGTCATAATGTTGGA  
 AAACATAGCATTTCTGCTTTTTTACATGCCAAAATTCATCATACTTCTTAAATATTACCTTGATGATCTCTTAAC  
 GGTGCCCCTGAATATTGGTAGA  
 GTTTAATAGTATATTTAATTTTCATAAGTAATTTGTGCAGGTTTGCCTACGTGTTCTATAATTTTGAATTCTGTTA  
 GTGTCCACATAAGATCCATCAAA

AATTAAATTTAAATCTAATAGCCTTTGGATTACGAGAGTTGAGATAGCGTCTTGCTAAAGAATAGAACAAGTTA  
 TGGATTGAATGCTTGAAATTACAGT  
 TCAGGTAAGCTTTACAAGAATATTATTTATTCTGCACTAACTGCAGCAAGTAGAAGCCTTCCACATTAGCATAAG  
 ATAACATGGTGGCTCCCAGAAAGCA  
 TAGTAGTCTGATCATACTAGTTCAAACCTAAAAGTCTGAATGACTTTCTGGCTGAAATTTATCAGCTTTCAGAGTAG  
 TACATGGGATGCAAGCTATTTTTT  
 CCCTCCCTTCTCCAAAGTCACCATTGAGTGAATAATCTGGGTTACTGGTCTTGGGGAAAAAAAAAAAAACCAAAAAA  
 CAACCCAGGAAAGATTATGGATTCT  
 TGATTCCATGCAAGTCCTTTGAAATATTTTTTCAGCAACTTGAGAGATTTAAATAAATGTCAGTCCATTAATTTGAT  
 GTTCTTGMTTTTATCAGTATAAAT  
 ACTGAATAAAAGATACCAAAAAATGTTTGTGAAACATTTGAAAGCACCTAATATTACTTTGTTCTGTGTTCTGAAC  
 TGAATTTTAGATGTTTTATCTATT  
 TTCAGCCGTGATACCTATACATTTTTTTTAATGCAGCTTACTTTATGTTTCTTTTTCTAAGGAGAGCTATGAGACAT  
 GGTTCAAATAAAAAATGAGTAGA  
 TTCTAATCCCATAGTTTATAATATTGTATTAGCCTATCATTTAGTAATTAAGTCTTATATTTTTTCAGTCATAGAAA  
 TGTGTGGCATATAAATAGTATG  
 ACCATATAATGTTCTATCAAACCTGGGACACTCTTAAGAAGGAAAGGTAGTGCTAGTAATAATTACATCAGAATAC  
 CAGTACTGCTGTCTTGGATAAATGA  
 GAAAGGATGGTCACCCTATATAAGATGTCACCCAGAACTGATCTTTACTCCAAGAATTCTCATTAACTCTAGCTCC  
 CTAGCACTGTCCTTTACTTTATTT  
 ACTTTAATGATTTATTTGTTGAATTTTTTCTCTTGGCATAAAATGTTTGCAAGCAGACTCGAACTCTGACTGTTCC  
 TGGGCTAATGTTGGTCTAATTGT  
 CTCCAGATTGTATGTTGGCTGATTTTCTCTCTCCTTTAACTCCACCCAATCCTTGTCTGCTGTTTCTCCTATCAGAC  
 TTTTTTTTTTTTTTCCAAAGA  
 ATATTTTTGGCTGATAGAACAAAGCCAGTTAGCACAGATGTGCAGAAATCATTACTAGAGTTCAGATCACAAATA  
 TATTTTCAAAGCAGTTTTGGACTT  
 TTCTCATGGACTACAGTGTTTGTATTTTTCTAAGTATGCAGGCATGATTTTTAAATAATGCGTATGTCATAGATT  
 ATAAGAAGAGGTGTTAAGCAGCC  
 TCCTCTTAGAGGCATTGAGATTTATCAAACACAAGACAGATACTTTTTTCTTCTCACTCAATGCTTCATTTTAGG  
 CTGATCCTTTTGTGTTTGCATTC  
 ATTACCTGTGCACCAACATGTCTGTTTTATTCCAAACAACCTAAACCTACTTAACCCTTACAGCCACAAGAGAGGA  
 CATTGAGAATATTCCTGGTGCAAG  
 TTGCAGGGAAGAATGTTCTATATATCCTCTGCTATTGTTTCTTTCTCTCGTGTGTGGCATAAATAGCTATCTCTAA  
 CTTTTAACTTGAAGTTAGCCTTC  
 CACTCTGCCAGAGTTATTTCTGTTAAGGGGCCAGTGCTTTCTGCCATCTTACTTGCCCTTTCCTCAGCATTGCTG  
 TGTGGGTCAATTTGCTCCTTGAAA  
 GACTTTCCTTTTGTGTTACTATGGTTTTTAGTTTCCTTTTTTTATTGGCTGCTGCTTCACAGTAATCTTTGTGGGCT  
 CTTTTTTACCCAACCTCTCAAG  
 GTTTGGAACGTGAACTTTTTTTATTTTCTTTCTGCCTAGGTGACTCATCTATAATTTTCATGCACAGTTCATATCTAT  
 ATGTCCTGATGACTCTCAAATT  
 TATGTTTTCAGTCCAGACTTCTCTTCTGAGTCTCCAGAGTCACATATCCAATAGTCTCTTTGACATCTCTGTTTTAA  
 CATTTCAAACTTAACATGTGCAC  
 TGTGGAACTCAGTTCCTCTCCCTGTTTTACTTCCAACCAATTTCTACAACCTGTGTTTTTCTTTCCATTGCTTAAG  
 CCAGAAATCCTGATTCCTGTCT

TTCAACAAGTGTTGTTGATACTCTTCCCCAAAATTGTCTTGAATTGGTCCACTCCTCTCCGTCTCTGCTGCCACAAA  
CCTAGTCCATGCCATCATCAGTT  
CTTGCCAGGATTGCCATAGCTCTTTCCTAATCAGTGTCTCAGCTTCCCCTTTTGCCTCCATTGAGTCCATTTTTGGT  
ACCATGGCAAGGTATGAATGTAA  
ATTGAAATATATCACTTCTTTATTTGAAACACTTTTATGACTTCTTATTACATTTAGTAATTCCTTATTATGACCTGG  
TCTTTTATTATGACCTTATTAT  
GATTCCTTTCCAGCCTCATCCTGTTCCCTCTCTACCTCCCCACTTGCTTGCAACTCTTCAGCCAGTGGTCTTTTTAAA  
GTTCTTTGAGCCACCTTTCCC  
CCACCTGCCTCTTTTATGCCTAGAATGCTCTTACCCCAGTTTCTATAGAGGCGGCGTCTTCTCATCATTTAGGCCTT  
GATGTAACCTACCATCTCCTTCTC  
GGACCATTTTATCTAAAGTATATACTACCATTGCTTTTCCATTTCAACCATTTGTTTGTTTCTTTCATAGCACTTACT  
AGAGTTCGGATTATTTTATGCA  
TTTATTTTCTTCTATTTGGTCCATCTTCCCCATTAAGGTCTTTGGGGACAGAACTTTGATTCTCTCATTCCCTACTG  
GATTCTCAGCACCTAGCACAAAT  
ACCTGGTGCACAGTTGGTATGCAGCAAGTATTTGTTGAATGAATAAATGAAAAGTGTGAGTCTCACTATGTTTCAG  
TTTTATATGTTTACATGTTTTCTTT  
TGAGATATTAGCATATCTAGTTTTCTAAATCCAGATCATCCTATATAAAGAAATTCAAGAGAGATGCAAGCTTTGT  
TAGTGAGAGAGATTAAAGATTGCAG  
ACTGAAGAGCATAAATAACAGTAATATCCCCATCTCAAATTCTAACTTTCAACTCCAGTTTTGCTCCCTACGTCAC  
ACCTTGGATATGAATTCACATGAT  
CTACATCAGTGCGATGGCTAGGAAAGTATATATTTTTGAAGCAAATGGAGTAACTAGTTAATGTCAGTCACTCTG  
TTGTTTTATTAGTGATTACATTCT  
TCCATGATCCCATAGTTTATATTCGGCTTTTCAACTCTTTCACCTTTCTCCTGCCAAGCATTTTATGTCTCACCTAT  
CCTTAACATACACACACACACA  
CACACACACACACACACCCCTACCTGTTTTACAGATACTATCCAGTTTTATCATAGAATTGTGTCTCACCTATCCT  
TAACACMCACACACACACACACA  
CACACACACACACACACACACCCCTACCTGTTTTACAGATACTATCCAATTTTATTATCATAGAATTGTGAATGAC  
CCCTAACTTTAGTCTGGTCACAT  
AATACCACTTGATTTTCGCCTCAATTACCCCGTTACTGTGTTACTACCTCTGTACCAGCAGCATTGTGCATGTCTTA  
AGCTCATTGAAGATATGATGAAC  
CATGTGTTATTGTATCGCTTATTCCAAATTAATTGCTCAATAAATGGTGATTAGACAAATGTGAATTACCATCCTG  
TCTCTGTAGAAAATCAGAAAATGA  
CATTTTCATGTAAAATAAGGTTTTGTATTTTATACTTTATCTTGCAGTATATTCTGTTTACTATATTTTTGTTTCCATG  
AAAAGTTAACTAATGATGAT  
TTTCTTATAGCAAACCTTCTTAGGAATTTAAGATACAAGGAAACAAATTTTGTGCATCCATGTTTCAGATATGATCA  
TTCTGTTTCAAATGACAAAGCCCT  
GATGGTGCTAACTGAAGAACCATTGCTTTACATTCCCTCCACCTCCTTGTCAGCCCCTAATTAAYACAACAGAGTCT  
CTCAGGTGAGTGTTGTAGATTATT  
GCAGAAACATCTGAGTTGGTTCCCATGTTTAGTTTTGGAACCTAGACAAGTAATTGGTCAAGTTCACTAAGTGAC  
TGAGAGAATCTCTGAATTATTTGTG  
GATAACTGTAGGCATATTAGTTGTCCAGTTTCTCCACCAACCTCATGGAGATGGTGATGGGTCTATTTTATGTGAT  
AGTTGCTTACTTAAATATGTCTTC  
TTTAAAACTACAGAATGAAATTAATATTAATAATTCATGGTGTTAGGCTTACAGTTGGTAACATTCATAGCAGA  
AAGTACCTTTTTCTTCCAATTAGGA

AGATAATAATTTGTGGGTATGTGATTCTGGAGATACTTTACCGTTAGTACACACTTCTTTATCAATTTTTCCCTTGA  
 ATTTTTTTTTCTGTGATTTTGG  
 ACCTTACCCTTAAAACCTCAAAAACCTCAGATTTTTGTGATATGTAGTAATCAACTACAGTTAAACATAGTTT  
 AACATCTTCACAGCTGCTGTAAA  
 AAGTAAGCAAAATAAGGGAAGTTATATATGAGCTATGCTTTTATTTCTGTGTGTTGAAAAATTTGAATTTTTTTTTTA  
 CCCCAGGCAATTGAGTTCATTCTT  
 TTTCTTCTCCCAAAGAGCATGATTTTCAAGAGGTTTTAATGCATCTATTTGTTGTTAGGGGTCCTATGACTTGTTA  
 AGTTCTCTACAATGTGTCTCAA  
 TGTTTGTGTTTATATAGCCCAAGAACAGAAATAGGAAATACTTTTTATTTTGTTTTGTTTTATTTTATTTTATTTTA  
 GACAGGGTCTTGCTCTGTTGCT  
 CAGGCTGGTGTGCAGTGGCTGGATCATAGCTCACCGCTTCCCCAACTCCCAGGCTCAAGTGATCCTCTTGCCTT  
 CGCCTCCTGAGTAGCTGGGACTACA  
 GGCACATGCCACCATGCCCAGTTAATTTTTTAAAAAAATTTTCTGTAGAGACAGGGTCTCACTGTGTTGCGCAGG  
 CTGATCTCAAACCTCCTGGGCTCAAG  
 CAGTCTTCCCACCCAGCCTCCCAAAGTGCTAGGATTACAGGTGTAAGCCACTGTGCCTGGCAAGGAAATACTTT  
 TTTGGTGGA AAAACCTTAGCATCC  
 AGTGTTTATATTTGGGAGGTATTTAGCTGTAGTTTTCTTTGAAGTTTGAATGCATGGTTGACAGGTTTACCATGTA  
 AAAATGCATGTCTTATTA AAAACA  
 ACGTTAAGTGTCAACTTTTTCAAATCAAAAAGCAATATTGAGTCATCAGGCACTTGGAATGCTCTTTGGAACCTAC  
 AAATGATTAGATTCATTTCCACCC  
 ACAGGGTGAAACATTCATTTCTATGTTTAATTAATTTCTATGTTTAATTAATTA AACTATATTTAWAGGTAAATC  
 ATGAACTTCGAGGATGGGTTTCAT  
 AGACATGAAGTAGAAAGGACCAAGTCAAGAAGAATGACAAATAATCAACAGAAAACCCGTATTCTTCAGGTATG  
 TTTCTGTTTGTCTTTGAACAATAGTT  
 GCGTATTTGTTGGAAGGCTTCTCCCAGCTGGGTTACTCTCTCCCTAGAAGGTTGGTAACTTGAGCTTTAGCCT  
 CTTCCCACTGTCTCCTGCTGCTTC  
 CGATGCAGCTATCATTGCCACTGAAGTCCTCCTCCCAGGTTTTATAGCACTAGTTTAGGTTTTATAGTATTAGTAG  
 TTCCCTGATCAGGCATCGTGCTAT  
 TTCTTGCTCTCACCTTCAAATTTGCTTTCAAGCCAGGAGTTCCTATGCATGGATATCTATATCAAATGTCTTTGTCA  
 CAATGGTATAATAATAATACTG  
 CTTATCTCCTCTTTTCTTAGAGCTTTTCAAAAAGAGAATGTTAAGACCTGGGGCAGTTTGGAAATGAAGACATCC  
 ATTTCTGAACTTAGACATTTACTAC  
 TATGGTTTAATGACAGGTCTTACTTAGAATCTGTTAAGTGGGGGTTCCATCTCAATGTTGTTTGCCACAAAACRCC  
 ATTAAGTCATTCCGTCATTCATCA  
 AATATTTGTTGTTAGAATTTAGACTAGCAACTGGCATCTGCAAGTTCCCAATAATGAGCCAAGGACTGTTCTTGAT  
 CCTGAAATAAAGTAATGGGTATGA  
 TAAAGTCCCTACCTTAATGGAGCTTACACTCTAGTGCAGAGCTAAACAATAAGCAAATAAATAAATACACAAATA  
 CACTGAAAATTAGAGCTAAGTTCTC  
 TAAAGAAACACAAAGCAGCATAGGAGGTTAGAATTTGAATTGAGAAATGAGTGACTAGACTGAGTGGTCAGGG  
 AAGGCTTCTCTGAAGAAGTTCCACTTT  
 AAGACAACATTGTTAAGGTAATTAATGCTAGCTGTTATAACAGCCCCAAAATCTCAGTAGCTTAATGGAATAAAA  
 TTTTTCACTTGCACAAAATTTGAC  
 TTCATTTCTAGGATTGCTTTTCTCCAAGCAGGGAGTAAGGGATCCAGGTTCTTCTATTGTGTAGCCCCAACAAC  
 TTGAAATTCTTGCTTCTTGCTTG

AGATGTTAGGGAAGAAAGAGTTTGGTGGATTGTGCAAAATGTTTAAAGAATCAGGCCTGGCTTTTAGCCAATGT  
 TTAGTTGGCTAAAACTCTGTCACTTT  
 GCCCAAACACTACTGTTAGAAAGACTGGGATCATGTATCTTTCTGTGTGCTCGAGAAGAAGACAAGGGTTTGGTGA  
 GAACATTGCTTTTGCCATAGGCAGAA  
 AAGTCACTGAAATAACAATGTAAAGATTTAGGGGAGGCTGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTT  
 TGGGAAGCTGAGGCTGGTGGATCACG  
 AGGTCAGGAGATCAAGACCATCCTGGCTAACACAGTGAAACCCCATCTCTACTAAAAAATACAAAAAGTTAGCC  
 AGGCGTGGTGGCGGGTGCCTGTAGTC  
 CCAGCTACTTGGGAGGCTGAGGCAGGAGAATGGCATGAACTTGGGAGACGGAGCTTGGAGTGAGCCGAGATC  
 ACGCAGTCCAGCCTGGGTGACAAAGCGA  
 GACTCCGTCTAAAAAAGAAAAAAAAAAAAAGATTTAGGGGAAAATCTGGAGAAAGAGCTTCCAGACAGAGGGGAAC  
 AGCAAATGTAAAGACCCTCAGGAAGGA  
 ACATGAAGAAGCTGAGTACGACTGGAATGCAGGGGCTGAGGGGAGAGTGGTAGATGAGATTGGAGAGGCAG  
 CAGATGTAGGCTCATAGCAAACCTTCATAG  
 GCTTTGATGAGAACTTTGGGTTTCATCATGGGTTTGAGGGGATGCCATTGAAAGGTTTGAGTAGGGGAGTGACT  
 TGATATGATTTATATTTTAAAAAATT  
 AGTAGTTGCTACATAAAGTGGACCATGGTGGGGAGTATGACTGGAAGTAGGGAGACAAGAAGCCATTTTAGTG  
 GGCTAGGTGAGAGAAATTGTTGTGGTC  
 CATACTGGTGTGGTAGTCGTGAAGGTGGCAGAAAGTACCTGGATGCCAAATATATTTTGAAAGTAAAGCTGATG  
 GAAATTGTTGATATAATGATAAATCC  
 TATGCTTAGGTACTGGATGATTGGTGATACCATTTACTGAGACCAGTACTGGGGAAGGAGCAGTTAGAGGTGGA  
 AGTCAAGAATTTTAAGTATGTAAAGT  
 TTGTTATGGCTATTAGGCATTCAAGTGGAGAAGTCAAGTGGACAATCAAATATCTGGGTCTGGAGTTTAGGAGA  
 GAGGTTGAAACTGGAAGTTTAATTTG  
 TGGTGCTACTAGCTAGAGAGAAATAAAGAGGAAAGAAAACAGAGTGAGGTACCAAGTTCTGGGCCATCTCCAA  
 CATTACAGGTTGAAAGAGAAAGAGGGAA  
 CCAGTACTTGAAAAGGAAGAAGAGTAGCTGGCCAGTAAAATAGGAGAAATTGTAGGGATAGTGGTGTCCCAAA  
 ATTCAAGCAAAGAAAGTTTTTAAAGAA  
 GGGGATGTAACAGTTAATTACAGGGGCTAAGAGAGCTAAGAGATGCTGATTCAGATAAAGAGTGAGAAATTAC  
 TTTTGGAATTGACCATCAGAACTTGTT  
 AGTGACCTTCGTGAAAGTTCTTTCATTGGAGTGGTGATGGCAAAAGCCAACCTGAAGTGGATTCAGGGGAAGTAAT  
 AATAGAGGGGGAAGTGAAGAAAATTA  
 TTTCAAGGTGATCCCTATACATGGGAATAGAGAAATGGAGTAGCTACAGGAGTCAAGAAAGAGTTTTATTTTG  
 GATGGGAAATATTTACCAGGTTTAC  
 ATGTTTACATGCTAATGGAAACATCTTAACTATTTAATTATTTCTGACTAATTGGAAGCAGTATGTGTTATGTT  
 CTAAAGTACATTAAAAGGATAACA  
 AGAAACATGGGTTTTTAACCATTATGGAACCTGTTTTCATACATAAAATGAGGGCTAGTCTAGGTGAGCTCAAAC  
 CTGACATTCTGAAATTCTGTAATTA  
 TCTACTGTGTGCTTCTCTTGGCCTCTCCAACTTTTTTCACCTACTATAGTAGGCAACCTTTATTGGTTGTTAATACTAA  
 TAGAGGCTGCAGACAGTCATTA  
 ACAATGTCACATGTACCCCAAAGTCCCAAAGCTGCTTATTGGCATTGGTATGTGGAGCCTAGTACTTCAAAAATT  
 CATGTCTGCCAAAGCCTACACGTCA  
 GCTGCTACCATGGTCAGAGGAAAAATGGACCCTCTCTCCCTTCTGCTTTCCACATCTGAAGCAAGTGCATCTTATT  
 GGCAGAATCTAAATTATATCCAGA

GCTCTGGTGGCAAATAGTTTGAGCGATATAGGTTTCAGGCTTCCATCCTTTGCTACACAGAGGAAGAGCACAG  
 GAGTCAGGAATAGATCATTGCCAGTA  
 GGTAATCCCACACAGCTAGTATAATTAACCTTAATGAAAAATTAGATTAATAATTCTTGCCCTAAATGCTTTTTTTTTG  
 GCCTTATCATTAGATGGAAGCTTG  
 TCATGTTTTCTTTCCTAGTCATTAAATGAATAGAGCATTACTGTGRAAGGGGTGCTTGGTATTCTTTAWTCTTA  
 AAGGAATTCCAATGCAAGAAAAAG  
 AATTTTGTTTAATTTTACTGTCTTTGTTCAATTTGGAGATGTGTAGTTTAAAGTCTTATTTCAATCTGTGCTATTTTG  
 GGTACACTTCTTTCTGTAGGGA  
 GCTTAGATATGCTCAGAAATGTGAGAATCTTAGCTACTATAGAAGGGAATGATGGGGAAGAATCTGTATTTGTC  
 AGTAGTATTTGTCAGTAGTAAAGAAG  
 TTCCTGGTTAGAGGAAGGAGAAGAATTATTCTGCTTTTATGCTGTTGTCAATGCCAAAGCATATCAAGATTTTAA  
 AAAAATTCTGTTTCTTRGTGTAAAG  
 AGCCTGTAAGTCTCTCTTTTTCTTTCTCATTTCACCTTTTTCTTCCGGTTACCTCTTTCAGTTGCTAAGAGTA  
 TCCCAAAGCTTATAAAATAT  
 TCTGTAATTTTAGTGAAGTACATTTCTTTGAAATATCTGTTGTATTTAAGCATTATGTCTTATAAAATATTCTGTA  
 ATTTTAGTGAAGTATTTCTTTG  
 AAATATCTGTTGTATCTAAGCATTATGTCTTTGATTTGACTTGTCTGTTTCATAGTTCTACCCTCTTCCTAGACTTT  
 GAGTTTTTAGAGAAGGTAAAGG  
 GGAAAATTAATGAAATATCTTTCAAAATGAAGGCCAAAATAAGGACATTTTCACATAGAGAAAACCAAGGGAATT  
 TGTCTCCAGAAGACCTGTGCTATAAG  
 AAATGCTAAAGATATTTTTTAAAAGACTTGGGTAAATGATACCAGATAGAACCCAGATTAGCAGGAAGGAAGA  
 GCATCAGAATGGGCAAATATTTGGGT  
 AAAAAGAAGCAGGACTTTTTAAAAGAACTATATTTATCTTATGTGTGTGTATTTAAATGTCTTTAAAAGACTATT  
 GTCTTTTTAAAGCAATAATAATGG  
 CAATTTATTATGGGGTTTACAGCTTATATAGAAGTAAATATAAGACAATAAGAACACAAGGGGGGGGAGGGAT  
 AAATGGAATAATACTCTTAATAAAATT  
 TTTATGTTGTTTGAATAATATTTCAAGATTGACTGCAGTAACTGATATCAAAGTATAATCTCTAGAGCAACTACTA  
 AAAGCTAACATAAAGAAGTACAGC  
 TAGAGCCAGGCACGGTGGCTTGTGCCTATAATCTCAGCTACTAGGGAGGCTGTGGCAGGAAGATTGCTTGAGTC  
 CAGGAGTTTGAGAAGCTGCGGTGAGC  
 TGTGATCACATCATTGCTCTCCTGCCTGGGCAACAAAGCAAGACCCTGTGTCTTAAAAAGAAAAGAAAAAATATA  
 GCTAAAAAGACAATGTAAGAAATAA  
 AAATGGAATACTAAAAAATATGCAGTTCAGTGAAAAGAAGACAGGAAAAGAGAAACACATAAAGAACAATAG  
 AATAAATTAATAATGCATACCAAGATG  
 GTAGACTTAAATCCAACCATATTAGTAGTTAAATTCAGTGGGCTAAACACATCAATTAAGGAAGAAATTTTCA  
 AACTAGATTTTTTTAAAAGGAAAG  
 ATCCAAGTATATAKTGTTTATAAGAAATAACACTTTAAATATGAAGACACTGACAGATGGAAAATAAGGTGAAA  
 AAGGATGTATGCCATGCAAACATGA  
 AGCATTAAAAAGCTGACATGGCTATATTAATCAGAAAAAGTAGACTTCCAAACAAAGATTATTCTAAGAGACAAC  
 AAAAGAAATTTTTTATTGATAAAAG  
 GAGCATACAAGGAAGATAAAACAATTCTAAATGTCTAGGCCCTAATAACAAAGCTTCAACGTACACAAAGCAA  
 AAAGTGACAGAAGTTGAGAAAAAGAT  
 CTACAGCATAAGTGGAGATTTTAACATCCTCTCATTCTGTGTAAATGCCTCTCACATTTTGTCAATTATGATCTT  
 TTGCTAAACCCATACAGCTGAAA

ATGCAAAAGCTCATGACTAGTTTTTCAGAGGTTCTTAGTTAATTGGTATAAATCTTGGATAGGTAAACGGAATA  
 TCTGTTTTCTGTTTGCCTGATATCT  
 CTGTAACCTGGCTTTTCCTAAGGCCTACCCATTGCTTTCTGAGACKAGGGAAATAAGTCTGGCTTCTTCAAACCTT  
 CTTTGCTCATTTTCATATGGGCCAA  
 AGTAGCACTAATCATTTAAAATTCACCAGTTAATCTGTTCATTTTCAACCTTATATTTATCAACTAGCTATAATTTA  
 GTTTTGGTTTTATTTACTACTTA  
 CTGGATAACTATGCCAGTAAGTATAGTATTTTATTATTTAACTGTGTTCTTCTTTTCATTTTATTGGTGATTTTTTTC  
 ATTTGTTGTTAAACAGAGCATT  
 TATTGAACACCTACAATATACAAGACCCACATTAGATTTTTATAAAATGAGATCCTTGACTTGTTACAGAGCCAGA  
 AGTGGGAACCCCTATATTCATCTC  
 TGCATATTGGTTCCATCCACCAAACAAAATTGTTTCCTTTCTTTTCTCCCTTAAACTTCTGTACATGCTTCTGTTTT  
 AGCGCTTATCGCATTGTATTGT  
 TTTCTATATCTGTCTTCCCAGTTARACTGCGAGCTTCTTAAAGGCAGAAGTCAAATATTTAATTTTGTAGCTTGG  
 GCACATGTAGGTGCTTTGATTACT  
 GAAAGATGGATATATGGACACAGTAATGAACAAATGATTAATAAATTATTTAAACCATTGTTTGACAGGGTAAG  
 ATTACTTTCTAATGAAATAGAACTTG  
 GTARTATGCCCTTGAAGTCTGATACCCTGACAAACCGCTTCTGAAAATCAGCTTACATAACTACCAGGGAGAATA  
 AAAAGGCATAGTGCAAAGCAAATGC  
 TGTGTGAAAGACAGGTTTCCAGTGTAGCATATGTGGGCCACTAATGGGTGTAGATTGAACTGTTCTCTTGCCAA  
 GAGTCATCAATAAGAACTCTTTAATG  
 ATGGCTTAGTTCATAGATGTATTTTGGCATGAAGATCTCTGTGTTTAACTCAGTAAACAATAGGATCTAGTTTCAT  
 TATGGACTTTTTCATACTTCTTAGG  
 AACTTGTCATTGGGAAGATAATTATTTTAGACAAGCAGACATTGTTATATGTTTCAGTAACACTTATATAGCTTGGT  
 CATAAAATTTGTGGTGCTTGGACT  
 AGCTTTTGGGCTCTTTCTTTAACTTAATGGAAAGTGCCTTAAATGCCTTAATCTCCATAAGACACTTCTCCCAGA  
 CTAGAATGGACCTAAAGAGTTCTA  
 ATATCTGTCAGTGTTTAATTATTTGAACTGAATATGTCAGAAATCTTAGCAATGGAAGTTTAACAAATTTAAGTAT  
 AGAATGAACCAGTATAGGAAGATT  
 TCATATAATTTTGTGATACAGTATTAACTTTTTTTTTTTTTTAATATTCCAGGGTGCTCTGGAACAGGGCTCAAATT  
 CTCAGCTGATGGCTGTTCAATAC  
 ACAGAAACCACTAGTAGTATCAGGTAAGACAGTGCAGAAGCTCTTGGCCAAGGAATTTAAATAGCTGGTATTT  
 TAAGAAAGTTGAAATTCACACTTCCC  
 TTTACCTTAAATTTGTTCTCATCTATGAGTTTTATATTAGTCATATACTCTAAGATATATATATATATGTATATATA  
 TATATAACATACTACTGAGCAG  
 TTGGAATAAAGCCAGATAATTTCAAAGTTCAACTCTCCCAAATCCCTTAAATACAAATTTAGTCAGAATGCTGTT  
 CCATCATAATTTTGATGAAAATCT  
 TGAATACCCATAGCTTTCTTGCCCTTGATTTGCATCTGTTTCATGAACCTAGGGAAGGGAATGTTATTTTTTGAGT  
 GCTCTGTACCAGGTGTCCTGCTGA  
 GTGTTTTACATGTATTTTATCTTTAAATCCATGTAATACTAAAAGTTTATTTTATATATAGAAAACTGGTCAAAT  
 CATTTATTTTTTTTGTAATTTTC  
 AGAGTCATAAGACCTAGTGGTAGACTCAGAATTCAGATCATAGCTGTTTTACTTACAATGCATTTTTTCACTACACA  
 ATAATACTTAAGATTTTGAAAGAT  
 TTCAAAGAGAATGCTGTATCTACAGGTTTAAATGTCTTCAAAGGGAAATATAATATGGGAAGCAATATGGTGTC  
 ATGGTTAAGATGCTAGCTTTATTTGC

CCTGTTGGGCAGAGTTCAAATTATAACTCTACCTCTTACTTTGTGGGCTCAAGCAGATTACATGCMATTGAAGTT  
 TTCTCATCTACAAAATGCTGGCTGC  
 ACAGAGTTGATATAATGGTTAAATTAGATAATATATTTAGAGCACTTAGCAYGGTACATAAGTGCCCCCAAAT  
 GGTAAATGRTAATGATGTGATGTCA  
 TTGACAACAGTATCACTGAAGTGAAATTTGAGGATCTCAGCTTCTATCTGTAGCTCTACCACACATTTGCAGTTCA  
 CTTATATGAATTTAACATTTACTT  
 AGAATATACTGTGTGCCAGTGTTCTAGGCACTGAGAAAAAATATAGATAGATAGATAGTAGTAAAAATGAAG  
 TCAAAAGCTATTTTATTCACTATACT  
 AATAACTGAGAATTCCTATTAAATGACAATGAGCAGCATATCAGCAATCATGAATGTTTAAAATCCTGAGATTGT  
 ATTATCCTGTTGAAATTGATAATTC  
 AAATATGTTTAAATAATAACTAATCATTCTCTGCTAACCAGAATATATCATTCAATTACATTTATTTTAAAAATATA  
 TCAATTTTTGGATGCATGGTTAT  
 AATGGCATGGTATTTCTTCATTAGTTAAAATGTTTTAGTGAATTACAGCCAATTTTCATTATTTGTGGATTCCATAT  
 TTGCTAATTCACCTACTAAAATC  
 TGCTAAAATGTATTTGTAACCCCTAATATTAGCAGTGCTTTTGTGCATCATTTGCAAACATGCGTAGATGCTCCCTG  
 GTGAGGTTGAACAAGGCAGTGTTT  
 TGCCTTTTAGTCTCAGCTATATAAACAAGGTCCTTTTCACAGTCCACATTTTTCATATTTTGTGCTTTTGTGGTA  
 ATTTTGTGTTTGAATGGCCC  
 CGAGGCATATTGCTGAATTGCTGTCTAGTGTTGCTAAGCCGAAGCGGGCTGTGATGTGCTGTGTGGAGAACTA  
 CGTATGTTAGATAAGTTTCATTACAGG  
 CATGATTTATAGTGCAGTTGGCTGTGAGTTCAAGTTAGTGAATGAATAATATATGTTAAAGAAGGTATCTCTAAA  
 CAGAAACACATAAATAAGGTTATAT  
 ATTTATCACTTAACAGAAATATTGTAACCAGAGGCTTGCAGGAACCTAACCCRTATTTCCCCCAGGAGGATGGT  
 TTAGGATTCATAATTCATTGCTGG  
 CAGTGACTTTTTAGAACATAACTATCACAAAATAGCAAGAAGCGACTATATATTCAAATATTGTAATCAGAAAT  
 GGCCTAGTGAACAAATGGAGTTATT  
 TCAAAGGAAATTCTTATGACTCATTTATTACTTCCATTTAGGTCAGTCTTTTTCTCCTCTTAATAATATTGATTGAT  
 ACATACAAGTTTTCTCTTTGAG  
 GAAAAATGAAGCATTTTTACAGGATCCCTTTAACTCTTCCAGATATTGGTCCTAAACCCTAGAAAATCCTGAACAC  
 ATTAAGAACTCTATTAGAATCATT  
 TGAATGATTTTTTTGGGGCATGATGTGGTTTGAGGTATTATCATTTTAGTTTTAATTTAGTAAGGCACATACAT  
 GGTACTTTTAATATCAATGACTAG  
 CAGTTGTTGAGTGACTCCTATGGGACAGGATCCATGCTTGGCACTTGACWTATAACATCTCCAGTAACAGTCCTA  
 CAAGGGTTGTGTGGTTATACTCTTT  
 TTATTGAGGAGGAAACAGTTTGTTTCATGGAGGGCAATTTCCCCAGGAGTGCCCACTTGAGAAGTGAAGCAGGA  
 ATTCTAAACCAAGTCTGCTTGTTTTCA  
 AGTCCTAAGCTTTTTCCACCTTTTTGCACAGCCTCGTAGGCAAGCATAACCTTCATATATTCGACAGGGTGGTGAAA  
 AATGACTTTTCCTCAGATTTAATT  
 TGTATTACTAAAAAATTAAAAATCGTTCAAACCTATGCAGAAATCTCTCTATTTTCTGATACTTTTATTATTTTAC  
 ATGCTTTGCAATGTACTTGACC  
 ATCTAGTACATTTTAGAGCCTTGATTCAGAAATCAGAATTTTAGCATAAGTGAATGGGTTCTTATAAAATATCCTG  
 ATCCTTGGGCTTTACAAAATAKTG  
 TTATAAAATTACTGCCAAAGTGCAACTAATATTTTGGTTGAGAGAGATTCTTAGAATTCAGACATAGCCTTAGA  
 ATGGGAAGTATTTTGGGAAAACAGG

AAAC TTTT TATTT TAGTAATACCTTATATTTTCTTACTTTAGCAGGAACTCAGGGAGTGAGCTACAAGTGTATTAT  
 GCTTCACCCAGAAGTTATCAAGA  
 CTTTTTGAAGCCATCCGCAGAAGGGGAGACACATTTTATGTTGTGTCATTTTGAAGGGTAAGTTCATCTTGAAA  
 GAATAGAGCAAATATTTTGTAGTGC  
 TTGCCGTACACAAGACATTGGGCTGAATATTGTGGAAAATACAAAGATGTGAGAAATAGGAAGCAATATAAAAT  
 AAATTATGTGAGTGGAACAGTCTAAG  
 TACTGTAATATATGCTTTACTTTGAGCTGTTGAGTATTATAAAGATGAACATGGAATTGGTGTGAGACTATGAGT  
 TTGTATTTAGGAAATATAAACTACT  
 TGTTGATTTTTATTGTTTGTAATCCTGATAGCCATGCTGTTTATAAAAAAAGGAAAAGTAAAAACATAAAAAATATT  
 CTATTTAGGCTTTTCCTATTTAGG  
 CTTTTAAACCAATTCCCATTTGGTTATTGAGAAACAGACTTCTGAAATGTAGGTTTGAGATGGATCATGGGCTGA  
 TTATTAATTGTGGAATGTTTGGATC  
 TATGTTCTAAGAATCTCTATTCCTCTTTTATCTCTTCTTAATTAAACATTTGGTTGGTGAAGAATGTGTTAGATAA  
 TCAGTTAGCTTATGTAAAAGTAA  
 ATGTTTATACTAGCACTTTTGAATGTTATGAATGTTAGGCTAAATGGAAAAAATGCATAGCTGTTTATATTTTCATC  
 TACTTTTTCTATAAAATGAAAATA  
 TTTTCTTCATAGAATTGCCATGTGAATATCTCTGTTAGAGTGCTGTAGGTTAACCCAGGAGCACATGAATGAATT  
 AGAATTCCTCAGTGGTGGTTTCTAT  
 CCCAAGCTGTTTTTCTTTCTTTCTCTGTTTTTAAACAAGAGAAGGGACCTCACTGTGTTGCCCAGGCTGGACTC  
 AAACCCCTGGGCTCAAGTGATAC  
 TGAGTAGCTGGCTTTCTTTTTTTTTTTAAGAATACACTGGGTATGGTTAGAAGTAAGTACAGTTGACCCTTGAAC  
 AACACAAGTTTGAATGAAGCAGGT  
 CCATTTATATGTGGATTTTTCTCAACCTTAGCAGGATGCCAAACCCACTTATATGCAGGGCCAAC TTTT CACACAT  
 GTAGATTCCACAGAGCTAACTGTG  
 GGACTTGAGTATGTACAGATTTGGAGTTATACAGGCGGCTCTGTAACCAGTCCTCCACATATACAGAGGGATGA  
 GTGTCTATACTAACAATAGAATGAGC  
 AAATTCATTCTCTGATTATTACCACTCACTGTCACTGAAATTCTCAGTTAACTAGTTAATTCTTTTCCTAAGGTATA  
 AATATCTAGCAGCTCAGGTTTTT  
 AGGACCAGAGAACTTTGATACTTGAAAAGTAACTAGGGAGATCAAACATCTTATATATTTTCAGTTATTAATCAAACAT  
 GCTAAGAGATTTGGAACACTTTAGC  
 CRTAGCGTCTCATGATCACTGCCTTTTCACAAAATTCCTGTGTTACCTTACTCTATCAGACATAACTGGTATGTT  
 TCTGTACAGGACCAGCCTCAAAG  
 TTTCTTTTCAGCAGCAAGGGGGCAGCTGAGAACAAGTTTCTGCAAGGCTCCAAC TGTACTGGTCTATGAAAA  
 CAAAGTATTTCCCTTACTATTGACAC  
 AATTATGTTCAAAGTAGGAGAAAAAAGTAACTTTTAAGAAGATTCAGTACTATGAAGAAGAAACCTGCTGGG  
 TGTTAAAATGATTGTGTTGAAGTATA  
 TTTTATGGAAGTTACTGTAGGAGAATGTTTTTAACAYTCAAATAGACAGTTTTTCTGAGGTT CAGGATAATTACTG  
 TATTATTTGATCAAAACAGTATAA  
 CAGATATTAGAAGTAACACTTGTC AAGCCATGGATTGAAGTCATGGTTAGTAATGTTTTTCTAAGGGAAGAGAA  
 CAAGTATTTTTTATAGAGAAGACAAT  
 CCTAGCTTTTTTTTTTTCTTTTGAGATGGATTCTCTCTGTTGCCCAGGCTGGAGTGCAGTGGTGCCAYCTTGGC  
 TTAACAACCTCCACCTCCAG  
 GTTCAAGTGATTCTCTTGCCTCAGCCTCCTGAATAGCTGGGACTATAGCCGCATGCCACCATGCCTGGCTAATTTT  
 TGTGTTTTTAGTAGAGACGGGGTT

TCACCATGTTGGCCAGGATGGTCTCCATCTCCTGACCTTGTGATCCACCCGCCTCGGCCTCCCAGAGTGCTGGAA  
 TTACAGGCATGAGCCACCGCGCCCA  
 ACCAATCCTAGCTTTAAACAGACTGTATTTTATAAAAGACAGTGAATGGCGGTTAAAAAATATATATGCATTG  
 CATTTTTTTGCATTTGTAAAAAGTG  
 ATTTCTCATTAGTTTGCTATAAAAGTGCTGTTAGAAAATTATTGATTTCCACTCCTCCCTAGTCCCTTGATGTTTCA  
 TCAGGAAGACTGTTAGGACTGGC  
 TCTCACACAATTACTATTAGGAGTATGTCTAACATTCTGTTTCGATTTTTTCATTTTCAAATTACTGATTTATCACATT  
 TTGGTGTTTTATATCAAGGAAG  
 AGTAAAGACAAATGAACTGAATTAGATAATAGTAATTAAGATAACATCCTTTTTATTTTTTTTGTGTTTCAATA  
 GTAGGGAAAAGGAAGTTAATTTT  
 ATCTTGATGACTGTAACTATGTTATGGGACAGGAATTTTGTCTTTAATACCTGGATTGTTCTTTTTGCTCTTGCT  
 GAAATGAAAATGAAATAATAAAT  
 TACCTCATGCTTGGTCTTCTTTCATCTTTCAATATAAAATATTCTAACAGTTTCATTAATAAGGTGAGTGGCATCAT  
 TTCTTGTCAGAAAGGATGTAGAC  
 TTTTCTCCCACTTAGGTTCTTAATTTTATTGGCAGAAATTATTATAATTGATACTACATTGGAAATTCAGAAGAAG  
 ATTCCAGAGCCCCAGGTTTCGGC  
 TTAGCATGAACCAGACATTTATAACAAAGCCAATAAGCAAGCTATGTTATTTTATTTTATGAATGGACATGTATTT  
 AACAAAATATTTAATATCTAAAG  
 ATGTAGAAGTAGCCATTGACTGTTTCGTACTGTTATTTAAGCCACTCTCTCCCAAACCTTACACTCCCTCACACCCTG  
 CATTCTGTACCTTTTAATCTCT  
 TTACAGTTGAGCGGTTGAAATTATTTTCTCTAAAGCAGACACACTCTTGGATAACTACTGAGTTCCTGTTATAGCT  
 CAGCAAGAAAAAATGCTTATGATC  
 CCAGTACTCCAGACAAAGGTCCAAAGTGCTTCTGTTCTTTTACTCTCCTTAAAGAGTTTTTACCACCAACATAGT  
 AATCTATTCGGCTTTCTTTGCTTA  
 GTTCCAGTTATTTGTAGCTTATTTTCTGTAAATACTTTGCCTTTTTCTCCTCTCTTTGATTTGTTTGTGTTTAGTATGA  
 TTTGGAGAGGGAGGTTGGAGA  
 ATAACTTTTCTCAAACCTCCAAGTCACTTTGAGTGTAGAATAACTGATACATTAATAAGAAAAAAATCCAATTAA  
 GTTTAATGCCTTTTTTGTGTTT  
 GTTTGTTTGTGTTTAAAGAGACAAGGTCTCTGTTGCCAGGATGGAGTGCAGTGGTACAATTATAGTTCACTTCAGC  
 CTCCAACCTGAGCTCAAACAGTT  
 CTCCTGTGTCAGTCTCCAGAGTATCTGGGACTACAGGAGAGTGCCACCACATCCAGCTAATTTTTATTTATTTTTT  
 GTAGAGATGGGGTCTCACTTTGTT  
 GCCCAGGCTTGTCTTGAACCTCCTGGACTCAATAATGCTTATTTTAAAAATACTTTTGTGGCCAGGCGCGGTGACTC  
 ACGCCTGTAATCCCAGCACGTTKG  
 GAGGCCGAGGCGGGTGGATCACGAGGTCAGGAGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCT  
 ACTAAAAATACAAAAAATTAGCCGGG  
 CGTGGTGGTGGGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGYGGGAGAATGGCGTGAACCTGGGAGAC  
 GGAGCTTGCAGTGAGCCGAGATCGCGCC  
 ACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAAAAAAAATGCTTTTGTATGATAGTCTATTTAG  
 GGAGGGAGGGAGGAGGTTGATTAAA  
 TAATATAACAAAATTAATTCAAGAGCTAAGACAGAAACAAATACTCAGGAAAGCTCAAATAAAGTTAAGAAGAA  
 GATGATTTTAAAAGGGCAGAAAAGAA  
 ACATTGAGAAGAAAGATTGGGAAATGATATTAATTTATTATTCAATACACATTTAAATACCATCCTTACATGTAG  
 CTGTCTCATCCACTAATATATAAG

GAAAGAATTTCTTCCTGGGGCAGTTTTTCATTAAAAACAGGTTCTCTTTGAATCAGATGGCTGTGAATGTCAGTTT  
 GCCATATTATCTCTACTTGTTTAA  
 TATTTTTCTTTTGCAAATGAGATGGATGCCTACTTTGGGAACATTTTTTATTTCTATTACATTTTCTGTTTTCAGTAC  
 CTGTCTCAAATTGTAAGTAAAC  
 ACGTAAAGTGTCTTTTCATCAGTTTCAAATACATACATAATATTACAGAAGAATCTGAGCTATAGCCATATGACTG  
 CCTGAGAGATGTCTCTTTCTCCTC  
 TCTTTCCAAGTATTTTATATATTCTAGCAATAGTACCATATTTGGGTATGCACCCCTTACTGGTCCCTTATAAAAAT  
 TATTAACACTGTCATATATACGG  
 CATGGATTATAAGGGTAGCCTTAAGTAATAGGGACTTATTGTAAGTTATGTGAGATTAAAGAAGCACAGGAAAC  
 TGGCTCAGAATCCAAATAATCCCTCT  
 GTGCTTGTGCTTCATTGGGTGGAATGTATACTAGTCTTCATGGAGAATTGGAATAGTGGTTTATTGTCATTTCTGA  
 TTTGACAGCAGCTCCAGTGATTCC  
 ACTTTAGTCTTTTTTTTACACTTTACTGATTCATACTCCTCAAACACAAAGATCTGATTGGGTAGTTCTGGTCACCA  
 TCTCTTACGCAGAGAATTACTAT  
 GGCAGCTTTGGCTTATGATCCACAGTTATTCATTTGTGACAAGAGAAAAACCTGGGGCCACTCTTCTGAGAAAAT  
 GACTTTGGYAGGCACTGGGAGCTTC  
 ATAAACTACTATACCACACAGCCTGTTTTTTTTAATTATTATTCTTTTTTCCCCGTTCTTAGTAAGTGTGCATTAAAT  
 TCTTTCTTGATTGGTTCCTAA  
 TTGTTTGATCTAAAAGGCCCATCTTCTCATCTATATGATAAGCCTTTTGAAGGCAGGAAATATCGTATGCTTTTGT  
 GACTACCCCTGACACTCAGCAATG  
 CCTACCACTGAGTGCTTACTGAACAGAATGGTTTTAGACATGAGCCATTGAGAGAAGTCACAGTGATGGAAGTT  
 TGTTGTTTTTTCATAATCATTTAAAAG  
 GTTTTTTTTTTACATTGTTTTGCTTAAAGAAATTGGAATCTATTTTACTGATCTAAGAAGTAATATAAGCTGTGCT  
 GAATGCCAGTAGAAAGTTTAGTT  
 AAAGATGATCTCTGTTTTGTCCATTTATGATTTGCACTAGCCAGATATCATCAGCTTATTCTTGTCTCTGAATCTAA  
 GCTGTAGAGCTAAATAATTCACC  
 TGAGTGTAGTGTTTAACAGCTTTCTATGTCAGTAAGTAGAACTTTGAGATACTTGCATCTTCATTTAAAATTATTT  
 GTTGGCAGCATTTTGGAGTAGATT  
 GTTGTTAGACCTCACTGATTTCTTTTTTGGAGAAGGGATGGGTTCTACTATTTGTACAGATTCAAGCAGTAGGAG  
 CCAGAAACTTGAAACTTTGGAAACA  
 CTGTAGGAAATACTGTGATTATTGCAAAAAAAAAAGGTTGACTAGTGGTGTTATTTGTTCAAATGTTTGCCAGAG  
 GTTTAGTCCCACGGATGGGCATAAA  
 AACGAATTTCAAAAATAGCAATGATATTTTCTAACTTAATACCTAGATTAATACCTAACACATGAAGAATGGCATA  
 TTTTCTTAATGCCTGTTATTTATT  
 TCTTAGAAATTTATGTTTAGTTTATATGCATATATAAAAATTCAACATATGTATATATATAAAAATTCAATTGTATG  
 TATATATATAAAATCTTCAGAGG  
 CAGGAGGAATTTCTTTAGTATCAAAGGAAATTCATTTATAAAGTCTACTTCAAAGGGTACAGCTCCTTCTTTTGT  
 TTCTGAGACCTCAAACCTTGAGTG  
 TACACATAGATGTTGGAGTTTGGCCATTTTCTTATCTCCACTTTCATGGCCTTTGAGCCAAATTCTAGCTGGAAAA  
 AAAAGTGAATAGATGAGTTTGGAA  
 AGATACTTGCATTTTAGAAAAGGGTCTTATCTCAGAAGAGAGATGTTTTAAAAAACCTTCTTTTAGAAGGCTTTTT  
 AAAATTAATTTTTAGCTTTAAAT  
 ATATTCTCATTGAGGTCTAAAGCACATTATAAAAAAATAGAGCAGAGTTGACTGAGTACCCTAGGTTAATAAACT  
 TAACCTAGATTTATTTAATCATTAT

GCCAATAATGTTCTTTTTATATGTCCATTTTCTATTTCACTTTATTTATACAAATATTAACACTTCAATGTATACTA  
 CCTTTGTAGTTTACAAACCTCT  
 TTCACATACATAATCTTAGCAATTCCTTACATATAATCACAAGTATTCTTAGTCCTATTTTATGGCTGAAAAAGCTG  
 AAGCAAAGAGAGATTAAGTAGCT  
 TATCTAAGGACACAAAAATGGTAAGCATTAGAACTAGGGCAATAATTGATAGATTCTAATTTAATACAGTGATTT  
 TTTTTCTTTACATTGKGTGCTTT  
 TTTCTTTATATCTGGCACTTTTAGATGTATTACAGCATATGTAAATAAATTAATCTAACCTCCAAAATCAAGATAT  
 TTCATTGAAAAATCTGCTGGACA  
 GTTTTGGCAGATTCATTACACAACCTTCAAAAATTGTTTGTTTGTTAATTACCTGATTATGTTTGTTTGTTAATTACCT  
 GTTAATTATAATTATAGATTAA  
 GTACTTTAATAGATTATGCTTTCATAATGCTGGAGAATAACTTAATGGTTTGAAATATTATAGTATAGCTTCAGAT  
 TTCAAGTCAGGGAATAGAATAAAT  
 AGGCCTTGAAAGTTTCTGGTTTCTTTTTAAAACAGTGTCTGAAATTTCCAACATGGTGGAAAAGAAATGTCTTTGC  
 TTTGTCTTAAAAAGGAAGTAGTTT  
 AGCAAAAACAGAAGCTGTTTTAATTATTGGTATTCTAGAGCTTCTCATAGAAAAGAGTGGGGTTTTTGTGTGTGT  
 AAATACATTTTATTATGGATTCTGT  
 ATATGGCGTGCTAGATCTTTGAGATCAGAAGCATTGATCTCAAACCTCTGTAAATGTGAGTTACTGTTGTGTG  
 GTTTTAGAGGATTCAACTATTTGT  
 TCTTTGTCTCTATAACATTGCTAGCTAACATAAAAACATGTTTTTTACTTACACTTTTCAATCACAAAGTCTAAAAT  
 AACTTGCTTTATGTTTGAGAAAA  
 GATTGTTCAGTAGTCTGTAAGTAGACTTGTATATAGCCACAAACAGAGAACTTATAACTCAAGAATAATACGAAG  
 TCTTCAAAATTCTCAGAATGGTGGG  
 CAGATCCATCTGAGTCATGATAATTTTTCAATGGTTCTTATTTGTTTGCTCTATTTGGTAGTCTTATTATTATGAAA  
 TGTGTTTTTCATATGCATTTGTT  
 TAATCTCCCTTTTCAGGTTGTATGCTTTCTCTGTGCAGGGATAAAGTCTATTCATTCTGTTTTGTCTTTTACAAGA  
 TCTATTGCAATGCATTGCAGGCT  
 CGGCAGAGTAAGTCCTAAATAAATGTTGACTACTCAGACCACTAACTGTTTATTTGTGTTGTTTACCCTTATTTTTT  
 ATGTCCTACTCTTCAAAAGATTA  
 CTAGAGGTGGGGTTTGCAGAAACAAAGGAGGACATTTTTCTCAATGAATGTTTAAAATTACTTCGGTGACTTGG  
 ATAGCAATGAAATTTTTAAGAATATA  
 TAGTCAGCTTTGGTGTGGGTGGACTTTTTGCTGTGGAATAGAATAATTCCTGAAAGACTTCTAAAGTTACTTTGA  
 AAAAAGGCAGGAATTAGGGTGGGAG  
 GTGGGAGAGGAGCAGAAAAAATAACTATTGAGTAACCAGGCTTACTATCTGGGTGACCAAATAATCCATACAAC  
 AAACCCCGTGACATGAGTTTACCTA  
 TAAAACAAACATGCACATGTACCCCTGAACCTAAAATAAAAAGTTAAAAAAGAAGAAAGGTAGGATTTAATGATT  
 GATAGCTTAAAGTTCTCATCAAATTT  
 TTTGTAATGAGGGAAGAAAATATTTCCCAAAGCAGAACTGCATTCCCAGAAAGAACATTTCTTCACATCATGAA  
 TATGTGAGGAAAAAATGTATTTTAA  
 ACAACTGATAGGTACAGACCCTAGGAACCTGCATTCTAGCCCCAATGTTGCCACTCAGTGACATCTGCTATGTGA  
 CTTTAGACAAATCACTTAGTCCCTC  
 TCATCTTCAGTTTCCTTATATGTTAAGTGATAAAGTTAATGTAGATTTGCTTTTTCTTCTGGCTCCAGAAATCTA  
 TAGTTTTATCTGTGTGTTTGTG  
 TGTGTGTGTGCTTTCCTCTTAGAAAGCTGTAGTATTGCTTTAAACTCTTTGAAGTTACTAAGTGATTCAATCTAGT  
 GGAGTCAGTGTATTTAAGAGTTT

TGAAAAACAGCCATAACAAAAACAACAAAATGTGAAGTTACAAGGTGGGGCTATAAATTAATGAGTTTGTTTTTT  
 GAATAGCTGGCAGTTTTGTAATTTA  
 TTTAATAACACTTATACAGAACACAATACGATAATCTTTAATCTTTATAATTGAGATAGGTGCAGTTATCATCATTT  
 TACAGATGAAGAACCTGAGGCAG  
 TCGTGTCCATTGTGAACTCCAATAAGAAAGATTAAAAAGAAGGAAGAAAAAGTCCGCTGTAGAAATGAAAAAC  
 AAATTTATCCATTGTGACATCRTTA  
 GAAATAGCAGTCAGAATTTAACCAAGAAAGTCTGGCTCTATGTTCTTGATCATTTCCCTCTATTGCCTGTTTAAGA  
 TATCCAAATATCATATCTTTGAAG  
 GTAGTCAACTTGGAGTTTAATTCCGTTGATCTCTCTTTTCAAACATATTTTGAGTTGTTTCRGGATCATTTTCCAG  
 AGACTAAGACAAAGCCCTGTGGA  
 AAACCAAGCCACATGTTACTCAAAGTAATAGTGTAGGAGGAACAAGAGCCAGCAGGGAAACACAGAGGAAATT  
 ATTAGAAATATAGGAGAATTGGAAGAG  
 TACAGATGAAGGAAAATAAATCAACTTGTTTATGTTATTTTAAACCTAGAGGTCAAGGAAGAGTTCAGAAAAC  
 GAGGCTGTCAGTGGTATCAGATGTTA  
 TAGCGCTCCCTAAGAATGAGAAATGAAAGAAACTCATTTGTCTTGTTGTAACATTAAAATCATTGGTAACTTGAG  
 ATTACATTTATTGTAGAGTAATCAG  
 AACAGAATCCAGAAAAGAGATAGAAGAGAATGGGAAAGAAAAACAATGGTTTGGTAAATCTGACACTGGAAG  
 GAAATGGAGATAGCCTAGATATAGATAA  
 GGAGAAGAGAACTAAGAGTGTGTATGTATGTGTTTGTGTTAAGGACAGGAGAGATATAGGGATAGGGTATAAG  
 GTAAATAAGGACTATTGTAGTTAGTTT  
 ATATACATATAAGAATAGAGTTTATGTTTTATGCTTGACAGATGTTCTAAAGGAAAATAAACAAGTCATTGTTCT  
 CCTCTGTATTGTTTTCTGTTTAA  
 GTAAAAGCATGGAGAAAGTTGAACAGTTTTTGTGTTATTTGCAGTAGGACTACAGAAAAGAAAAATAAAAGAGCG  
 TATGGAATTAACCAGATTTTTGAA  
 AAATCAAATCATATTTAAAGTACTGCTTGAGATATCAAAGTCTTTTAAATCTTAAATACAGAGCTGGTTTTTAAT  
 AATGGTAAAGAGTTTTCAAATGT  
 TAAAGTTTATAGAACTATTTCTCATTTAATTATATGTAAATGGTTTTATAATTTTTCTCACATTAAGTTAGAAAAC  
 TAGCAAAGTAACCTACATAAAAT  
 TTGGTATAACCTTTCTTATGATGTCACAAACTTAATTATGGAGGAAGCAATTATTATAATTAGGAAAAATTGATG  
 ACTTATGATCTGACTCAGGATCAA  
 GGTCTGAAACACTCATGTCATTTCCCTCCCCTCTCCTATATAGTATACAGTGATTTGTTTATTGACTCTTAATAT  
 TTTAACTTTTGACACCATCAGCA  
 AATCTAAACACTGACAGTTGGTGTGTTTCTGAAAAATGTTAAGGCTAATAATTGTGTGATATTTATTCAACACAT  
 TGCTGCTAGTCTTCTATAAAGATC  
 TCATTATGGGTAGGCTTTGAAGAGATATCCAGTTCATAATATGGGCATCACACTGGAAGCCTGCAAGATGGATG  
 ATTGGCTGAGAGTAGGAGTGTAGTTT  
 CTAATAATAATAAGTCTGTTTCCCTGCTAGGAAGTACCCTGAGGGCAGGAACCCTCTATCTTAGTCATGACTTT  
 GTTCCCAATACATGGTATATCAGG  
 CATGTAGCAGGTATTCAATAAACATTGAATGAATAAATGAACAAATTAATTAGATACCAAATCCATCAGTTTGAC  
 ACTATATTCTAAATTAGTTTATATA  
 GTCACCGTAATGTGATAAAGAAATTGTGATAGTCTAGTTTACTGAGCATTGTTATCAGGACCATCTGTGTCTCAT  
 GTTTAAGTCTAGAAATCAGTCTTTG  
 AAAATGAGTGAATTCAAATGCATTTTTAAAGTCTAAGTGCCATTTTGCATGAGGGAAATTATATACCTAGTAAG  
 AATTTCTCTGGGTTTCAAAGTCAGA

ACACTGATCTTGGTAATACATGGACAGAAAGGATTATTTTTCAGTTAGGGTGTTAAACAGATTGTGCTATCATTTT  
AACCTCTGGAATGAATAAGCCCCA  
GATTTAATAAGCTCCTTATGTATAAGATGCCTGTTATAATGTGTCTGTAATGTCCTAGAAAGATGAGACTACCCCT  
GCATAGTGCTTATTTCTTGCATGG  
AATGAATTTAAGATAAGCTTGTCAAATTTTAAATCTGACATAGACTATCTTTTGAATAATTTATTTGGAGTCATCTC  
TGTCTGATTATAATATGCCCTAA  
AATAAATCTTGAAAAATTAATAATATATCATACAGAATAAACAGTCTACATGTAACAGTTAACAGTTACATGTAAA  
GTGTTATTACAGATTACATTTTACAA  
CTATTTCTTTTCTGATTCCATTTTTTATTGAATAGAAGTTCAATAAATTTTATTCTGTTAAGAATGAGTCTGTCTTGG  
AGAGGGTCAGTGTACATATTT  
ACTTTGAAGTTCCTTAAAGAGGTTAGTTGAATTCTCCCCCAAATGTGCTTTGAATTAGCAGCTTTTGAATATCT  
ATTCTAGTAGAGGTAGGAGCAATA  
TGAAGTAGAGAGAATGATATGCAATAAAAAATAATAATTTTCTTTGACTTTGGAATGAGGATAATCTAAACCACA  
GCCATGTAGGCAGACATTTTCTGAT  
GTATGTAGGTATGTTTGCATGTCTGTCTATCTATCCATCTCCTTGTTTTCTATATATTTGACCCTAACCTTTCAGGCT  
CCTATTTTGGCATCAGGTCCAA  
AACCAAAGTGAAGCGATAAGAGAGTGGAGATCCTTAAATCAGTGGTTCTCAAACCTTTTGGTTTCCACATTCTTTA  
TGTTTTCAAAGAAACATTAATGAG  
AAGAGTAGCATTTTGTTAACATTAAGAAAAACCTCTTTAATAACTGGCTTAATAGAAACCAGCTGGATTCTTATAT  
CTGTTCTTTTGGTCAAAATATAT  
GAAGAAAATCCAGCCTCACACAGAGATGTAATTGGAAAGGAAGGACTATTTTAAAAGCCTTTCAGATAATTGTG  
GGTATTCTTTGATGCTACACCAAAAT  
GCAACAAGAGGTAGTTTCTTAAAGGTTAGTTGCAACACGGAAGCCAAATCTGTGAACTTTTAGTACTCTGAAATT  
AAAATCCATTGGTTGATCTTGCCT  
TTGAATGGATCTTTTACCTGTGCAAGGTTTTGTTACATCATACATTGGTCATTTGAAAATACTGGTTCACTGAGTT  
ATGCAGATCTTCCAAATGTTGACA  
CATTTTATTATGTATTAATAATGTCACATTCTTTAATACTATCACAAATCTTATCAGAAAAGTCTTTAAATATTGTGA  
GCAGTTCATGGTAGCAGATAGAA  
GTTTTCCAAAATTCTAGTTTTCACTTGAGGGTTCAAATTTTATCATTGTCAAAAATATGTCAGTTTTCTTGAAGTG  
GTAGATTCACTAATTTATTTCA  
AGAACATATCTGTCAGATACCCAAGGCTAAACAACCTATAATTTATCAGTCTTTCAAGTAAAAGTGGTGTTCCATG  
AAGAAAGTAGCTAGTGCAGCTCACA  
ACTCAGTTACACAAGTGGTTTTCTCTAGACAACAATTGTACTTTGATATGCAGCAGAAATGCTTTATGTGTATTT  
CCCATTTTGTCTGACAGAAATATTA  
AAAAGACTTCTACTCAAGGGTCAAGATTTTATAAAATTAATAACTTTTATTGCTTCATCAAAGACTTCCTTCAATTA  
CATTGACTGTGTGTGGTATGTGC  
ATGTATATGAACATAGGGCAGTAATATGATGAACTATATAATACAGTTTGATGTTGCTGTCTTAATTCATGCTAG  
GGCACCATCATTGCTTTTGCATCAT  
CAGTGTAATATCAACACAGTTGTTTCAAGATAAATGATGAGATTTAAGAAAGTTATTCAACACTGAATATTTCA  
GTACTGTTAGTGTTTGTGCAAGC  
ATTCCTATAAAAGGTCTTTGTTGTTGATTACTTGGGGCTCATACTGAATAAATCTAAGCAAAACATAAAGTTTATA  
GATTCTTCCATTTGTAAAACAAAA  
GCACAATTCTGTAGTTGAGATATTAACCTCAGCCTTCGTATTTGCAACTAAGTGTAATTTGATGAATTAACATATGG  
TTGGAAAGTGTCATGATTCTTAT

ACTGACTTTTCAGCAGGCATTGAGCAAGGTCGATTATCCACAATTTTATTAGTCTCTCAGCTATTATATGCACTTTT  
 TCAGTCATTGCATTATATTAATA  
 ACTTACAATGTAGGATACTTCCGTGGTCTTTTCATTTCCAGTATGATAAGCTGTAATAGAATTTTTTTTAGAGGA  
 TTCTTCACACCTGTGTTTAAAATA  
 TCCAGTTCCATTTTCTTTGAACTCTACATGATTGGTTTCAGAATGACATCACAACCTTAATTGGCACCATTATAGTAT  
 TTGAAAGATTTTGTTCATAAGC  
 GACAAGAAGATAGATTATTCATATGTCTGAAACAAGAGAAAGATGGCTTTCAATGTATGGTCATTTTTTATTAAC  
 AATTTTGACCAGTTTCTTTGCTAGA  
 TTCCACCTTTCCGTGAGTCATAAAATTCTCAAATATATCAGTTTTATCTTTTCTGTATCTTATATGTAGACAGTATA  
 GTTGTGGGTAAAGGAGGCAAGTT  
 TTCTAATCTCCTGTCTTTTAAGCCAGCAATTCATTCTTAAATATAAGAAAAGTATGTAATAGATTTTTATATGAAAT  
 ATTAAAATTTAAGAAGCTGTAGA  
 TAGATTTAGAAAACCTTTCCAAAAAGATTTAAAAATAATATTATAAAACAAAACAAAGTATACATTCTTCTTTGTT  
 GTATTTAGATAGAAGTTTCAGTAG  
 TGGAGTCCATATAACTAAGCTCCATAATAATGTGTTTCATAACTTTCTCCATTATTCTACTACTAGATCATAGAAG  
 GCTTGTTTCTCTTAAGGTTTAGTT  
 TTTTCATGCACTCAGAAAATAGTTTTTCAGGAAGTATGTACTGAGTATTCTGAGGATATGCAAGGAGCATGGGTT  
 CCAGTGATAAATAACATAGTCTC  
 TGTGCTCAGGGAACCTTATAGTTTAGTTGGTGCTAACCAGATAAATTAGCAAAATAATTGACAGATATAGTAAGCA  
 TGATGAAAACACACAATGGACTCAG  
 AGAATAACAAGGATGCTCTCCTGAGACAGCGACATTTAGCTGATATCTCAACATGGGGGGAATTCAAGGATTC  
 CAGGCAGAGGAAAGAGCATATGAAGG  
 CAGTTAAGTAAGGCAGGAAAGAGCTTTATTTGAGGCACTGAAAGAAAACCAATGTGACAGGAGACTAGTGAAT  
 GGGAAAGGGGGATGAGTAATGAAAATGA  
 GGTGTTGGGAGATTGGAAGGGGCTAGTTCTACAAAGCCTTATGAGCTATAGTAAGTAGTTTGGGTTTTATTCTGA  
 GGGCCAGCAAAGGGATTTAAGCAGCA  
 GAATAACACGAATGATTTAGTTCGTAGAAGATGTTCACTGCTACATGGACAATGCATTAGAGAGAGACCTATTA  
 ACAGTAGGAGCAGAGAGAGCAGAGAT  
 ACCAATTAGAAGGCTTTTTTCAGTAGCTCCGGCAAATGATGATGGCAGCAAAGTGATGAATGGATTCTAAATAC  
 ATTTTGTGGATAGATTTAGTGAGAAT  
 TGGAGAGATATTAGATGTGTTTCAGATTTCTGTCTTGAATTACTAGGTAGAAAGGACTATCATTTTTTACTAAGAT  
 GAAAATATTTGGGGGGCACGGTAAT  
 ATGTTTGGTATAAATGTGGAGAGTTATTTTGGATACATTTAAATTTGAAATGTCTGTGAGACATTCAAGGGAAT  
 GTGTGCAATTGGCTCTGTGAATCTG  
 GGACTCAAGAGAGGATAGATCTGAGCTAAGAAGAGTAATTAAGAATTGTCAGCTCATACACATTTATTTATTTAT  
 TTGGAGACGAGTCTTATTCCATCTC  
 CCAGGCTGGAGTGCAATGGTGTGATCTCAGCCCCCTGCAACCTCTGCCTCCTGGGTTTCAGATGATCCTCCTGCCT  
 CAGCCTCCTGAGAAGCTGGGATTAC  
 AAGTGTTTGCCACCACACCTGACTAATTTTTGTATGTTTTAGTAGAGATAGGGTTTCACCATGTTGGCCAAGCTG  
 GTCTTGAACCTCCAACCTCAAGTGA  
 TCCACCCACCTCAGCCTCCCAAAGTGTTGTGATTATAGGCATGAGCCATCACACCCGGTCAGCTTACACACATTAT  
 CTCATGGATTTACATGGGAATAAT  
 GGAGTCATCAAGAAGAGAGTATATAGAGAGAGGAGGGTTCAGGACCAAATTAGATTTTGAATTGGAGAAGGAT  
 GAGAGCAGCAAAGGACGGTTCAGGACC

AAATTAGATTTTGAATTGGAGAAGGATGAGAGCAGCAAAGGAGGAGGGTTCAGGACCAAATTAGATTTTGAAT  
 TGGAGAAGGATGAGAGCAGCAAAGGAG  
 ACTAAATAGGAGCGGTCAGCAAATAGGGAGAAAACCAAGAGAAGAGACTTTTGAATGGGGAAGGAGTGGTTT  
 ATTCTTGAATACTGTTGAGACGTTAGAT  
 AAGATGAAGGATTGGTTTTCTACTGGATATGTCAACATGAAAGTCACTGATACACTTGACAAGAGCCTTTTAGTGA  
 TGACATGAAACATAAAAGTTAAAGA  
 TGGATTAAGTAGTAAGGAAGTGAGAACAGCATAACTATTTAAAAATGTTTGCCTATAAAAGAAAATAGAGAAAA  
 GGTAGGGTTTTGTCTTATTTTCAAAC  
 TCTAGATATCAATTATATATTTTAAAGGAATGAACCCATAGATAGGAAGAAAGCAATGTTAAAAGAGAAGTCGA  
 GTCCTTGAGAAGGGGGCAAAGGGTGG  
 AATTCAGATAACAGGTGGAAGGATGGCCCTTTGCTAGTAAGGCAGGACATTATGGTAGARGCAAGATGAGTTC  
 ATAGATAGGACCTTAGGTTCCAGGAAG  
 ATAAAGAGAAGACACTTCTAACAGTTTCTGTTTTCCAGTGAAGGACATCTCAACTGATGTCATCAGCTGAAAGT  
 AAGGATATTAGAGGAGGTGTAGAAG  
 ATTTAAGCAGAGAGGCACAAGTATGAAGTAGTTGTCTTGGCCAGTGGGATAGTGAGCCTACTAGAGACATTTGG  
 ATTGTGAGGCCTTGGTGGGGTGTGCTAG  
 TATGCCGTATGCTTTAAAGTCAAATCTCTGCAACARAATGCCAGGTGTAGAGTTACAGTTGATTAATAAGACACA  
 GCCCCTGTACTCAAGGAGTTTGTA  
 TCTAGAGAATGAAAGACATATAATTACAATATGGTGTGATGGTTGTGATGAACAAGGTAAATATAGGGTGTGAT  
 AAGAGCACACATTAACCAATCTTGGA  
 GGGGCTGTCCTATGCAAGGAGGTCTTCATGGAGACCTAACAGCTAAGCTAAGACCTGCCTGTTCTGCAAGTACC  
 TTAATAAGCTATTACATAATAGCT  
 GCTGGGGATGAAAATATTCTAGAAGGTATGATAGTCACACTGTATTATTATATCTTTATGCAGTAGCTTTTATTTT  
 TGCTGACATCATCAGGCATCTTTG  
 AAATTTCTTTTTTTTTTAACTTCGAAATGCTGTGAAGCAAAGTATCATCCCTCTCTTCCTCACCTCTTCTCTTC  
 CTTAACACCTTTCTTCAGTTCT  
 AGCTGTTGGCAGCTAGATGCAACTTAATAATGCCATTTTGTCTTCATTTGGCATGTCATTCATCATTTTCAAATAA  
 ATAACTTTAATTTGTGCTGCC  
 ACAATCAGGAATATTTGAACATTTAGTGCTTATCTTGGGTATCTGCCCTTATTACTTTCTGAGGAGCTGCCCTTTAT  
 AGCAATCAGGGAACCTATACCACT  
 TGGAAGACTAATTGTCTTTATACAAAAATAAAATCCTTGTGTTGAGATTTCTGTCTTGAGTGGCTGGTTTGAATCT  
 AATTTCTGTTATATGAAGCAAAG  
 ATTTATTATTTAACATAGTTTCAAATTAATAATTGATAGTGAAATTTTATGTGCCATGCAATGTCATGTGTGTTCTT  
 TTAAGGTGAAGATTCATTTTTTA  
 ATAGCTTTATTGAGGTATAATTGACMTATAATAAACTGCACATATTTAACACCACATTTTATGTGTGATATGTAKG  
 TGTGTGTGTGTGTGTGTGTGTATA  
 TATGTATATACTTATATAGGTGAAATCATCACCATAGTCAAATAATCAGCATCCCCATCAGTCCCAAAGTTTTG  
 TTTCTCATGCCCTTTGTAATTTT  
 TCCCTGCCCTCTCACCAGTCAACTGATCTTTCTGTTACTACAAATTAGTTTATAGTTTTAAAAATTGTATATAAAGG  
 GATCTTAATTATATATATTTTCC  
 TTCCAGCAGAATTATTTGAGATTAATCCATGGTGTGCTTGTATCAGTAGTTCGTCTTTACTGCTAATATTCTGGT  
 GTAGGGATATAGCACAATTTACT  
 AATATATTCACCTTCTACTGTTAGCTACTACAAACAAAGCAACTGTGACTATTTGTGTACAAGTCTTTGTAGGGAC  
 AAATGCACTCACTTCCCTTGTGTA

AATACCTAGGAGTGGGATGGCTGGGTTATATAGTAGGCATATGTTTAGCTTTTTAAGAACTACCAAACACTTTT  
 GGTACAAGTGGTTGTACCATTTTAC  
 ATTCCCACCAGTGGGAATAAGAGTTGCAGTGCCTCCATATCCTTGGCAACACTTGATACAGCCTATCTTTTTATTT  
 TTAGCCATTCTAATAGAGGTATGT  
 AGTGGTATCTCATTGTGGTTTTAATTTGCATTTCCCTAATAACTATATTAATTTTGACCAACTTTTTATGTGCTTATT  
 TGCCATCCATAAATCTTCTTTG  
 GTGAAGGGTCTGTTCAAATCTTTCCGATTTTTTATTGAGTTATTTTCTAATGATTGAGATTTTCATTTTGAGGATTT  
 TGTGTGTGTGTTGTTTTGTTTAG  
 CAGCTTAGAACACACACATTTATTATCTCACAGTTACTATGGGTCAGGAGTCCAGGCATAGATTTGCTGGGTCTT  
 CTA CTCCAGGGTTTCATCTGAAAGT  
 AACAAATCACAGTCTTGGCCAGGGCTGCAGTTGCATCTCAAGGTTTTACTGGGGAAGGTTCCACTCATAAGTTCAC  
 ATGATTGTTGGTAGAATTCAGCTCC  
 TCGTGGGATGCCGAACAAAGGGTCTCAGTTATTTGCTCAGTTGGAGATTCCCCTCAGTTCCTGCCACATGGGCCT  
 CTTGATATGGCAGCTCACAAACATGA  
 CAACTTGCTTCATGAAAGCCAGCATGGAAGTGTTTATGTTTACTGGTTTATTATGATTAACAAAGGATACAGATG  
 AATAGCCCGATGGAAGAGATGCATA  
 GGGCAAGCTATGTGGAAGGGGTATGGAGCTTCCATACCCTCTCCAAGTAGGCTACCTTCCAGGCACCTCCTCATG  
 TTCAGCAATCTGGAAGCTTTGCGAG  
 TTCTTTATGTATTCTATATACAAGTCCTTTATCAAATACATGACTTGTAATATTTTCTCCCAGTTTTTTGCTTGTAT  
 TTAAATTTCTTTTTAGTGTTT  
 TTCAAAGGGCAGAAATCTTAATTTTGATGAAGTCCAATATGTCAGCTTTTTATTTTATGGACTGTGTTTTTATGGT  
 TTTAAGAAATATTTGCCTAACTA  
 AAGGTCACAAATATTTTCTCATATTTTTTTTCTTAAAGTCTTACAATTTTAGGTTTTACATTTAGGTGTATGATCCAT  
 TTTGAATTAATTTCTGTATATG  
 TGCAAGGTGTATATCAAAGTTCATTTTGTGTGTGTGGATATCCAGTATTGTATGATATTCAATGCTGTGTTTTCAC  
 CTAATAGCTTTTAAAGGAAGAATT  
 AAAGCAAAGCCAAGTTAAATAGGTTATACCAGACTGTTAGTTTTCTTTGTGTTATCCTTGACTTTGTTAGGTTGCA  
 CATTTTGCTGGAACCTTCTTAAAC  
 AGTATCCTATGATTGAACTTCTTCAACCTCAAATTTCCAGTCTTTATACTTAACCAATGAAATGTTATTTATATACTC  
 TTTGGAAACATAACATATACTG  
 TAGATAATTATARCAAAAGTGTAATGGGTAGCAGAAGATATTGCCAAACCAAAATTCTGGAGACCAGATGAGAC  
 AAACACATAATATGTGCCTGTATGGG  
 TCACAGTGCTAGGAAATACACTGAAAAGTGAACAAAGAACTAAAAATAAGCCTCTTATAAGAAAAGCTGCTACT  
 TCTGGGCTAATCTATTAGTAGTTTGT  
 TTTCTTAATTTGAAAAAGAGCTTCAGAATCAGACTCTAGAGTGAATCTTTTAAACTATATATATTATGCATGTA  
 TACACATATACATACTCTATGCAT  
 TGGCTTAGATAATATAAAAACAGCACAGGAATAGTTTATGCTGTAGTATTGTGCAGTCTTATTCAGAAATAAGAA  
 AAATAGTTTTTCCATTTAGCTGTGA  
 GTCATTACTTTTTAGTTTCGTGGAGAAATTAAGTTTTAGTGGGGGTCGATACAGCTTTTTTTTTTTGTAGAAAAT  
 CAGGTAGCATGGGCTTTGGAGTCA  
 GACAAATCTCGGTTTTAATTTTCAAGTGTCTTATCTCTAATATAGTAATATTAATAGTTACCTCTTCCTGGGTTGTT  
 GAGAGGTTTAAATGAAATAATAA  
 GTTAAGTACTGGCACAGTGCTTAACACATAATAGGTATTCAAGATATGTATACTTCTTTATCATCAGGGGATTGA  
 CCTTGAAGCATATGGGTACATAGA

AAGAACCCATAAAAGGATATGACAAGGGGAGAGGAGTTTGTTTTATTAGATTAGCAAATTCATTCTTTCCCC  
 AACTCCCCAGAAAAATTCATTTTAA  
 TGCTATTATTTTCCTTATCACCCATAAACAGACAGAAAAATTGGCATGTTTCGAAAGAGTGTTAAACTGCAAGAG  
 TATTACATGGTATCATTACACTACC  
 ATCCATTATCTAACTGGGTTTATTTATGAAGGTTAGATTTAGATAATTTAGACTTACCCTTTTTCCATTTTGTGGGG  
 AAAGAGAATCAGAACTTGAAAG  
 GAAGAGCTCTTCCAAACAGAAAGATAGCTAGAATCTCTCTCAATAGTCTATATAATCTCTTCTAATAATCTTTAGC  
 AAAAGTGACTTTTGAGTTCTTTTA  
 AATATTTCTGTATTGATTCTAATTCAGTGCTCATTTTCATGTGATGTCTATAGAACTACAGACTTAGATTACAGTCA  
 GAAACTGAGTAACAGCAGACAC  
 CAGGTTTTGAGCTGTTTTTAGGGTTTATTCAGATTGACTACATCACATTCATTTGGATGGATTTAGTGACATCAAT  
 TGATAATTGTAAAGAACTAGCCT  
 TTAATCTGACATTTTCATATTGATATTAGATAGGAGCCACATGACTATGTCATGGGTTCATTCACAAGATATAATA  
 AACATTATTTTCACTTGTAGCATA  
 AAAGCAGTCTTTAGTGTTATCTTAGAAGCATTAAACACACACATATATAAAATCATATAGATTAAAAAATTCTAT  
 CAGCATAGGTTCTGTAATTGTAA  
 TAGATGTAATTGCATAAGGTCGCTATGCATTATTTCTCATTTTCAGATGAAGCCTGTCACCTTGACATTTTTATATA  
 TAATAGGAATCCACAGGTTGTTA  
 AAAACATAAGGAAATTAAGATAATTTGAATTCAGATTATAAAATAGTTCAGGTCTATTCTGATGAAGCAATGTCA  
 CTTAGGTCTGAGGCTTTTGGGTCAG  
 CTTTTAAAGTTTATCTTATTTTTCATTAAAATATGTATGTATAATACTCTATCATGCAAAGAGAAATGGTCAACAG  
 AATTTCTGTGCATATATGGTTGC  
 TGGTTATTTGAAAGCCATTCTCCTCCTAGGATTATTTGACCAATTTTGTATTCAGCTTAGATTAGGGAGAATATAG  
 TGTCTCCTTAGGTAATACTTAC  
 TAGAACCAAAAAGTCCAGATTAAATTGTTTTTATTTTATTTTACTGCAGTTTTGTGCAGTTGACCTGATTTTAAAA  
 TATAAACTAGGAAGGCATATTT  
 TTCATCTACCAAAATGTGAATCTTACTTTTCAATTTGAAACAAACACCTCTGCTTTAAGAAGATGAACGGTAGTGAT  
 AAATACAAAAACGTTAAAATTTT  
 TAAATGAATAGCGTTTGTTGATCTTTAAAGAAAACATAAATGTGAAATAACTTTAGAGCTTAAAAGTTTTCTTCCT  
 CAAACCTTTTAGGAATTGATTTTA  
 TTATCCAAAACAACCTTTATTCATACTAAGTGTGACTATCTGAATAATTACCTTACTTTTATAGTTTTTTTTTAAAAA  
 AGCTCTTCTAACTGTAATTATC  
 CTCAAAGAATTGGATTAAGAAATTCATAAATTTTCTGTCATTAGTCCAGCGAGATATTTACTGAATAGCCATTACG  
 TTCCTGGCTGTTAACTACAAGG  
 ACTCCAGCAATGAGCAAAGCAAACAAGCCCTTGCTTTTCCAGAGCCCACATTCTAGAGTCAGATGCTAAACAAAT  
 ACACAGATACATATATAATAAAATA  
 TTGGTTTAAGTGTTGAAAAGCAAATCAGAGTAAGAAGGCAGGTAGTCATTTTAAATTAACATTGAGAAATGTTT  
 CCATTTTAAGCATATAAATCTTTG  
 GCTTCTCTTTTTAAATTTGCTTTTGACTGTTTGTTGCTGGCGAAAATTATTTAAAGTAATTAGGAACTCATCGTAAA  
 TATTTGTACACTACTCTAAAAA  
 TCTAAYTCCAACCTAGTGAATTTAGCATTATATAAATGGAGGTGTCTATAGAGGAATAAAATCTTTGGTTAACCCTA  
 CTTTACTTTTAGCAGTAAAGTCTG  
 CAGCTCATGAAATTTGTAATTCTGGAGTGGTTTCAGTTTGCACCATCCCCTGGGGAGCCTGTTGATACAGCACTG  
 AGCTACAGATTATTATTCAGCTGTA

GAAGTTTTATGCTTAGGCATTTTCTTACATCTTCTCTCTCAGTCAGTTGAAATCTTGTATTTTTGCCAAATAAAAC  
 AAATCCTATATTAATAAGAAGCT  
 ATAAATAAACAGTAGGAGCATGAGTCAGAGCCAGCATCACCCATTCTTGTGTGTACCTCTCTAGTACTCCCA  
 GTATGTTCTTTTGCACAAACATGT  
 AGTACTGCGTAACTACTGTTCTACAGATGAACTCCAGTTATTTAAGTGAGGCCTGTTTTTTAGAACTAAGTAATAC  
 TGTAAC TAGTAGAATTCTGGCATG  
 TTACCACAGTCAAAAGCAGAATGTGAAGCTGTATGCTATTGATGGGAGGAAAAAAGAAATATATTTCTACCTTT  
 CCTGTATAAAGCCCAGAAAGATTCTT  
 GTTAGATTTCCAAAATGGGTTTTTATATAATATAATATAATTCAACATTTTTTTTCAGAGGAAAAAAGAAAAAGATA  
 AGTTGCTTGAAAGTCCCATCTCTT  
 TTGCTAGTCGGCTTCTGTCTGATAAAATTTTATTCTGTTTTGCTTCATTGTTTATTCATT CAGTAGATGTT CATTGAA  
 CACTTACTGTTACCAGGCACTG  
 TGCTACCATAGACGGTATAAAGACAAATAAGCTAAAATCCTTGTTCTCAAGGAATTTATTATTGTAGAGAAAAGT  
 GACTTGTAATCCAAAATTT CATA C  
 TGGGATAGGAACTATAAAAGAGACATCCTGCAGAATACTGAGATTATAAATAAACTAGTATGTTTGAAGCATT  
 TAGTGTCCTAGCTGTTTTATAGAAGT  
 CATTTCAATTAATCCTTTCAACAGCTTTATGAGGTAGATATCATTACACTATTTTACTTGTGCCAAGAGCACTACAT  
 TCATCAAGGAAGGAGGAGGTAGT  
 ATCAAATTTTTTCAGAAAACAAGAGAGTTGGTTATAGAAGGATGAGTAGGGTAGAGAGTAGGAGGAAGATGAG  
 GCAGTGTGGCATTTCAGACAGAGAGGTG  
 AATATGTAAGCACACAGGATGAGTGAGAGAATTGTTGGAATGTAGTCAGTTCTAGGAAATAAAAGATGGCAGG  
 GTCCAGATCATGAAGTTTGGTTCTTAA  
 CCTGTAGAAGACAGAAATATGATCAGACTTGCATTTCAGGAAGATAACTCTGTAATTGTTTGCAATTCCTGTAA  
 ACGAAGAAATTTTCTCCAGCATTGA  
 CTTATGTTTATTAATACATATAAAAACTAAATTCTAGAAACATCTTATACAGTTATACATCACTTAGCAATGGGA  
 TACATTCTAGGAAATGCATTGTCA  
 GGCAATTTCTTCATTGTACAAACGTCATAGAGTATACTTACACAAACCTAGATGTATATCTTTGTTTATATTTATT  
 TTT CATATGGAAAACCAAATTT C  
 TTAGCCATGTTACTGAATGTCAGTCATTTCCCTTACTTGATCTGCATCACCAACATTAAGTGCCATGTATCAGATT  
 CTATATATGTTCCATTATAATCT  
 CATGGGACCACCATAGTACATGCGGTCCATTGTTGACAGACACATCATTATGTGGTGCGTAAGTGTACTTATAGT  
 TTTACAAC TACAAAGTGAAGCACAG  
 AGAGGTTACATAACTTAGCCAGTGTCACACAGCTAATAAGTAGCAAAGTCAGGATT CGACATGTTAAGCCAAC TT  
 CATGGACTTAACCACAGTCTCTGTAT  
 TGCCACTCCTCTTCATTTAAAGGCATGCATGCACATATTAGATTTGGAGACAGTTCTAGAGACACTGAATTGGCA  
 ACTACAGTCGACAAAGAGACATTGC  
 AGCAAGTACCTCTACGTTGTTTGTCTGTACCCTTAACTTGCAAGTGGTCTACAAGAAAGAATGAAGGAAGGGAA  
 GTGATATATTTGTCAACCCTAACTA  
 CATTGTAAAATGAAGCTTTTTTTTTTTCTTTTTTTGAGAGGGAGTCTCGCTTTGTACCCATACTGGAGTGCAGTG  
 GTGCGATCTCAGCTCACTGCAACC  
 TCCACCTCCCGGGTTCAATCTGTTCTCCTGCCTAAGCCTCCCGAGTGGCTGGAATTACAGGCATATGCCACCACAC  
 CTGGCTAATTTTTGTGCTTTTTT  
 TTTTAGTAGAGACGGGGTTT CACCATGTTGGCCAGTCTGGTCTCGAGCTCCTGACCTCAACTGATCCGCCTGCCT  
 TGGCCTCMTAAAGTGCTGGGATTA

CAGGCTTGAGCCACTGCACTCGGCCTGAAGCATTTTATGATGCTCACTACCAAGTGTTTCATAAGCGGTAAATAAA  
 GCCCTTGGGCTAGGTAATGTAAGAT  
 CCATTGCAACCTGGAGATGAAGAACTTTGTAAAGGGGATAATTGTTATCACCCATCAGAGCATCAGAGTTTTAAG  
 ATGATCTAGATTTAGTCCCACTAAT  
 TATATTTCTCCTAACTTTTATAGACAGAGAAATCAGTGAAAAATATGTATTTCTAAGGATTAAATCCCAATCCATA  
 GCTCAAGGCTACCTTTATGTCTAA  
 TAGTTGTGATATAATATGTTGCTTAGAATAGGTTCTATAAATGGAATTCACACATAGGACCCCAGAATCGCTTCTT  
 CACAGGGTTTCAGAGTTGGAGGAC  
 ATATAAAAGGATAAATCAATCCATAAATCTCTAAGTATTCCTAAAAGAAGAGTGGTGTGTCAGTTATTGGCACCA  
 GTACATCATGAATCCTGCTGGGCTG  
 TGGAGAAAGGAGAACTAACCTGAACAAGACTATACTCTACACAGCTGTCCCTGACATTCTACCACCTGTTAGCA  
 TTCAGTATTAATTTTCATGGATCAAA  
 TCAAGGCATGCAAGTGCAATTAATAACTTTTTAAAAACAACCTGCTGATTGCTGACGTCTATGAGGGAAGAAAA  
 ATTAGGGGCTGTTATTTTCCTATCAA  
 CCATCACTCTCATTAAACAAGAACAGGAAACATTACGAATAAATAAACTCTTACATCATACCAGAAAAAAGAACA  
 GCTGAATCCCCTAAAGGATTTTCATA  
 TTCTATCATCAGACAACGAAAGAAAATAATTCCATATAAATGGCTGCATGAATCATTGAAAGAACTTCCCTGCT  
 TTTTAAAGAGCAACTAGAAGTAT  
 TTAATTTTGAATAAGAGTAACCATTTGTAAACCGTATATACCATTGAAGATGTTTTGGGTATAAAAGGGAATT  
 TCCATAATAATTACATAGTAATCT  
 TTTATAGCCAAAGTTATGAAGAAGATCAGGATTTGGGAAAATTGTGGTCAATATTTGATTTTCCTTTCTGTTTGA  
 TAGCCTTATCTAAATTCCCCTAGC  
 TTGAATGTAGTAAATATGAGTGTGTAAAAATGTGTATCTAATTCACCTCTTGGGGGAGGTTTAATGTAACCTATAA  
 AAAAAATTCTTTCCCAGGTTAAA  
 AAAAAAGTTACATAGTGTTTAATTCTACCTTTTCTCATTTCTCTATATCCAAAACCTATCACCAAGTACTGTTAATTTT  
 ACTTTATAAAGCTCTCTTGATA  
 TGTCTCTTCAACCTCATCACCCTGCTTTAAATGAGGCATACTTCCTCATTTGGATTATGGCAGCAGTCTCTCTTCC  
 TTAGTCTCTCCTTCAAATCCAC  
 CTCCATCCTTCATATCATCCATCATGTATTTTCCCTGCCTAAATTCCTTGGTTGGCTTCTATTAAATCCAAGTTCCT  
 TAAGCATGGCAAACAGGGGCCA  
 GCTTCCACGAGCTGGTCCATACCTCTCCTGCCTTATTTCTGACACTTTCCCTTATACCTTGAGATCCAGCTTGCCA  
 GTATTTCCCAATACCTCCACCG  
 CATATACAACCACATACATATACACCCTTATCCCCTCCCTCATGTACCACCCATATATAATTCTCTTTACCGCATCCT  
 TTATCACACTCCTACAAGCCTT  
 GAAGCAATCTGAACAYAGAAACCATCTTTGATTTTCAACTCCTAATTGGGACTGAACATACTTGTCTTGAACACAG  
 ATTGCTGCCTTGTTAGTATAGGTA  
 TCCTAATATGGCACATGTGTATAGTTGAAGCTATAATGCATTCCCCAAGTTCCCTTTCAGGACTAAAGGGTTATTT  
 CCCAACAGCTGGAAGTCCTGCTGC  
 CAGACAGCCTTCTTCAGGAATTGCCTTGTCCCCAACTCAAGACAACCCTGAAGGACCTTTCCCAGTTTTCCATGGG  
 GTCCACTGAGGCCTTGAGATTGTA  
 CTGAAGTCAACTTCTCCCTCTGCTGTGCTTCCCTTCCACAGGTGTTTATCACAAGAACAATCCTTAATAAACTTCC  
 TGTAAGCACACTCTCAGAGTCAG  
 CTTGCTGGGGAACCCAAGCTATAACAACATGTTTATCTTGTTACACTTTCTTCTGAGTTCTAAATCATGGGAAGT  
 TACAGTTATCCAAAATCTTACTAA

TTGGAAGATCATTTTCATCAGACGTGTATTGAGCACCTGCTTCATGTCAGGCATTGTGCCAGATACAATGGAAACA  
 AATCAAGACTTGGCCTCTGTCCTTG  
 AGTTTACAGTCTGAGGTGACAGTTACATATGAACCTAACTGTATTAAAGCATCTTGAGAATCTTTTTGATTCAAAT  
 GTCAATAATTTTAGTATAAAATTC  
 AAAAGCATGGATTTTGTGGTGTACAAAGAACAAAAGGAAAATATTTTCAGGGGAAGGAAAATGAATATAAATA  
 AGTAGAATGAAATTTTAGAAGAGCACA  
 GAATTTGTCTTTATAATTCTGAAAGCCTTTAGGGAAAAGGTAATTTTTTCACTTTTTATTTTGACATAATTTCAAAC  
 TTACAGAAAAGTCACATGAATAC  
 TGAACCTCCCATACACCCTTTATGCAGATATATCAGTTTTTAACATTTTCCCTCATTCTGTTTCTTTACACACATACAA  
 ATTATTTTTGTGAACTCTCTGA  
 GAACAGGTAAATACTACATTATATCACTTCATACCTTAAAAATGTAGTATTTCTAGAAACAAGGGTATTCACCT  
 CCATAACCACAATGCAGTTGTCAA  
 ATTCAGGTAATTTGACCTTGATCTTTAATCTACAGTCCAGATTGAGATTTTGTTAATTGTCCCCAAAAGATGAGT  
 TTAGGAAGAGTCATATAACTTAAC  
 ATTTGTAATGTAAATTAAGATCTCGGAAGTTTGCTAACCTGATTAAATACCACAGTGAAATGAACCACTTCTGA  
 TTTTCTACGGAGAAAGCAAGGGTT  
 AGGTCCAGGTTACCTGTAAAGGTGTGACTTAGGGCTCCCCAGAAGTGAACCTTGCTCTGGAGACCCCATTTGGT  
 GAGGTTCTTTTCAAAGAGATTGTAT  
 ATTTTGTACCCAGGGATTGTGATTTGTTGAGCTCCCATGTAAGTATGCCGCCTTGCCATTGACAGGTCTTCACAA  
 TGCTTTCTTAAACGACATTCTTG  
 GAACTGTCTTTCATAGTCAGGCGACCCGCCAGCCAGGCCGCCAGGTAGGTGGAAGGGCGGCTTCGGGAACTAG  
 ATCCGGCTGGTTCTCTGCTAACATGTG  
 GTCCTCGATGCTCTAGCGGATTGGCCCCAAACCTCAGCACACAGAGCTTTGGTAGGGCTGGGGATGCMGGTCA  
 GCCTGCTGAAGGCGTCTAGAGGACTTG  
 GCGGGGGCGTGACTTCCTGCTGCACCACTTCCTTTCCAGCCACGTGGGCTTCTGGGACCCTATGACATAATTC  
 TGGAGTGGTCTGTTGGGAGTTGTGG  
 CGGCGGCGAGGCCCAAGCACAGCTGGCCAAGGCGGGTGTAACGCGGGTCGTGTCCACCCCAAAGCTCGACG  
 CCTCTAGAAGCCATGGCTTTTCTGCTG  
 AAGCTAGAGCCGGACATGCAGGGGAAAGTAGATTTGTTTCGAGCCCGTATCGCCCAGGAGGCCGAGGATCTCG  
 TGTCCACCTTCTTTCCTCAGAAGGTCT  
 TGGAGCTGAATAGCCGAGTTCAGGAGCTCCGGCTGCAAGACCTGTCCAGAATCCATTCGGTGCCAACCTCGGAG  
 CCCCTGGCCACCCAGGCAACAAGGG  
 AGATGGGCCCAACCAAAATCTGCCAGTTCTGCTGACTCAGTTCTCCATCAAGGTCCCAGCACTGCTTGGCAGCGA  
 GGGACAGCTTCTGCGGAGCAACCAG  
 CATCTGGTGGAGTTGACTGAGTGGGTGAAACCTGAGATCAAGCTGCTGAGAGAGAAATGCAACACAGTGCACA  
 TGTAGGTGCACCCGCTTATCCTGAAGG  
 TGSAGGACGGCAACAACCTCAGGGTGTCTATCCAGGAAGACACTGTGGACCAGCTGTGGACCGTGGATAGCAC  
 AGCGGCCTCCTCTCTATGCGGCTTCTC  
 CACCTACTACAACCCAGGCCAAGTTGGTATCCAAGATAGTGAAGTACCCCAAGTGGAAGATTATCGCTGCA  
 CTGTAGCAGAGGTGACGAAAACGAG  
 TACCTGAGTGTGCGCCAGATCCTGCTGCACCTACGGAACGAGTATGCCACCTTGCGTGATGTAATCCTTAAAAAC  
 ATTGAGAAGATCAAGATCCCTCGGA  
 GCACCAACCGTGATAACCTATACTGATAGCCTTTCTCCACCCCATTTGCCTTCAGGAGTTGGCGGAGCTGTTTAG  
 GCACCTAGAAGGCTGGTTATTAATG

ACTCTACAGGTCATGAAATTACTGGAAATGTGTGTTACCATTTAGAATTCATACAGCTTATCCTCATACTGAATC  
 ATCTTTTGGTGTAATTTTATTTG  
 ATTTTCCTTTTGTGAGGAGAATGGGGACAAACCATCAAAAAGAAAAGACTGCCATTTTAAAAGCAGCTTTATAG  
 ATAAATGTTGGCCCAAGGTTTCTTT  
 ATTATGGCCTTAGCGACATTGAATGGTTTTAGGGGCTACTTAAAACTAAATCTTAAAGTTAAATGAAGTTTTTCC  
 TAAAGTTCTGCTATCCGTTGAACT  
 TTGTTCAAACCTTACAGAAATGCTCCTTTTACAGCATCATTTTTCTTTTTTTGCGAGGAACGGAGTTTAAAGGGATA  
 TGTTTAGGTCATTGTATGCCTCCA  
 GTTTAACAGATTGAAATAAACTTGGTTATAAAATTTTGTGATGGCCTGAGTGTGCTTATCTGAGCTTTAATTTTC  
 RGATGTGGACAATTCAGATGATAA  
 TTTCTACTTCAGGTGGACTCATACACTTGATATTTTTCAACTCTGATGGGTAAGAAAAGTAAGCAAAGAGTAAT  
 GTATTTTAATTATTCTAAACAAATG  
 GACATCTACTTCCCTTAGAATGAAAAGAGGAATCTGTGGAAAATGACTTCCTTCCATAATCTTCTATCATCTCAGG  
 ATCTTCATTTCTTTTTACACAATT  
 TAGAATTCCTTCTGTATATCCGTTTGTCTCTTTAAGTCTCTGCTAAATCCATCAAGCTGTGTATCCTATAATACTTTT  
 TCCCCCAGTTGATTTTCATTGC  
 CCTTTCTTTTTGTCTCTTAGATATTTCTTAAAGTCCCTCATTAGCATTTCCTCTCAGGTTATACAGGTCATTCTTTTT  
 TTTTTTTTTTTTTGAGATGG  
 AGTCGCGCTCTGTCATCCAGGCTGGAGTGTGGAGTGCAGTGGCGTGATCTCAGCTCACTGCAGCCTCTGCCTCCC  
 AGGTTCAAGAGATTCTCCTGCCTCA  
 GCCTTCCTTAGTACCTGGGACTACAGGCGCATGCCACCACACCCAGCTAATTTTTGTATTTTAGTAGAGACGGG  
 GTTTCGCCATGTTGGCCAGGCTGGT  
 CTCGAACTCCTGACCTCAGGTGATCCACCCGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGC  
 GCCTGGCGGGTCATTCTTTAAAGAC  
 YCTAAAATGTGGCTTTGTGTTCTTTTGCTACGTCGTTGAGGGGAAGCCATTGTCTTCTCAACAGTAAGCTATATGG  
 TTGGTATTTTTAAAGTTCAGCATT  
 TCCAGATTTCAATTTGCAGGTAAGATAAATTACTGCACTGTGAGTCTATTCTAACCTATGTAGATTAGAAAAATCAT  
 ACACTGGTTTAGCGGAAACTTTAT  
 TTTCTTTCTTTCTTTCTTTCTTTTTTTTGGAGATGGAGTTTCGCTCTTGTTGACCAGGCTGGAGTGCAGTGGTGCGA  
 TCTCGACTCACTGCAGCCTCCAC  
 CTCCCAGGTTCAAGGGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCATGTGACACCATGCCCAG  
 CTAATTTTGTATTTTAGTAGAGAC  
 GGGGTTTCTCCATGTTGGTCAGGCTGGTCTCAAACCTCCTGACCTCAGGTGATCCGCCAGCCTCAGCCTCCTAAAG  
 TGCTGGGATTACAGGCGTGAGCCAC  
 TGCGCCCCGCCACGGAACTTTATTTTCAAAGTAACTCCTGAAAAATGATAACAAGTGAACTATTCTTTGTGCG  
 ACCTTCTGAAAAGCCTGAGTCTT  
 CAAATCTGTGCTTAAAAAGCAACAACAGCAAAAGGTATTGCCTAATAGAAAGACGTCTTACTTCCTTCCCCTTCA  
 AGTAAAAAGAGTTCTTTTTATTTT  
 CCTGTAAGGTGTTCTATCTGATTCAGAAAAGTGCTGGGAACCCTAAATCAGTTCATACCATTTGACAGGCTTTAAT  
 TGCACACAAGTGTATAATGGACAA  
 TGTAAGGGAAGAACAATGTTGCTTAATTTTCAACTAATKTTAATCATCTAAGACTTAACATATCTGTTCTCAC  
 ATAATATAATCATCTTTGCTTGT  
 TTCTTTTACTCAAACCTGTCAAGCTTTATGACAGATTGAAATAAATTTTCTCCAATCCCTGGCATGCCTTTTACTGAG  
 AAAATATGTCCTTTGTAGTTAAA

GATTTCTGGTATATGCATTTTTACCTTTTTCCAACCTCTGTATCAAATTAAGTTGGCTCTGGAATACTTTATAAAGT  
 CCCTACTTCTGCAAAAATTAATC  
 TCTTTGTGATTTTCCCCTTTCTTGCAAATGGTTGGAAAACTTTTATAATAATRAACTTAGGAACTGGTGTTACAAT  
 AYGGTATCCAATGTGAATCATT  
 GCCTGTTATTAGTACTTGTACTTGAAGCCACGTTTTTGTTCCTTTTAAGATGGCAGCTGAATTGAGCAGGTGTTT  
 TTTTTTTTTTTTTTAAGAAATGGT  
 TTTAAAACTTATTTTGGAATTAAGTAATTTTAGTTTTAAATTCTAAAACATAAAGACACTAAAATTTG  
 ATTGGAAGTTGAAATTAGTAGATC  
 TAATTTTTATGATTGAGTGAAATGTCTTAAACCACATAAATGATGGAGAAGAGTCTTAAAGTTCTTAACTTCAAA  
 CTTCTTAAACGTTTGGGCTTAGTT  
 TTAATAAATTACTGAATCGTTTATAAATTTTTCTTTGGTTATTAGAGGCAACATTTGTTTTGTAGTCTTTTTTTTTT  
 TTTTTTTTTTTTTTTTTT  
 AGACAGAGTTTCACTCTGTCACCCAGGCTGGAGTGTAGTGGCGTGATCTTGGCTCACTGCATCCTCCACCTACCA  
 GGTAAAGCAATTCTCTGCCTCAGC  
 CTCCCGAGTAGCTGGGATTACAGGCGCCCACTAAGCCGGGCTAATTTTTTTTTTTTTTTTTTTGTATTTTAG  
 CAGAGACGGGGTTTCACCATCTC  
 AGCCAGGCTGACCGGGAACCTTTGACCTCGTGAGCCACCCGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCG  
 TGAGCTGCTGTGCCTGTCCTATTTG  
 TAGTCTTTATGAGGTTTACATATTTATATAGAATTCTTACCAAGTAGTCTCAACCATTAATAGATAACTGTGTGATC  
 AATTTAAGTGATTTTTAATTTCA  
 GGCCAGATGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGGGGCCAAGGTGGGCAGAACACTTGAAGTCA  
 GTTCAAGAACAGCCTGGCCAACATAGT  
 GAAACCCTGTCTCAAGCAAAAATTACCAAAGTTAGCTGGGCATGGTGGTGTGTACCTGTAGTCCCAGCTACTCTG  
 GAGGCTGAGGTGGGAGAATTGCTTG  
 AACCAGGAGGCAGAAGTTGTAGTAAGCTGAGATCATGCCACTGCACTCCAGCTTGGCTGACAGAGTGAGAAC  
 CTGTGTCAAAAAAATAAATAAATAAAA  
 AGCTTTTTAATTTCAAAGTACCATCTCAATGGTCAGAAATTTGAAAAGTATAAAAGGAGAAATCACCTATGGTT  
 TACTAGCCTAACAAAATCATTACTA  
 TTTGTGTGTGTTTTACTTATTTTAAAGGATGCATATATGTTTTTACATCATTGCAATATAGTATATGCCTACTTTTAT  
 AGAGGATAAACATCTGCATTTA  
 AATCCAATCTAATATTATAACTTCGTTTGATAGATATCCCCTGATTCAGATAACAGATACATCCATTCCCCCAAATA  
 TTTTAAAGGCTCTGTTTCCTAAT  
 TGTGAAAATGATACTTGTCTGAGGCAGGAATTGTGATTAGAGTATGGAACATAGAGATCAAGTAGCTGGGAGG  
 AAATGTGCCTATATGACTGTGAGAATG  
 TCTAGATATTTAGAATAGGGTTTTACTCTGCTTTTAGGTTGGGGCTTCATAATAGCTTTAGCAACATATCTGATCA  
 TTTTATTACTCAGTGCTTTTTCCC  
 TCCCCCTAGAGCATTTAATTACCTCTAAAAATATGTACAAGAAAGCATTGCCTTATACTGATTAGATGAACTG  
 TTTCAGCTTTCATGATTTGGTATG  
 AGTAGGAGGGAATTCTACTTCTCTGTCAGCACTTAAACATTTCTGGAAAATGCTGTAGTGCATTGACATCAGGC  
 ATTTTAAAAAATTACTGATCTTTG  
 AGAAAAAATAAAGGACMCAGTGAGTTACAGAAATAGATGCTGGACCATCCCTGTTCTCTGCATTAGAAAACTAT  
 ACTGCTAACACATAGAAATAACTGCT  
 CACCTTGAAGTGAATTATCCATGGGAAGCCAAATACATCCGTCTATCTATCCATCAATCGATAGATAGATGTATT  
 TGTCTCATTCTGTGCGCCAGGCTG

GAGTGCAGTGGCACAATCTCGGCTCACTGCAACCTCTACCTCCCGGTTTCAAGTGATTCTCCACCTCAGCCTGCC  
 AAGTAGCTGGGACTACAGGTGTGT  
 GCCACAACCTTTTGTAAATTTTGATAGAAACAGGGTTTCTTCACATTGGCCAGGCTGGTCTTGAACCTCCTGACCTCAA  
 GTGATCTGCCTGCCTCGGCCTCCC  
 AAAGTGCTGAAATTACAGGCGTGAGCCACCATGCCCAGCCTCCATATTGGATGATTATACATTTGTATATATTTTT  
 CTTTGCTTGCTTTATTTTGATGAG  
 TCTTCAAGTCAGTAGCAGAATTTTATTAGTGAAGTGTAGTAATCCCAGTTACTCAAATTTAAAATGTTTCACTAGT  
 AAAGTGAATTTAACGAATTCTCTT  
 TCATAGAGAATGTTGTGTTATTTTTMCTTATTGGTGAAAACCTAGCTGGTGCAGATGAACAGGAATCGAACCTAT  
 TATCAGTAAAATTAGATTCCTTTA  
 CAATTTGGCCATTTATTTTAGCTTTATTATTTAATCTCATTTTTACAAAGTTATGCAATAGTATATTGGAAAAATAA  
 TATTGCTTTGAAAGTGTGAAGTA  
 TTTAGAATTTCTTAAATAAATATATAGTTTCCAGTTTAGATGTTTGAGATTTAGAGGTGTACTTTTAGCCCCTTGTT  
 TTATAGATGAAGAAATTTAGAAT  
 CAGAGATATATATGGCAGCACACAACCTCATCTGTGACAAAACAAGGTCTAGAAGTCAAATCATCTTCCATTTTCAG  
 GCTAGTGCACCTTTTCACTGCACTTT  
 ACTATGTTCTAATCAGTAAATAAAATTACATTAGAAGAGAGTTCTTTCAGAAATACGCAAACCAGTAAAATTTCA  
 ACAAATTAACATCCCTTTCAAGTTG  
 TGTGTTATTTTTTTCTGCAAAATGTAAGAGAAGTTGAAATGTTTCCAACCTCAGAATGGACTATAATATTTTTACTG  
 AATTAAATATTATGTAAAGAGCTG  
 GGCTCGGTGGCACACACCTGTAATCCTAGCACTTTGGGAGGCTGAAGCGGGCGGCTTGCCTGAGCTCAGGAGTT  
 TGAGACCAGCCTGGGCAACACGATGA  
 AACTCCATCTCTACTAAAATACAAAAAGTTAGCCGGGCATGGTGGCATGCACCTGTAGTCCCAGCTACTCACAGG  
 CTGAGGCGGGGGTGGGTGGATCACT  
 TGAGCCTGGGGAAACAGCGATTTAGTGAGTCAAGGTGCCATTGCACTCCAGCCTGGGAAACAGAGCGAGACC  
 CTGTCTCAAAAAATAAAATAAAATAAAA  
 ATATTCTACTGGCCAACCTCCTCCCCACCTTCTAACCTTCAGCACCAAAGAACTGCTACATAGATTTGTCATGGCT  
 GCTGCTCCTAAAAAACCTACAGAA  
 TAAAAGGGGAAAGAAGAGCTGTAGTTTTCTGTTCGGGAGGAAGCCTGAAAGGGTCATTTCCCAAAGAAGTTT  
 ATCTTTTAGAGCTGGACTGGTCAGTAC  
 AGTAGTCACTAGTCTTGAAATGTAGCTAGTCCTAACTGAGATGTGCTGTAAGTGTAATAACACACTAGATTTTG  
 AAGACTTAGGGTGAAGAAAAAGTAA  
 AATATTTTAATAACGCTTATGTTGCTTACGTGTTGAGTGATAATATTTTGGATACATTAAATAAAGTAGATTATTA  
 AAATTAATTTATATGTTTCCTT  
 TACTTTTTTATTTGTGTGCCCGAAATTTTAAGGTTATATATGTGGCTTGCTTATATGCCTCACAGTATATTTCTATT  
 AAACAACACTGCTCTGGAACCTA  
 AGATTTCTAAGGTAACTGAGATGATCCTTGAGTTGAGCCTAAACCTGCTCTTGTGGTATATTTTGGAGATGTGG  
 GTTCCATTAAGTATAGATTTTTT  
 GACTCCTAGTAAATTTTTATTGATTCGTCATCTGGTTCATAATGTACAATTCAGTAAAAGCCATTCAGAGAGAGGC  
 TAAACATTGTAAGCCAAACCTTCT  
 TACAGTGCAACTTGGTCTTGGAACCTAGAATGCTTACAGGAATGGATGGCTAGCATGTGCTTTATGCAAGCCTTT  
 TAAATATCATATCCAGGCTTTTGGT  
 GAGAGACTAGAGTACTAATATTCTGTATTGCTAATTCAGGGCTGCTCCATATGCTTTATAAACCAAATAGAGGGT  
 AGATTTAGCATGAATATTGAGACAT

TTACCCAGTGTTTTATCTGGATGAAAGAATAACGTACCTCTGGTGCCTCGAGAGTACCTATAGGCAGTTAGGCT  
 TTTTCTCAAGGAGCAATATTTGAAC  
 ACTCAATTAAATCCCATGATTTTATTGAATTCCCATATATAGTTATATAGTACTCATTGTATTCTGTGTGTGTGT  
 GTTTTTTTTGTTTTGTTTTGT  
 TTTTTTTTGAGACAGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACCTC  
 CACCTCCTGGGTTCAAGCAATTCT  
 CCTGCCTTAGCCTCCCAAGTAGCTGGGACTATAGGCGCACGCCACCACGCCAGCTAATTTTTGTATTTTAGTAG  
 AGACAGGGTTTTACCACGTTGGCC  
 AGACCGGTCTTGAACCTCTGACCTCAGGTGATCCTCCACCTCGGCCTCCCAAGGTGCTGGGATTACAGGTGTGA  
 GCCACCGTGCCAGGCCTGTATTCTG  
 TATTTTATGTATATCCTCACTAAATAAGTTTCTTGAGAATAGTGTAGGGTTTTTTATTACCCCTTTTCCATTATGG  
 TGTTTTGCATATGGTAAAATTC  
 TAAAGTATTCTGAATTAAATTGAGCTAAACCAAACCTGGATGAACTAAAGAAAATCGTTTCTCCCCTTTCTTATT  
 AGAGGAGGCACTTTGTTACAACAG  
 AGGATATATGTGGATAAGTGTACCCAGGAATATGTAACAATGTTGAGACATCTACAAATCTGATTCAGAATTAAA  
 CAAAGGCCAATGAAGTTTCTACACT  
 GTTTTCTGGCTTCTAATTAAACATTTTTTTCATATCACAATAATATAAATATTTTAAAGTACTCCAGGCAGTTCTCCT  
 TCAGTATTCCCAGGATGTCATT  
 AAGCCTATTACCCTGCAACAGCTGCAAGTTATTCCAAACATTTGTACTGAATCTTTACACTCTACAAGCCCATT  
 GAAATCAACTGTGTCACCCTATTA  
 ACAGCTTCAGTAATTCTCTATTACTTTTCTAACCAAGAGTTAGGGAGAAATAAGTCTAGTATATCAGAATGAAGA  
 ATCCTGAAAATTGCTTTTCAAAACC  
 GATTCTCTGTGGTCACGAACATAATGTTTTATTACCCTTCCAGATTTTGGTACTTAAAAAGTTAGTTTCTTAAAT  
 GTGTAAAATTTAGATCATTTGAT  
 AAATGATTAAAAAGAATTTCCCTTAAGTTTATTAAGTCTTATTTGGTAGTAAATATAAAAGACAAAACATTCCATG  
 TAAAAATTATTTACATATTTTAA  
 ACAGTTTAAATAAAGTAAGTGTAGTCAACTCAAACTCCACTGCCTTATTCCTCTGGGTCTTGCATATTTATTGGT  
 GGAAATTTTTCTCTGCTATTTTT  
 TACATTTTACATTGTGGGATAGGCACAGAGTTATTGAACCTATGACCTATATATCCACAGCTTACTTCAATGCTT  
 CTATTAGGCACCCGTCCAGAGTTG  
 AAGAAACGTCTCAGATCACAGCAGGTTCTCAAAAAATGTCATTTTGTTAGCTTCATTTCAATATAACATTGATGA  
 GAAAAAAATCAGTTCCCAGCTGG  
 GACCACTGTCTGTGTGGAATTTTACGTTTTCTCCATGTGTCTGCACAGGTTTTCTCCAGTTTCCTCCCACATCCTT  
 ATAACTTTTGCTTCACAAACATT  
 TATTGACTTAACTTACATCACTACGACCACCATCACTTACTGATTCACCAAAAATTGCATAAATTATCTTACTTGTT  
 TTTATTAGTTTTTATTAAATGTA  
 TAGCTCACATTTATTTCAATATTTAATATTAGAAGTGCTTGGGGTCTTTATTTAGAAGTTTGGTGATGTTTTTATAG  
 CCAGAAATATGCTGTAGGAACGT  
 AACTCTCATTTATATCAATTAGGCTATCATAAACTGGTTTCCTTTATGTCATTTGCTTAAAGAATTTATGGACGA  
 CGTGAGGAATAGTACTCAACTTT  
 TGCTTATAAATAATTAACAGATGGCATGAAAAATAATTAACAGATGACATGAAAAGAGACCTCTTAGGAAGCAA  
 TTTTACGAGCCTGCTGAACTGGGAG  
 TTAACATTTTCAATAAGAAGCCATTAAGTCAAATATAATCTTTTGCATTTAACTCTACATCATTATATTACTAGAA  
 GTCTAAAAACCAAACAGTATGGT

AGGAAGAATAAAATGTTACTTATGAATATAATTTCTGATTATACTTTTAGAAAGAGGATTAAAAGTATTGACAAC  
 AAATGAAGTCAGAATAGCTTGATTA  
 AAGTAGTTTATTTAAAGTATCTCCTTAATGTTATAGTGTATGCTAATTATTTGCTTTGCATATGTACACATCATGAC  
 ATACATTTATGCATTAGCTGAGA  
 CATGTTCAATTCTGTAAGTTTTGTACAAAATAATGATCTTCAGTTACACAAGAACATTTAAAAGTTTTACATTTGAA  
 ATAGATAATTAAGCAATGCTGTG  
 TGAAACTGAAAAATGAAACTTTGATTAAAAAAAGTATTTTCAAATATAGGATAGTACATGTATATGAGAATACT  
 TTTTGGGTATAATCTTAGTAACAT  
 CATTATGGTGTATGTAAACATATGACATATGACATATAAAATTATAGTAGATAGATCACTGTGTTACCCACTTAC  
 ACATTCTTCTGTCGTATTAAGCAA  
 CAAACATTATTCTTATCTATAATAAAGCAGATATTTCTGTCTGTCATACATATGAAATATTTGTATGTATTTATTGG  
 AAGTAATTTTAAAGGCTAKTCTT  
 TTTTATAATTTCAACTTTTATTTTAGATTTGGGGTAGATGTGCAGATTTGTTACATTGGCATATTGGGTGATGCT  
 GAGGTTTAGGGTATGAATGATCCC  
 ATCACCCAATAATGAGTATAGTACCCAACCAGTTAGTTTTTCAACCCTTGAGTAGACATTGCTCCAAAGGAAGAT  
 ATACAAATGGCCAATAGCACATAAA  
 AATGTGCTCAATATCATTAGCCATTAGTAGAAATAAAAATTAAGAAATGTTTAAAATATTTAATGTTTTATTGAA  
 AAAAAAACAATTAATAATTTTAAA  
 TAACATTTTAAATGGAAAATAATTATATCTCCCAAAGCAAAAAAAAAAAAAAAAAATTAGTGAGAAGAGTAGCACTGTT  
 TTACAGTTTTGCAGATTTTAAATAT  
 GTGATTTAATGGAAGATAACAAGATTCTTATATATTTCTGTATTCAGTCTGTTGCAATATATTGTTTTAGTTGAAAT  
 ATATGAAGAAAATCTACCCTCAT  
 GCAGGTAAGTAGTTGCAAAGGGAAGGAGTACTTTAATAGCCTTTTCAAATAATTGTGAATTATTGTTTTGATACT  
 ACATAAAAACTCAGCAAGTGGAAT  
 TTCTTTAAAGATTAGTTGCACTATAGAATCTAAAGCCATATCAGTGATACTTTTACCCTCTGTAACATTAAAAATCC  
 ATTGATCTATCTTGCAGTTTGAA  
 TGGAACCTGTACCCATGCATGATTTTGTAACATCATTCAGTCTCATTTGGAAGATACTGGTTCCTAAGAGTTAG  
 GCAGATCTTCTAAATATGAACATA  
 TTTCATTATACAATGTCAAAAAATCACATTCATTAATATCACCACGAATCACATCAAGAAGGTCATACTCATGGTG  
 GTAGATACATGGTTTCCGAAATTC  
 GGATTTTCACTCAAAATCCCAAATTTTATAATTTGCAAATACTGTCAAGTATTTTCTTTGAAATATCAGGCTTACTG  
 CATTTTAAAAAGTGCTTAATGG  
 GTTACCCTTGCCTGAATAACAATAATTCGCCATTATTTCAAGTAAAAGTGGTATTGAAAAAAGTGGGTTGTTGAG  
 CTTGTAAATCATTACCCAAGTGCT  
 TTTTCTTGAGATAGCCATTATGCTTCAGTAGGCAGCAAAACAGTTTTTTATGTTCTTGCCCTTTCATCACAGAGAA  
 TGTTAATAATTTTACTGCTTCAA  
 CAAGGATATTCTAACTGAAATTGGCTTTAAAATATTTTATTCATTTATTCATTTGTTGGCTGATTCATTGTTACTG  
 TATGTGGCGGTGAAGAATACAGT  
 GACTGCTAGACTAGCTTGGTGCCACTACATAGATTTATACTAAGGCCTAGCAGTTTGGCTTTGCTTTTTCACCAAC  
 AGAGTAAATTTCAACAAAGTGCAA  
 AAAGGCTTGAAAATAGTTTGACCTGAAAGGGTCTTAGAGATTCTTGCCTCCTTAGGGAAAAAACTAGTTAAAAAT  
 CTATTATAATAGCTCTCCTTTAGCT  
 AGTTTACATTATAAATCGTTTGTGGTTACCTACTATATGAAAGGTTGTAGTAGATCCTGGGGATTACCAATGAA  
 CAAAACAGATGAAAATGTCTGCTGT

TATGGAGCTCATGGAGCTCATTTATAAAGGGTAACGATATATAATAGGTGTTAGAGAGAGAAAAGTAAAGCAGAG  
 AAAGGAAAGGAGGTGTGTTTGGTGGGA  
 ATGTGTGTGTACAATTTAACTAAGGTGGCTAGGAAAGACCTCACTGAGAAATTGACATTTGAATGAAGATCTGA  
 AGTGGGRGCAGGTAAAGGTGTGTGCC  
 TCTACAAAGGCCACAGCTGAGGCCRCCCTTCTTTATCTACCTCTTTTCATCTTTCCAACCTCCAGACCTTTTTTTCT  
 GCTGATAACTCATACCTTTTAA  
 TTAAACTTTAATTTAATTTAATTTAATTTAGTTTCTTCCAACCTCKGCTCTCCACCACTTCCTTGCCCCTAGCACACAC  
 ACAATTAATCAGCACTTACAAA  
 AACAAAAATTTACCTCTTTACAAAAGGTACAGGSATAGAGTAGAGGCTTTGAATACTTGTGCTTCAATTATTTTC  
 TCCTGTCATCTTTTATATTAAAGG  
 AAATAAAATATAGGGGTGTCTCCTTTTCCTGGTTGAATTGGTAGAGAATATCTATGGCAGTTAGTGCTAATGTTT  
 ATAAAAAGGAATAAACACAAGTGAG  
 ATCATGATGGAGAGAGACCAAATCAGTAGGGACCAAATTATAAAGAGATTTTCATTTCTAGGCTAAAGAGTTTGG  
 ATTGGTCATGTGTGCCTGTCATCCTT  
 TATTTCAATTCAGTAGTGGAACAATGGCTCTCAACTATGTGGTTGAGGGAATAAACTTAAACAACAAATCTAAATT  
 TCCTATATCAGTTGCATTTAACTGT  
 TGC GTTTCTCAGGGACTTCACTCTCCTCAAACCTTTTTGGTCCCTTTTTTTAAAAAAAATTATTGAACACTTATA  
 AGTGTTGAAATACTAACTTGTCT  
 AATCTTTATAACAATCGGTGCAATAGGTGTTGTTACTGTCCCTATTTTACTAAAGAGAGAAAATGAGGACACAGAT  
 TAAGGAACTTGCTCTATATCCTACA  
 GCTAGAAATTAAGTGATGGAAGCAGAATTCAAACCCAGGCAGTCTAGCTCCAGAGTCCATCCTTTTAATCTCTGT  
 CTGTGCCACTCTGGCCAGTGTTTGT  
 GAATGCAGAGTCCTGGAACATCTGCATCAGAATCACCAGGGATACTTGTTAMTAGTGAAAAATTCTGGCCAC  
 CTCAGACTTACCATGTAAGAAATTGT  
 TTTTCAAGAGTTGAGGAATGTGGGTATTTATATTATAAAATCCTGCAGGTGATTCTGATATGTATCTTTTGTTAAG  
 AACTTCTGTTTTAGAAAGGACCTAC  
 CATTGGCACTGTTTTTGCTTCTGATCCCAGCTTCCTATGATTAAATGCTTTCTGAAAGATACATAGTTAGTACAGA  
 ACTTAATACTATTATACTATTCTA  
 TACTATTTTACCCCAGTTTGACTTTTAAGCTGTTACCACCTGAAATTTAAATGAAATCCTTTTTAGAACAACTTG  
 ACCAAACAATTTAAAAATTGTAT  
 TTCCTACAAATGCATCATACCTTAAAAATGCACTTTAAAACTGAATGGAAGCCAGGTATGGTGACTCACGCCTGT  
 AATCCCAGCACTTTGGGAGGCTGAR  
 GTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATAGCGAAGCACTGTCTCTACTAAAAA  
 TACAAAAATTAGCTGGGCATGGTGGC  
 GGGTGCATGTAATCCTAGCTACTCAGGAGGCTGAGGCAGAAGAATCACTTGAACCCAGGAGGCAGAGTTTGCA  
 GTGAGCTGAGATTGCGCTACTGTACTC  
 CAGCCTGGGTGACAGAATGAGACTGTATAAAACAAAACAAAACAAAAACAAAAACAAAAACAAAAACA  
 CACACACACACACACACAAAACTGA  
 ATGGAACTGAGAACTTACTGTTCTATTTATGATGAAAGAAAATGTTAGGCTTTAGAAAGTACACCCTATGTGA  
 TATGAAAAGTGAACCTTCTAGATT  
 GTGGAACAGCTTTATATAGTATCATACCTCCATCTTTAACTAGGATTGCAGCTGGCAATAAAGCTCAGCTTCTAAT  
 GCGAGCAAGTGTTGAGTTTGGCTA  
 TTTTGTTAACCAAGATCCTGAGAGAAAATTTGTTCACTGATGTACATAAAGCATATTATGGTTAAAATTTCTTAT  
 TCCACTTTGTAGTTTGCCTTGTTA

AGACTAGTCTTTCCTAAAATGTGCTTTAGGGGCATTTTATAGTCCTCCATATTTCTTTATGTGACGCCAACATCAA  
 AATTGCTTTGAAAAGGATCAAGAT  
 CTGGAGTGAAGGCTGGAGGAATAGCTAGACACTGAGGCTCATGCAAGGAGAGTTAACTTTAGAAATATATATTA  
 CATGAAACCCCTTTTTTTCTCCCCTT  
 CCAGAAATTTGTAGAATCATTGATATGAAAGAGTTGGCTAGTATATCTCCCTGTGTTCCCTTCCCACCCTGCTGTG  
 CTATCCACACAAACACAAGTGGCA  
 TTTGAGGTAACTTTTGAGATTCTGCTGGGTAAAAAATATATATTTTTGTCTCAACCCAATATGTTCTTTTTTT  
 CAACAAAAAGTATAATGATTCTC  
 CTATATATCTATGTTTAGCTTTTTGCTAGAAGTGTTTCCTAATAACTTTTCTCAGGAACTTGTTGACTCAATTTACT  
 GTAGTCATAGAGCCTCCTACCT  
 GACCCCTGTGGTCTGCTTTGGTAACACACCTGTAGGAGGCAATGTAATCTTCAGGAAATTCACCTCCTACCAGT  
 AGTCCTGAGTAAAGTGAATTTTCTA  
 AAATGAAACATCTTGTAACATTATATATGACAACTCCCAAATGGTTAAACTCTTTATGCATTTAGTGTTTGATC  
 TTTTCTTAAGAAGACTTTTTTT  
 TGGATTGCTGTCCTGTTGGCACTCAACAAATTACAATTGCTAATAAGTTTCAGAGGCCCTAACTTATTTTCCTAAA  
 TCATATACTTCCCAGTTTTTGAAA  
 AGGCTCTCTATTATAGCTAACAAAGTTTTGAACTTTGATTAACCTCAAGTGTTTTTGTGTTGTTGTTTTGTTTTTGT  
 TTTTGTTTTTTTTTGAGACGG  
 AGTTTCGCCCTGTCGCCCAGGCTGGAGTGCAGTGGCAGCGATCTCGGCTCACTGCAAGCTCAGCCTCCCAGGTT  
 CACGCCATTCTCCTGCCTCAGCCTCC  
 TGAGTAGCTGGGACCACAGGCGCCTGCCACCACGCCTGGCTAATTTTTTTGTATTTTATTAGAGACGGGGTTTC  
 ACCATGTTAGCCATGATGGTCTTGA  
 TCTCCTGACCTCATGATCCATCTGCCTCGGCCTCCGAAAGTGCTGGGATTATAGGCGTGAGCCACTGCGCCTGGC  
 CAACTCAAGTTATTGAATGAAGTAA  
 TATGAAGCACATGGGCAGAGACTTCCTGGTCCTATACCTGTCCAAGGCATGTAACCAACATTTTGTTTAAGAGCT  
 TTCCCTAGGGAGATCAATCCTGACC  
 TGGTTTAGAAACATTTTTCTGAAACTCTCATGAAAAGAGAAGAATGATTTTGGATGCTACTGTCTTATCGGAA  
 AGTATTTACAAGAAACCATTTTGAC  
 CAAGAACTCAAAATTAATTCTGGTTTTTAAAAAGAGAGGCATATATTACTCTTAGATTGAACCTTATGAAATTAC  
 TGATATTTGACTCACCAAAATGAC  
 AGTTTTTTATGGTCCAACCTAATACTTTTTGGTGTTAATAAAAGGGGAGGCAGGGTTAGTTGGGCAGTTTGTGGG  
 GGACATCCTACAGAGTTTGACATGA  
 AAAGTATATCTTAAATCAAACTTGAATTTGTACTTATTAGGATTTCAAGCTCATTTTTCTCTGGTGACCTTGCCA  
 GCATTATCAGATTTGTATTAAGT  
 GWCAAAAGTTTCCTCTTGTCAAAGTGTAACCTTGGCTCATTCTGGGCCATCTGTTGTACACTAATAGTTTTGAT  
 TTATTCTTCTGGCTGACATGCGTA  
 TTAATGGTACAGATGCAAGCTTTATATTTCTTTTATGGGTCTGTAGTTTTATTGCATACATTAGTGGCACAATA  
 CAGCATTATATTAGCTATTTATA  
 TGAATATTTCTAGTTTTGGGGTGACTAATACTTTGCTTATGTTTTACTATGCTGAGAAGCCCTGTATCGTAG  
 ACATTTAATAGACTTTAAATGAT  
 TACATTCTACTTAATTTACCACCCCTAATTTGTTTTTGTAAATTTTATTTTCAGGATATGTCATGATGTATGTCT  
 GTATATTTTATGATTTCAAAT  
 TCCCAAATTGCCATTGCTTAATGAATGAGCACATTTGGTTTTTGATCAGGAATTACTATAACTACTAGAAAACAA  
 CACTAGTGCAAAATGTGAATTTAG

ATAAATAACTCTGAAAGGCCTACTTCTAGTTGATGCTTTGGCTCTGATTTTTGAGGGTTAATATGTTCTTCAACTTC  
TTTGACAGCTTGAAAGTTATGAA  
ATAGAGTGATAGAAAACAAAAGCTCTCTAGGACAGGGACATAGGTTAGAGAAAATTTAATACAAAGCTGTTTTT  
GAAATAAGGCTAGGTGCTATTTTTAT  
TTTTGATCTGCTTTTAACAGTTTCCTATATTGGAAAACCAAACCCATCTCTGGAATCCAGTGGATATGGGCCAAAT  
ATAAAATGTTAAGAGGCTGGTGTT  
ACTATTGGTCTGCTATGCAAACTTTTTCCCTAATAGAGGTCTGTGTTGATCTGTGTGAAAGCTGCTCAGTGTAC  
AAGTGCAGCCTGAGTTTCATTCAT  
TTCTAGGGGAAATTATGATAGAAATAACATTCCTTTCTATTAAAGATAAAGGGCTCCAACCTTGAGCCAGAGTTT  
TCTAATACAGTAGAAGCCATCTTAT  
CTATATAGTGGGTACTAGTAGATAGTCTGTTAATAACAATACTGGCTAAAATAATTTTTTAAAAAGATATGTATC  
TTATTTTCATTTGATTATATTAAT  
GTGCACTCATTGGCTAGTTTTGTATGCAGGGCCAATTTCTTTTCTTTGAGTCCAAATCCACTCTTCAGAGTTCCATG  
TTGCCATTCTTGCAATAAACTATA  
CTCACAATCAGCAACCATAATTTCCAGTCTTGGTTACAGTTACTGTATATTGGAATAAGAACTTAAAGGTACTTG  
TTAAGCAATCTGAGGGCAGACTAT  
TATGATTTTTATGGTTGCTTTGCAAGGAGAGAACATTTTAAAACTTACCTTTATCTTTTTGCTCTCTTTTCACACTAT  
TAGTTTATGTAATATTTTATTA  
CACAATTATACTAATAACAATTAATGAAAGTAGTCTTGAAAAAAGACACAACCTCAACTAGTCGTGAATGTTCCCTG  
TGCTATGGGCACCAGTCAGTGAGTT  
TTATAGGGGGAATGGGAAGGGTCATTTAATCAGTCTGTACTCTACATAGATACGTAGAATCAACATTTGATAACT  
CTCAGTAGGGCCTGAATTTTATTGT  
GAGTTACCATAATTAATTTAGTGGAGCAAAATGTGGTCCTTTTCTTTGATAAGCCAGTCTCCAATGCTTTTGTTT  
TCTTAAAAAGACAAGAAACGTAA  
CTTTGAAGGCTGTGTTATACACTAATATTCTGTTGGTTTCTAACAAATTGCTTGGAAGAGGGAATTGGGAGAACT  
ATATGAAAGTAGCATATTTTCTCAT  
GCACAATATTGCACAACATTAGAAAACCTGATGCTTTATTTTCATGGTGAAATGTTTTGATGTTTCTCCCCGCCCC  
TCCCCTCCATGTTTCTCTAACTTT  
TTCTGGGAATAAGAGTAACAATACATTGATTATAATATTTGTCCACATGATTTTCATTCCAGAGGAAATCCTATAGT  
ATCTAAGCCCTACGGAAAGCCTCA  
TAAAGACATTTAGGAGTCCATTGGAAGATGTCTTAAATTAATGACATTGTTTTGAGTGACAGTGCCGAAATAT  
ATTTCTTTCTATGGCAACATTCTTA  
TTCTACTATCCTTTTTATTTTTAAGTAGTTTCATTTTTGTCACTGTTGGGTTATTTTAAACRATTTAATTGGAGGGAA  
AGAAATACAAATTCATCAAATT  
TTACCTCCAGAATTGAATTTCTTCACTAGTTCCTTAATAAGTATTTAAGTAAAAGGGGGTGCAATATCAAGAAG  
ACTAAATTATAAATATGGTTTCTAT  
GCTGTTAAAAAATAACAGTCCTGTCCTTAATTAACCATTATGGAGATTCCTCTTGACCAAATACCTTTAGAACTGT  
TTCACAGATTTATAATTTAACTAG  
TAATTTGGTTTAGGTTTAGGTTAAGTGACAGAAAATCCAAAATAATAGTGGCTTAGATAAATTATAAGTTTATTT  
CTGTCTCATGTATAAGTTTCAGAG  
ATCTAATGATAGCTACTTAATCATCGGGATCCCTGTCTCCTTCTGTCTTCTTGCTCCATAATCCTCAACTTGCAGCT  
TCTGTCTCATGGAACAAAGACGG  
CTACTCCAGCTCTGGCAGTCATATCTGCCTTCTAGCTAGCAAGAAGGAGAAAGGGAAGGAGGAGGGAAGCCTT  
CCACCATTACCACCTTGAAGAAGTTGC

CATTCTTCTTACAGTCCGTTGACCAGACCGTGTAGCYTTATTTGGTTGCAAAGGAGGCTAGGAAATATAGTCTTT  
 ATTCTGGGTGGCCTTGTGCCCAAC  
 TAAATATCACAAGTCCTGTTGCTATTTTAATGTTACAGAGGAAGGAGAGCATATATTGGGTAAGAAAAAAAAAA  
 AGGCTGTTTCTGCCACAATTCAGTTT  
 CCCCACAAAATAAAAACTTGGATTTATATTCTTTAAAAGAAAATATCCTCTGTAGCACAGTGCTGAATACATGAC  
 AGGACACTGTGCTCAGTAAATTTT  
 TTTTAAGTGAACAGTAAGGGCCGGTCACAGAGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGG  
 GCAGATCACGAGGTCAGGAGATCAAGA  
 CCATCCTGGCTAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCATGGTGGCGGGCGC  
 CTGTAGTCCCAGCTACTCTGCAGGCT  
 GAGGCAGGAGAATGGCATGAACCCGGGAGGCGGAGCTTGCAGTGAATGGAGATCGTGCCACTGCACTCCAGCT  
 TGGGTGACAGAGCAAGACTCCATCTCA  
 AAAAAAGAAAAGACAGTAAGAGGCCAGTAAATATTTATTGAGTGATTACAGCTCCTTTGTTTCTGAAAATGTG  
 ACTCACAGTACTAATGATCTGAAAAA  
 GAGTAAGATGAACACAGAGAGAAAAGTAATGATTCAGGATTCATTGTTGTGAGGTTCAAGAAACATTTATTTTGA  
 GGAAGTGGAGGTAAATATTCAATGAA  
 GTTAACTACATTRTCCCAGATATATAGATAAGATAGATAAAATGTATATATTATATAGACACTTATTTTTATTATAA  
 GATATGTACTGTTAGAGTTGGTA  
 GTATACCAGCTTTGGTTCCCTGCAGTCTGATTTTTGGGCCTCTTGGTAAACAAGTACTTAAAACTGTTACCCAG  
 TTTGCAAACCTTTAATAATTAGAG  
 ACAAGGCAGGTAACTGTAATAATTCGTATGCATGCAAGATATGATTCTTCTTTTTCATAAATCATACCTAGTTTTT  
 AGCTTTCTGCTGCAAAAAGTAACT  
 AGTTTGTTAGAAATTTAGTGAAAAGAAAAAGTAGGAGACATTGTGTTTCAAGATCATGTTTAAAAACAACAAT  
 TAGTATGGCATCTGTATCTTAGCAGG  
 AAACCTCAGAATTAAGTCTGAGCCTGACCTTAACCCATGACCAGTGGAATCTAAGTTTGTTTATTTTTACAATAGCC  
 TTAATTTGGGGCTTGCAAGTTTTT  
 TGGTTTGTTCACATTCAATTTAGAGACCCTGAARCAGTATTTTAATGATAGGAATTAAGCTATGTATAACTTTC  
 ATGGCTTGAAGATCAGMAAATTCT  
 GTTTTTCTTGGTTATAATGTTATCTTAATCAGAAATGTGTGATCTTTGTGTCTAACAATGCACTTAAAAAAACAAA  
 AAAAAAATTCCTTGCTGCAAT  
 GCTGGAATTCTAACAGAATTTACTCATCTTCTCTCCTCCCCAATTTTTAAAAACATGTTCTCATGGAAATGTTGC  
 CTTGAATTACACAAAATACAAGC  
 AGTTACTGGAAATGTTTTCTACACAGTTCTGTTTATGTATATCAATTTCAATTTGTTTTCACTTTTATAGATTTTTT  
 CTTAGTTTTGCTGATAATCACA  
 CTTCATAATCATATCAACTTTGAACCTTTGAAGTTGTGTTGAATTTTTATTGCTCTCAGCTGTATCAGATCCTACTTTT  
 TGATGAGTTATTTTATATAGTA  
 AGTTTGTGGAATGAATGGTGAAACAATGAACACAGGAACAAGAAAAGCAAATGCATCAAACCTTAGGTTT  
 AGTTTTATCTCCAGTTTAACCATGCT  
 CCCTTAAATGGAAATGCTAATATACTTCTTTTTCTTTTTCTGTAAACAGTTCATATGTTTTTGCTAGGAAGAAGA  
 GGAGTTACTGAAAACAGTAGTTA  
 ATATCAAAGTCATAAGTCATTTATAATCAGCCTTAGAATGCCTACAGATGTATATACATTTTTCAAACACTTATA  
 CTCTGCCTTTTCTGAAGCCAAGCT  
 TCTCTTGGCTTATTCTTACTTTTAGTACTTGTGTGCTGGAAATTAGAAACCTCAGATGCCCTGCACAGTGTAGGAC  
 TATACAATTGTGCTAGGCATGTAT

TTATAATTCTTTAACTCTTCTACCAAAGTGTGTGCTTTTAAAGGGCAAAAAGTGTACAACTTCTTTAGCATGCACCT  
 GTCACCTGCAAATAAATGTCTAA  
 TTCTGTGGTCCCTTCAAAGGGAGAAGAGGAGGCTAAGCAGGCAAGGCTCTAGGCCTTGGCCCTGCTTCAGTCAG  
 AATCTTTCTTGTACATGTTGATAATT  
 TCTTTCTTCTACAGATAGAAAATTCCTGGGATAAGAATTTTGATGGGAAAAAGAGTTTATTGCCTTTAAAAAAA  
 AGTTTGAGAATGACTGTTTAGGGGA  
 GATACTCTCAGTATCTTGTGACATTATCTAGTCTAGCCCCATCATTTTACAGGACTGGAACTGAAACACAGAAA  
 GGATACGTAGTCTTGTCCAGATTA  
 TACAGCTATTTAGTGGCAGAATTAGAATTAATCCCTTGTCTTGACTGCTTGACCAGTATTCTTTATAGTGGTGC  
 TACAGCATATGTGTTGTAAAAACC  
 TAGGAGTTTTCCAAGATCACTGCATAGTGCCTGACAACGGTAGACACTCTGTAAATGTGTATTACATTAATGAAT  
 GAATGATTCTAAGCTGGTTTGTAGT  
 TAATTAAGCCAACTTCAAATAAAAACTATAATTTTTCATCATCCCAACGAAATATTGACCTCTTAGAAGCCCTGAA  
 TTTGGCCTGTTCTAGGACTAAGCC  
 AATTGTTATATRTAACAGATTGTTCTTTGTTAATTTTAGGATCACCTGCTGTTACCAGCTACCACCCATAACAAGAC  
 CACAAGACCAAAAATGTCAATTG  
 TGTTACCAGCAATAAACATAAATGGTAAGTTGAAATTCTGATGTATGAACAATATGAGATTTCTTTGTTTAACCAG  
 GATTCTCTTCATTAGTTCAAAGAC  
 ATTTAAATGTTTCAGTTAAATGAGTATATATCCGTACATTATTTCTGAAACATTTTCCACTGTTTTATTATTTTAA  
 GATCCATTATTCTTTTAAAAAA  
 AATTACTACATCCCTAAAAGAAAAAGTAGCTTTCTTTTACTTACAAAATTGGAAAAGCTAGCATCTTGACTGCTCT  
 ACAGACTTGCATCTGCGACCCTTC  
 TAGTCATAAATGTAGTGTAAGAAGCCCCTGCTGTGTTCTGACAGTGAATTTGTAAATTAGATTTTGGAACTCAA  
 AGCAGTTTCCTGTGTTACCTAAATG  
 TTGATATCATCCTAGACCATGCCACAGAAGCCTGTTTATGTTTAATTGATAATACTGATGAGTTTGACACTGATAG  
 TATTACTATTGCATTTTGTATAT  
 TACTAGCCCTGGACATTAAACAAAATACAAATAATAATATATGCACAGATGTTAAGGCATAATATACCTGCTTA  
 TAGTTTAGAAGAAAAAAAAGACAA  
 GCACAAAATCACTGTACTCCAAAGAGAAAGTCACGTAAGTAATATGTGAGGTACTAAGAAATCTAGTGAAAAG  
 AGATTACTTTCAGCTTGATTCTTCAT  
 GGAGAAGTGAAATAACTCACTCCCATTACCAGTTGTGGATTATAAATGAGCACAGTGTATGTTATTGACTAATTG  
 CTTTCTCTGACACATTATATTGAAA  
 ATCGGAGAGTTGGCCATCTTTTCAAGTTTGTGTATCTCACATAATGTAGGTACTAGTAAATTTTCTGTCTGGTAG  
 CAAAAACCTCAGTATATGAAGAAG  
 ATGGAAACGAAAAGCCCTTTGAACTAACCTGCAGTCGAACCAGTTCACATTTTGATTGATAAATATAATTAAACC  
 TTCAAATACAATTATAAAACAAAAC  
 ACTAGAATTTTGGGGGGATTCCAGATAAGTTTCAATAAAATTTAGTCTTTAGGGGAAAAGTGGTATTCATCATAC  
 ATTTAAGACCATAGAGGACCTTAG  
 AATTTTCATGATCTGTATATAGACAGCTTATTAGAAATGCTTAGGAAGTAACTTCATAGACATAAACATAAATGGG  
 CTTTTGTAAAGCATAAAAATTAC  
 AATATTTACTTAGCCTTACAAAGACATTGAGAATAAGTTTAAGAAAAACAGTAACATTGAGAAGAACATCTCTTT  
 CTTGAGTAAAGACATACCTTTGAAG  
 TTAATCATGGAATAATCATTTACTACCTGAAGAAGAATATATTCAAAGTGCAATTTTATTACATTCAAAGAATGT  
 ATTAGGAAGTCTCACTTATACTC

AATCTATAGGTACCTAAGTTTCTCTGCCTAACTAGAAATAGACCAAAGGGTCAGCTTATGCTCAGAAAGGCTATT  
 TTCTAAGCTTTATCAAGGCAGAATT  
 CACTTAATTTTGTCTTCTGATATTGGGTTCTTATGCAAGAGAATGCCTGGTCTTATTTTGTAACAGTGAGCTATCT  
 TTATATTAAGCCAAAACCAACAA  
 GATGCTGTTTGACTATGTTTATCTTGGCCAGGTGACTGATTTTTGAAGAGGAGCCTTTCATGGATTTGGAAAAAT  
 CCAGTCATTAGAGGATCATTGTCTT  
 TTTAATCTGTTTCAGACTCTTTGTAAATAGTCTTTTTAAAACTTCCCTTTTTAGGATAAGTCCTATTAATAAATCA  
 AAATTCAAATTCAAGAAATATTG  
 ATTGAATATATACCGTGTTTCAAGTACCATGCTACATCCTCAGCCTTTTCTCTTAACAGTCTTGCTCTGAGGATACT  
 TCTTAAATTAGAGCCTGATTTTT  
 TCCCCTTCCTCATGTACCTCAAAGTCTGCAGCAGGGCAGGTGGTATCCTTGCTTCCCACTACAACCTTCAAATTAT  
 TTCTCCTTCAGTCCCCTTAACATC  
 TTGCAAATCTGTCACATCACACTTTACTCCTCTAACCCTTATTTGTTGCTGCCGTCTACTCCACCTTCACTGACGACT  
 CCTCCACTGCCTCTGTCATCCT  
 GCTGTCTAACTTCAGCATCCGCATAGAGGAGTCATTAGTACTCTAGCCTTTTTTGATCCAGTCATCTTTTATTCCGC  
 GCCACCTCAGCCTCCCATCCATA  
 TGGTCATAATCTAGACCTTATCATCATCAACAGCAGCACTCTCTTGAAATCTCCTTTTCAGGCATCCCACTCTCTG  
 ACTGTTACTTACTCTCTTACTTT  
 AAAACTCACACTCCAACACTCTCTTAACTTACAAGTCTACAGTCCTTTGACCCTGCCACTTCTTTTCTGTGTCCATC  
 ACATCCTTAATGGCCATACATT  
 CCTCCTTACCCAGCTTAAATTCCATGGTCCTGTCACCACTCTCTTGTAACACCCCTTACTTCCTGGCTCCTTTCTTA  
 GGTCCCACTTTCTTGAAAAAC  
 TGCTTTCTAGGTAAACCCAGCTCTTCTCTCTATGCCTGCTCTCAAGCAGCTGAAAATTGCTAGAAAGTACCGTTCA  
 GTGAGAACAAGTAGATTCACTTCA  
 AATTTGTGACCATTAATTTCAAAGGGATATTTAATATGCCAGTCATTACTAGTAAATTCATTTCCCTCTCTCTGAA  
 AAAAAGAATTTCTCACCTTCTCT  
 TGTCAAACTCCAACATCCTCTCTCCCCACCCCTCAGCTGATGAACTCACCAAATACTACACTGAGAAAATAGAT  
 GTCATCATGCAAGAGCTACATCAT  
 GTTCCACTACCACCGGCCAGCTGCATCTACACCCACAACTCTGCCTTCCCTCTTGTTGCACTGAAAGAAAAGT  
 CTCTTCCATTGGGTTCTAAATCCC  
 AGCTTTGCTACTTGCTGGATATGTATCATTGGGCAAGTTAATTAACCTTTTCTGTACCTCAGCTTCCTCATCTGTAAA  
 AGAAGATATTAGTAATACCTTCC  
 TCTAAGAGTTGATGTGAGAATGAAATGAGTAAAAAAGCTTTAAATAGTACCTGACTATTTTGGGGTTAGCTGCTG  
 TTTTACTGCAATAATCATCATTGCC  
 ATTGTCCTTAAGACATCATTGCCAACTTYTAAGGTAATTTCTTCCTTTTGGACCCTGACTCCTAGCACTTCTCACAG  
 TGCTCAAAGACTTTACTCTGCAT  
 TTATAMCATTTCGTCTTTTATACAGTTTCTTCTTCTCTACTGGGTCATTCCCATCAGCATACAAATGTCTTTTTTTTTT  
 TTTCTCGAGGGAGGGACTCAT  
 TCTGTTGCCAGGCTGGGGTACAATGGCGTGACCCATGGCTCACTCCAGCCTCCACCTCCTGGGCTCAAGCAGTT  
 CTCCCACTTCAGCCTCCYGAGCASC  
 TCCTGAACACAGGCACATTCCACTCCCAGCTAATTTTTAAATTGTTTGTAGAGACAGAGTCTCCCTATGTTGCCCA  
 GGCTGGTCTTGAACCTCTAGGCTC  
 AAGTGATCCTCCTGCCTCAGCCTCCCAAAGTGCTAGGATTACAGACATGAGCCACTGCACCCAGCCCAAATGTC  
 TTAATATTACTGTTACCTCTGGTGT

AGGGAGACTTCTTTGACTTCACATCCTATTTCTCTGTTACCCTTTAGAGCAGAACTCTCAAAAGACTTACACTCACT  
 GCTTCTACTTCCTTACCTCTGAT  
 TCTGTCTTCTTTTACCCTTACCAAGCTTTTGCCATTTCACTTCATGAAATGCTCTTGCCCTTCCACTCCACTGAAAT  
 GCTCTTGCAAAACGGAAGTCAA  
 TGACAAATATCTCGCCAAGTCAAAGGTCATGTCTCCATGGTCATCTTACTTGACCTCTCAACAGTATTTGACACAG  
 CTAACCTATTCCCTCTTTTTTTTAG  
 TCCACCCAATTCTTATTTTTAAAAATTTCAAATCTATAGAAAAGTTGACAGATGAGTGCAGTAAATATCTAGAGAC  
 CCTTAAATAGATACCACCAATTGT  
 TAACATTTTTCTATATTTGTTTTATATCTATCTATGTACACTCACCAGATTTTAAACCTATTTGAGAGTGAATTGC  
 AGATAACATGGTGCTTTACCCCT  
 AAATATTTGAAAATACATCTGCTATAGAAAGGACTTTCTCTTGCCATAACCACAATAAAATGATGATACTAAAAAA  
 AATTTAACATTGATACAGTAGAGTT  
 CAATGTATAGCATATATTCAAATCTCTCCATTTTCCCCCAAATATTCTTTATAACTTTTTGAATATATGTTCCAAAG  
 ATTATACATTACCCCACTCAT  
 CATACCTCTTTGGTTTCCTTTAATCTAGAAAAGCTCCTATCCTTTTTACTGTCTTTGATGACATTTACGTTTTTAAAA  
 ATTAGACTTATTTAGGTATAAT  
 GTACATACAATAAAATTCACCAATTTTTGTATACAATGTAATGAATTTTGACAAAAGTAAACCATTGTGAAGCCA  
 CCATTTCAATCAAGATATAAATGT  
 TGTCACTACTTCAAAAAATTCCCTCATGCTCCTTCCAGTCGTTCCCCTCCTGTACATTTTGGACATCCCTTCTTTCTT  
 GAAATATGTTATTTTCTAGGC  
 TTTTCATGATACCATGACTTGTCTAGGTTTCCTCCTGTATCTCCTCAGTCAGCTTATCCAGCTTCTCCTCTTCTGTCAG  
 GATGGAATCTTGTTTGTTATGT  
 TCGCCACTTACCACCATATCCCCAGCCCCCTGGAACAGTACCTGGTGTAGGAAACCTTGAAAGATGAGGCACTTTC  
 ATTTACATTAACCTCATGTCATGCTG  
 AGCTGCCAGTTAAGTGCAGGCTAACTGTTCTTTATTCAAATGCTGGGATCATGTGGTGTGAATCCATTGGTCAAA  
 AAGGAAAGATAAGAAGGMCAAAATG  
 CATGGGACCAGAATTATTTAGATTTTGGATTGTTTCAGATTGTGGAATATTTGCATATACATAAGGAGATATCCT  
 GGGGATGGGACCCAAGTCTAGACA  
 TGAAATTTATCATTTATGTTTCAGATACACCTTATATACACAGTCTGAAGATAATTTTATACACAATTTTTTTTTTTT  
 TTTGAAGGCAGAGTCTCGCTCT  
 GTCGCCTGGGTTGGAATGCAGTGGTGCGATCTCAGCTCACTGCAAGCTCCACCTCCCGGGTTCACACCATTCTCC  
 TGCCTCAGCCTCCCRAGTAGCTGGG  
 ACTACAGGYGCCTGCCACCACGCCCGGCTAATTTTTGTATTTTCAGTAGAGTTAGGGTTTCACCGTGTTAGCCAG  
 GATGGTTTCGATGTCCTGACCTCG  
 TGATCCGCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCCTTTATACACAAT  
 TTTTACAATTTTGCTCATGAAACA  
 AAGTTAATATACATTGAACCATCAGAAAGGGTCAGGTGTGGATTTTCTGCTTGCGGCATTATGTCAACACTCAAA  
 AGGTTTTAGATTTTGGAGCAATTG  
 GATTTTCAGATTTTGGATTAGGGATACTCAACTTTTATTCTCTTTATTTTTTATAGATAATGCCTCAAGAAAGTAAT  
 TCATACCATTGTCACAACCACTA  
 TATAGCAGAGCCAAGATTACATACATTGTATAAAATATCTTCTAATTAATTATATTGTTAATTCAATATTAAAGAC  
 AGATGTGCAAGGATTGAAAATAT  
 TTGGGACTCCAAAATTTAAGATAGATTTTATACCATTTTATTCTACACTATAAGGTGGGAAAGATTCAATTGGGA  
 GCAGTATAACATAATACAGTCATTT



GTGCAGTGGCGCGATCTTGGCTCGCTGCAACCTCCACCTCCTGGGTTACGCCATTCTCCTGCCTCAGCCTCCCAA  
 GTAGCTGGGACTACAGGTGCCCCG  
 CACCATGCCCCGGCTAATTTTTATATTTTAGTAGAGACGGGGTTTCACGATGTTAGCCAGGGTGGTCTCGATCTC  
 CTGACCTCATGATCTGCCCCGCCTC  
 GGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGTGCCTGGCCATCAGTTGTTTTTGTGTTGTTTTATTAT  
 ATTGAGAAGTCCACAGCAAAGAAT  
 CACAGTGAATGAGCCAGTTGCTGTGTTGAGTCAACGGCTTAAGAGTGGCTAAGGCCGAGATTGCAGAAATTCTT  
 TTGGCCTTTCCAGATGATCACAAAG  
 TGA CTGTTGCAGTTCAAGCCATCATATTTGTATTACAGGCAGAAAAAAGGAGGTGTCTATCAAGAAAGCAGACA  
 TTTCCAGAAATTTCTAGAAAGCTCT  
 GCTTATGTCTTATTGAACATAAACTGTGTCAACTATTTAGGCCTGGCTTCAAAGGAGGCTCAGAAAGATAATATT  
 TTA ACTAGGCACATTATGATTTGAA  
 AAAAAATTTGGAGTCTTTTTAGTAAGGATCCACCCTCGGCCTCCCAAGGTGCTGGGATTASAGGTATGAGCCACC  
 GTGCCTGGCCATCCGTTGTTTTTG  
 TTTGTTTGTGTTTATTATATTGAGAAGTCCACAGCAAAGAATCACAGTGATGAACCAGTTGCTGTGTTGAGTCAAC  
 AACTTAAGAGCAGCTAAGGCCAAGC  
 AGCAGCTGCTTGCTGCAGCACTCCATCATATGCTAAATTGGAATTGAGCAAAGTGTCTGGGAAGAGACAGTAAA  
 GAGAGGATTCATGTTGAAAAGAGAAA  
 TAAATATTTATTTATTGCATGGGATGCAGCCAGATGATTTAGGTCATTAATGAGTTGAGGAAGTAACTTTAACT  
 ATCAAACATAAGAATAATACAATGA  
 AGACTCTATTAAGAGATTATTAATCTGGCTAAGAATTGTATATAGAAAAAATGAAGAGCGAAGATGAGCTAGG  
 ATAGATACTAGGATAGATACTTTGGA  
 ACAATCTCAAGATGAGGTGATTCAAAGCTTGGAATGGGAAGAAAGGAAAAATTTCCAAAGAAATTCCAAAGGA  
 AAAATCTATAGCATTTGGCAGCTGATG  
 ATTTATAAAGGAGAGAGAAGGAAAAGCTGATTCCATGATTTTGAGACCAGACCACTGAGAGACTAGTACTGCCA  
 TTCAAAGAAATGGATAAAATGGGAGG  
 GACAGTGTGAGATTTATGCATCTTGAATTAGAGATGTCATTGGTGGGTATTTTTGTGTATCTCTTGTAACAGATTG  
 GTGTGTATTTCTGTGTRTCTCTTG  
 TAACGGAATACCAAGTAATTCTTGACTGATGAATTCATTTCTTTCCTTTTTGTAGATACTACCAGCSCCCTCAC  
 AAGTTAGTGTCCCTTTCCATCTG  
 TTGTATTGTCTTAGACTTCTCCTCATTCACCACCTTGATGCTTTTACAAGATAACAAAGCTTATGGAGATAGGCAT  
 GGGGCAGCAAGAAATATAGAGTGG  
 GAATCTGCAGAGAAAAGAAAGATGAGCCCTCTTCTTTGCTTGTTTTCCCTGTTTGTTTGTTGTTTTTCTCTGCT  
 TTGCCTTGGGCATGATTGGGATT  
 ACTGAGATTATTCCTCTTGCTATATTGTTGCATTAGTATTTAACTGGTGTCCCTAATTTATTATTTTCCATCTTA  
 ATTCCTACAATTCACCCTGCAT  
 ACTGCTGCCAAGATGAGTCATCCTTACATACTGCTTTTGTTATATCATTATCCTTCTTGAGAACCAGCGGTGGATT  
 TTCAGTGTCTATCAGATGAAATCC  
 AAATGCTTTAGCCTGGCATCTAAGGCCCAGGACAGTCTGTCCCCACATTACCTTTTCAACTCAGTTCTAACCAAAC  
 TGTTCTACTAACAGTTCCTTGAAT  
 CTGTATTCATTAGTTCATTTAAAAATACTCATTCACTGATTCTCCTTTATGCCTAATGCTGGAGATACCAACATAGA  
 ATAAGGTACCCATGATTTGTACT  
 TTA CTGTAGCCTCTAAGTGCAAGAGACAGAAGAAAACAAAAAAAATTAACAACCCAAACAAAACAATTGCAGC  
 TTATGACAGTTTGGGAGAAAGAAAGA

AATATGGTCCCTTCATGAAATATGGTGGGAGTCAGGGGTAGAGAAGACCTTTTTGATTGGGTAATGAGGGGATA  
 ACAGTTAAACCAAGATCTGAAGGATG  
 AGAAGAATCAACTGTGAAGAGTCAAGGGGCAGAAGATTCCAAGMAGAGGGAGCAACAAGAGAGCAGAGGCC  
 CTGAGGCTGGAAAGAATGTGTTTCGTTTTA  
 GGATTTGTCAGAAGACCAGTAAGGCTAGAGTAGGACCCAGTAGGGAGAGGAAGGAATGATACGAGATCAGAT  
 CGGGAAGGTAAACAGGGGCCACATCAGA  
 CAGGGAATTGTAGGCCACGGTGAAGAGTTGTGATTTATCAATATAGTGGGAAGGTGCTGAGGAGTTTGAGGTG  
 GGAAGTGACTTGATCTGATGTACATGT  
 AAAGAGATCACTCTATTTTTGCACATGTGGTGTCCATGTCTTGGATTACCTGTCCTCTTCCCTCCAGGCTATCTCCT  
 ACCCATCCTTCAAATCCAGCTG  
 AAATCCTACCACCTCTGCAAACTTTTGCTTTACTTCTGTGGACCACTAGAATACTTCTGTAGGTCCTAGAATAAA  
 TAGTCTTTCCTCCACTGAACMTCC  
 TTAATGTAACCACCTCTCATTTATGGCTTTATCTAACACTACTCAAAGATAATTACCTTGAAGGTGGTGGTGYTTTT  
 TTTTTATTTAACTGTGTCTCTTC  
 CTCCTCTTCATCCCATGCCTAAAGAAATGCTCTGATACTTAATTTTTAATATATACTTAGAGTTGGGAAATTTTTGC  
 CTTTTTATGATTCACAAAAAATA  
 GCTGTTCTATAGCTGTATCTTACCTCCCTCATTTCTCTGGTTGGCCTTACTTCCCTCTTATTTCTTGAAGTCATTAG  
 TGTATCCATTTTTAACTCTTCC  
 CAAACCATTAGATTTTCTCCCACTTAGCAGAGCCTCTTCTGAACTCTCTACTATAAGGTAAACCTGGAAATAAATA  
 GATTTCTTAAGTGGAATAAATAGG  
 GTTAATGGTAGAGCTGCCTTTATATACTTGGAATGTTATTACTTTGCTAAGAATGATTATTGTGTTGATGATTTAT  
 TGGCAGTTGCATGGGTAATTTTTTC  
 CATGAATAGTTTATATTCATTGTATTTGATTTACTAATTAATTATTTAGTTTATTAGAACCTTCACAAGACCCCAGA  
 GGTTTAGATTCAATCAGCTTGCA  
 TTGATCCTTCTAGGATATTAGTTACTTCTTTGACCCAAAGGAGCCAGTTGCTGAGCCCCCTGTACCTAAGTGACCT  
 TCAGAGTCCTGTAGCTTTTCAGTG  
 AGATCACATTCATCTTTCTGCCTTATGGCAAAAATCTGTTTCTACTCAAACAGATGGTTTAGCTTTCAACAAATAA  
 AAAGTAGAGATCAGCATCTACTT  
 AAGCCAACAGAGTTGGTTTGAGCCTCCCACTAACTCTCCAAAGTCTGTTATTTTTTATTTTTTAATGTTAAGCTCCT  
 CCTAATGGATTTATCTTTGTACG  
 GGGAAATCAAGAACCTTGCACTTTGTCAGAAGATGAGCCACTTCATATAATCTCAAGTTTCTATGTTTTAGAAGA  
 ATTGCTTAGTAAGCACAGGTTTCGAC  
 TGCTCTGCGTTCTCTTTTTGCAGTTACACCAACTGGAAACATTGGGGATCATAACTTTCTGTGGAGTGTTACATA  
 GGTGTGTCAGCAGTTACATGATAC  
 TTTCTTTGTTATTTAAGACATCTGCAAGCTGAGGATGGCCAGTGTCATTTCASAGATGTACATGCCAAAGACATG  
 GTAGTTTTTCATTGACATTGGGAAAT  
 ACTCCTTCTCTTGAGCTTTTACAAATAAAAGTTAGAGAGTGCTAAGAAATCCACATATTTAAAAAAATCTAGAATC  
 TTGACATTGAAAATATAAATGCAG  
 ATTGTTAATTATATGTTCTGTGTTAAGAAAGCATTCTTGCTTTGGGGATCTTGAGTACATCAGACAAAAGTTGCC  
 TTTGAACTCAACCAAGTATCCATA  
 GAGATATTTGCCTCTATATGTATTATTAATTTTACAGAAACAAATGCCTAGCTTAAATATAATATTTAGATTAAATA  
 GTGAGACAAATGTAACCTGCACA  
 ATAGTTTGCTGCTAGGAAAGGAATTGTTAGGAAGGCCATATAAGCAGTTTCTTTGAAGATACCTAGGACCTATCT  
 CTTTCTTTTCACTCTATTATGTAGG

ACAGAACCTTGTGGGAAATGTGATCTACTTGGCATCATTATGGTATGGATTAGGACAAAGAAAGATGCATTAAT  
 ATTCCTGTACTCATGCCATAAAATAT  
 CATGGTTAAAAAGGCTTTAGTACAAAGTTATCCATGGTAACTCCCTTCGAGGGATTGGATTCTTTCTGCAACAT  
 GTTGCTCTGTGGCCAACCAGTTTAG  
 ATATAATCACCGCCATGTATAGAGACTGCCTGCTTTAAATAGCTCTGGCAATTTGTAAATTGTTCTTTCTGTCTC  
 TAAATCCACAGGATTATATATA  
 AATATAATATTTCTTCCACATGACAAACCTTCAGCTGTGTGTATACTATATACCCTCTGAGTCTTCACATCCCAAGA  
 CCTCCGGTTTTTTGGAAGCTATT  
 CTTGTGACCTGGGTTTCATGGTGCCTCGATTCTGGTCATTCTCTTCTGAATGGGCTCTGGTTTTTCCACATCTGTAA  
 AGAGGGCTGATGAATGCTGAGCA  
 TAGTGGAGGATCTCCTAATTCTATATAGTGTATTTCTATTGGTGAAACCAAGGTGTTTTTTCTAGTTACATCATAC  
 ATTGTACCTTCTGTCAAATATAG  
 TCTTTAGCTTTTTCATATATGTTACTACCAAGCTTTTTAATTTTATTTGCTGATTTGGATTTTTGGCCTGAATTTAGA  
 ACTGTGTTTTTATCCTGATTAG  
 ATACAGTTTTTCATTAAGTTCCTTACTCAATAAACATTCAGAGAAGGCTTCCCTAAGTCCATGGAATAGGTTAGGA  
 TCCCTTCCCAACTTTATGTACCCA  
 GAATACTCCAACTTTCTTATTTTTGTAACATTCATCTTACTTGTAATCAGTAGTTATCATCAACTTCCCTGCTAAAT  
 TGGAGGCTTCTTAAGGTTAGGA  
 CCTTATTTATTCAATATTCTTTGTTATTCTTACTGCCTGACAAGAGTAAGTGCTTAATACATAGTAGCTGCTCAAAA  
 ATGTTTGTTGAATTGAATTAAT  
 ATATGAAGTATCTATATAAGCCAGGCACTGTCCATTGTATCTAAGAAAAAATGTAATTAATTAAATGAACACAC  
 AGTGAAATTTGAACTGAATCTTCAA  
 TACTGGCCAAGTGAATGAGGCATTCTGAGCATAGGGAACAGTATAAGTAGGGATATCAGAGAAAGAAAAGTAC  
 TTAGTCCATTTGGGGAATTAAGTTGG  
 ACTAAAGTGCACATGGATCAAGGTACAGATGAAGCTGCAGAGGCAAGCAGAAGCCAGTTCATCATGGCCTTGT  
 GTAGGACATGATACAGAGTTTGACTTA  
 AGCCTCTGAGTAATGGGGAACATTAATAATATTTTGAGCAGGGAAGAGGCACAAACAGAGTTACATTTCAAAAG  
 ATTTCCCAGCAGCAACATAAAAGAT  
 AGATKAGAGGGGATTATGCTTTTATAGGGACAACAGTTAGAAAGTTATCATCTAGGCAATAGATGATGGGTAGT  
 TCATATTTAAAAAAGGATTCTTTCTG  
 TATACTATCTATTGTGTGTAACAAATTACTCCAAAACCTAGCAGCTGAAAGAAAACAAACATTATCTCAGACAGTT  
 TCTGAGGGTCAGGAATCTAGGGCC  
 AGCTTAGCAGGGTGATTCTGCTTCCGGATCTCTCTTGAGATTGAAGTCAAGCTGTTGACTGGGGCTGAAGACTTG  
 ATTGGGTCTCAGGGATGTGCCTCCA  
 CATTCACTCACGCAGCAGGTGGCCTCTTTCTGGGGCTGCTCACAACATGGCTTCCCCAAAGCAACTGTTCCGTAA  
 GAGTGAGAAAGCAAGTGAGTAAGAG  
 CACTCCAAGATGGAAGCCACTGTCTTTAACTTAATCTTGGAAGTCTCATACCATCACTTCTGCTGCGTGCTGT  
 TGGTTACTTAGACCAACCCTCGTA  
 CATTGTGGGAGGGGACTGTACAAGGGTATGAATACCAGAAAGTGGGGATCATTTGGGGGCATCCTAGAGGCTG  
 CCTCCATGCTGCTTGTCAAAGAATGT  
 AAAGAGTAAGTCTACAAGTCCATAATATTTTGATGGCTTATGGAGATAAGAGGGCAGAACACTCCACCTAGAAA  
 CTATCCTATTACCTAGAGCAGGTAAA  
 TAAAGTCATTTGCTTGACTGAAAGGTAAGAGGGTTAAACCCACCATTCTAAATGTCAAGTAATAGTTACGTTGC  
 AGTCWGCTTAGTTTTATACTATAAA

TTATGATGATTTGGTTCCATGTTTTTTTTTAGTATAATGTGATTGCTGTTGAGCTTTATATTCAGCGCATAGCACTA  
 TGCCAAGTAATGTACAGGTTTCA  
 AAAGAAGACTAAAATGTGATTCATGCCTTCAGAGAGTTAATTTACTGGAAAACAAGGCATACATAGATCATAGA  
 AAATCTTAGACCCCAGAGAAAACCTTA  
 GGGGAGTTTAACACAACAGCCTCATGAAACAGCTAAGAAAACGGGTTGAGGGAGGTGAATGGCTGATCCARCC  
 CAGCTACTGACAGAGCTGAGACCAGAT  
 CTCAGACCTTTAGGTCTGTAGGCAAGCATTTTATTTAAATATAAAAATGTAATTTGTTTACGATACCAAGAACATTC  
 TTTTTTAAATATAATGTAAAAAAC  
 ATTTATAGAGACAGGGTCTCACGGTGTTGCCAGGCTGGTCTCAAACCTCCTGGGCTCAAGCAATCCTCCTGCCTC  
 AGCCTCTTAAAGTGCTGGGATTACA  
 AGCGTGAGCCACCACACCAGGCCAAGAACATTGTCTGAGGACTTAAATCATTTTTCTATTTGAAATCAGAATGAC  
 TTAGTATTTGATGTGGGGAAAGGCA  
 AGACATAAATGTGTTAGAAGACATTTCTGTCTTCTCTTGAAAGACCAACTCATCCTGTCACTGATTACTGTGTGAC  
 TTTGAGCTAAATTACAACACATTA  
 GTAAATACAGAAGAAAATAATGAGCAATGCCACATCCCTTGCATTTTTTAAAGCAAGGACCASCCTAATTGACAG  
 TGAGATCTTTATATTAATTTTCCCA  
 TTTAAATAGATAATATTCAGTGTTTCGATGAAGAAAGAATTATGGGGTAGATTATGAAATAATAATTGGAATACT  
 ATTACTGTTAATAGTAAAAAATGTC  
 TGAACATACTGTTGGAATAATAAGAGTAGAAATGAGTAAATTAAGTATTTACAAGGTTTAAAGTATTTTGGAG  
 TTTTGTTTCCTTTTTTAAAAACAAA  
 ATGCTTTTTAACTGTTTCATTGCCTATATTACCAAATAGTTTTAAGTTGTACATTCAGGTTGAGCTCCCCTTTTTTGA  
 GTACCCACCATTTATGTTAAGTA  
 TATTACTGGGTTCTTTAACAGATATAATTGCATTTGGTCCTCATAAAGCCTATGTGAGATAAACATTTTCCCCATTT  
 TACAGCTGAGATAACTGAGAAAT  
 TAAGAGAACCACAACCTGAACTTTCTGCATGGTCCTTTGAGGTCATTTTTCTGAAAAGCAAGTGCTAACTTTTAATA  
 GTGTCTTTAGCCTTTCCCTCAGGC  
 AACAGCGTATGACTAAGTCAAGACAGCAGAACTAAATACGAGTTTATGTTACCATAATTTGAGAGTAACAGCTAC  
 TGTTTTATGTTCTGGTCATCTATTA  
 CATAAATATTTAATACTGAGCACATAATTTGGCATGATGGCAGCAAAAAGATGTCTATCCTGGAGTTTCCATGAT  
 AAACCAGAAAGGCTTTTTTTTCCCA  
 CCAGTATGTTAAAATATGGAGAAATAAATTAGGAGATGGTAAGCTATAGCTTTCTTATCATCTTAAGTTTACCTAA  
 TTAAGCACTTTAAAATAAATTCA  
 TTCCATCCAAATTTAGCAGCCATCTATTGCAAAGTTTGCTTCCTTTTTCCATGACATTTAAATATACAAAAGCCTTT  
 TTTTWWAAAAAAAAGCATTTTT  
 TAGTAGGTAAATGTCAGTCACTTTCTTAACTAGCCTGTCCTAGACCAACAGACCAGTTCTACTTTCAAAACAAAA  
 CATAGCATTTAATAGTAAAAATAA  
 ACTGTCTTCTCACTAATGAATTTACTGTCTGCATTTCCATAGATGTGGCCATGTGTTTTCCAGGAAAACCTGGCTT  
 GAACTCAAACCTAAGATAGCTACTA  
 AGATCCATTTAGACAAAGTGGGCTATAAACATGAGAAATCTTCGCCCCCTATTCTGTGTTTGGTTAATTTTGCTG  
 CTCTCATACATTCTAGCAATGTAC  
 CATTCTGTTCTCCCTTTTACTCCAGGGAGCCTCATTTGGCATATAGTTAGTCATACGCATACACCTTTGAAAACAG  
 TCAATGTAACCTCCAGTTATTAAG  
 AACCTGATGGGTGCTAGGCATTGCCTTATGAACTGAGAGGTACAAATATGAAAACCTATATAAAAGCTCCTGCCCC  
 AGATAAACTATGATACAGAAGAACA

TTTGTTAGCTACTTTTTAGGTGCTGTGTATCAATCAGTTTTCTTAGTTTCAGGTGACAGAACTAACTCTGGGTAA  
 CTTAAGTAGAATTTCTAGGAGGGC  
 CAAAGAATCTGACTTGGAGGTACGGTAAACAAAAACAAAGTCCTAAATTACACCACAGAGCTGGACATAGACCA  
 TAGCTTACATCAGTGATAGTGCTGAT  
 CCTGAACTGTACTTCCAGAACCAAAGCCATGGCTGGCCCTTGCAACTAGATACCACTCCTTCTAGGCCTTCTAGG  
 GCCACCCTTCTTTCCAGAAGAGTCT  
 CCTTTGCAGTCTGGAGCTGGTGCATCTGTGTGGCAGAGCCTAGGCCATATGCCTGTACCCTAGTTACAAGGAAG  
 TCTGGGAAATAAATTATCTGGTACTT  
 TATACACTTAACTAGTAGGAGGTCAGCTCTATCTCACAAGATAGGAGGTTCTGTAAACCTAGGTATGCAGTCAGG  
 CACTAAGTGGCCAAAACTTCTATC  
 AGCAGCATTTTGAATAATCAGAGAARATTCAACACAGGAGATGAGAACCAGGAAACAACAACCAAAAAAGAAA  
 TATGGTGAGTTTGGCCTTAATATCACA  
 AAAAGATGAAAAACAATGTGTTCTAAACCAGTTTACAAACCCACAGTTAGTGAAAGGCTAGGTGATAGCTTA  
 TATTCTCCAAAGTTATAAAGTCCTAA  
 AGAATGTAGCCTCTACTTGAGACCCATTTGGTCTGTGCAGGTACAGGCACTATCATCTCTCTCTGTAAAGGGTCA  
 GTATGATGTTTAGGAAAGCARCCCA  
 CATTTGGAGTCAAAGCCTTTTGTCTTACCCTGCGCAGCCCTTAACCAGATGTGTGACCTGTAAGCAAGGAGTTC  
 AGTTAGTATCCTCATCTATAAAATG  
 AGAAGGTTGAGGTAGGTGATCTTTAAAGTTCCCCCTTGGCTCTAAAGTTCTATACTTTGTACATGATAGGCCAGT  
 GACAACTAATGTTTTTACTCAAGCC  
 GATTGTATGTTTTAGAAGGCAAATAAATTATGAGCTTATTTTTTGATGACAGTACATTTTCATAAACCCAAAATAA  
 TTAAATCTCCACCTCCCACTTAGT  
 TTAGTTGTTTTGGGTTTTTTAAACAATTTTATTTTCCCACACAGTACAGTTTAAGCATTCAAGTTGCAGCAGGGG  
 AAATACTCATTGAATTTGTCATCA  
 TTATTTTTGGCTTCTATCTTTATGAGAGCCATTTTAGTTTTCTAGTTTCAGTGATCTATTTTTTAGATTTTAAGCTAC  
 AATATATTTTAACATTTGAGTA  
 TTAAATTGCATTTGTCATAATGAATTAGAACCTTACGTTTCTTCAAAAATAGCTTTATCTCAATATGAAATGCAAG  
 ATTATTCACATTTATTATAAAGA  
 GGAGAATTTTTGAGACATTCTATATCCACTGGTCAGTTACTGTTTTGATTTGGTGATGGAAAATCATAGCAAATT  
 GTACATTTAAAGTAAACCAAATAAA  
 TAACTTTATGCAGAGATAGAGATTGAAGTAGCCATTTGGAGATGAGAATTTTATCTCCATGCCTTCAAGAAAGAC  
 CATATGTAAACCAGGTCAGAAAAAT  
 AAAAATTACACCATATTTATTTTAATAGGACCATCTTAGAATACAAGACATAACACAAGAGAATTTATTTCTGGAA  
 TGACGTGAAGGAATTTATTATACA  
 AACTGCTCCATCTAAACATGTTCTTAAATGTAGGTTTCGTAGAATTTTTTTTTTTTTTCCCCTGGGGACATAGTTT  
 CACTCTGTCACCCAGGCTGGAGT  
 GCAGTGGTGTGGTCTTGGCTCACTGCAACCTTCGCCTCCCAGGTTCAGTCTCTCCTGAGTAGCTGGGATTACAGG  
 TGCACACCACCATGCCTGGCTAATT  
 TTTGTATTAGTAGAGACGGGAGTTTCACCGTGTTGGCCAGGCTGGTCTCAAACCTCCTGACCTCAAGTGATCCACT  
 CGCCTCAGCCTCCCAAAGTGTTGGG  
 ATTACAGGCGTGAGCCACCGCACCTGGCTGGTTCCAAGAAATTCTAATATTTGTTCTGTTTTTAAGGAGAAAAAT  
 TATTGAACAACCTTACCTCTAAACTA  
 TAGTCCTTTAATGATAAACACTACTGTTAAAAAATTCTTATAATTAACATAATCTGATATTGTTACTTAAGTCTTTG  
 AAGAAATAGAGAAAAACTGGCTA

ACATCCCTTATGACTGCATGTTTTTATATATTAAATAGCAGTATCGAATTGCCTTCCACTATCTTTTTCAGACGAAA  
TAATATATATTTAAACCTTGCAA  
GCCTTATCCTTTTTTATACTTCTTTTTCTAGAACAGCTTTATTGAGTTATAATTCACATACCATACATTTACCCATT  
TAAAATCTATAACTCAGTGGT  
TTTTAGTACAATTTTAGAACATATCACCCCAAAGAAACCCATCCCCATTAGCAGTGACTCCTCATTACCCCTTTC  
CCCACTCCAGTAATGCTGAGCAT  
CTTTTTATGTGTTACTGGCCATTTGTATATCTTCTTTGGAGAACTGTCTATTCAAACATTTGCCCATTTTTTATTGG  
ATAATTTATCTTTTTATTGAAT  
TATAAAAATTATATCTTCTAGGTACAAGTCACTTATGAGATATATGACTTGCAAATATTTTTACTCATTCTGTGAGT  
TTGTCTTTTCACTTTCTTGAGCA  
AGCCTTATCTTGAGCAAGCCTTATTCTTAACTCTGGTTAAATTTTGGTTTTCTGAACATTTTGCTAAATTTTCTAAA  
GGTCTCCAGTTGTGGAGTAGCCT  
GAAATGAATACTTTATTTTTTGAAGTGGAAGTAACTCTAATTTAAGTCTATAATAGAAAAGAGTAATTGCTATA  
GGAACCAACATTTTTTTTGGTTGT  
TTACCCTGTGCCAAGCATTGTGCTGAGTGCTTTGCAACCCTAAAATATAGGGTTGTTATTCCTGTTACATAGATGA  
AGAAACAAAGGCTTGAAGAAGTTA  
AGTAAGTTGCACAAAATCAAAAAGTGCTAATAAATGGTCAATTGGAATGACAATGCAAGTTATTTTTTTTGACCA  
ACTTTCACATTTTATGTGGGGTGTT  
TGTATTCCACTATTGTACTTTTTAAGACAATATTATTTTACAGCCTCATTTTCCATCCACTACAATCCCTAGATTTTT  
TCAGGCATATTGGAACCGAATC  
ATTCATTCTGTCTCTTTAAACTGGAATTTCTTGGTTTCTGAACCAAGAAATTTGATTTGATTTCTATAGCCTTTT  
GATTTTTTGACTGTCCCTTCCAA  
ATTGGCCCTTTTTAAACAAATTATAACAATTTCAAAGCTAAAAAAATTCTAAAGGTAACTTGCCATTAAATGA  
TTAAATGTAAATTGCAGACCATT  
TAAATGTCTTTTTGATTAGCTACTATTAAGAATTTAGCTGATCTACTTCATTAAATGTTATATTAGCTCAATGAAT  
AATTATATATTTCTGTGTAACAA  
ATTACCCCCAAACTTTAATGGCTTAAAGCAGCATTGTATTTCACAGTTTCTTTGGTTTAGGAATCCAAGTAGA  
GTTTAGCAATGTGGTTCTGCTCTA  
GGTCTCTCATGAGGTTTAAGTCAGGCTGCTGGCTGGTGCTGCAGTCTCATTTTTGGGGAGGATCTATGTGCAAGC  
TCACTCGTATGGTTGTTGGGAGGAT  
TCAGTTCCTTKTGAGCTTTTGGACTGAGACCTGCAATTCCTCACTGGCTGTTGGCCAGAGTTCTCCCTCAGTTCCTT  
GCCATATGCGCCTCTCTATGAGA  
CAGCACATGACATGGCCTCTGGCTCTGTCACAATAGCAACACAAGAGAGTGTCCAAGATGGAAGCCATAGTATC  
TTTGTAAACCTAATCTCAGAGTTGATA  
TATCATCACTTTTCCRTATTCTACTCATTAGAAGCAAATATTATGTTTCAGCCCACTCTCTATAGAAGGGAAATAC  
CAAGGGTATGAACACCAAAATAT  
GGAGCTTACTGAGGGCCATCTTAGAGGCTGCCTGTCACACTCAGTGAATCTATTGTTTAATCATGGAGAATCCAA  
AAGTCATGGTGTTTTTATATAGAAC  
RTTGATATGTACAGTTCCATCCTGCAAACAAGTGATAATTTTATCTGAAAATATATAAGGAAGTCTCTCCATCCTA  
AAATATGCATATGTGTATATGTAT  
TATTTTTAAAGCATCTTTAGTCTGCATATACATTTGCTACTTATTAATCTGTTTTATCAAGAGACAACATTAAAGTT  
GTAAGACACAATTAGAATAGAAA  
ATTTTATGTGTAACCTTTATGTTCTCCCTTGACTTATAAAGATTGCTATTGAATTTTAGTATTGGAAAACCTTGCC  
AATGTAAAGATGATTTTAATTTG

AACATGTACAACATGCAAAAGAATATTTTTCTCCTTAAATCATACTTTAGAAGTCAACCCTAATTTAACAAAATA  
GGTGATCCTGAATGGACTTTAATT  
CATTTATTCAGCAAGGCTTATTGAATTACTCTTCTAACTCATCATGAGTTACACTCATAATAAAAGTCTGCACTCCT  
AATTTTGGCCTCCCATAGCTCAR  
TCAATCTCTGTCTATGTTTCCAGTTTCATTTCTATTGCCATCTCCCTTAACCTTCCACTCCAGCACAGTTTTTCTTGC  
TCCTCCCCATGTTGATGCTCA  
TGCATGCCTCAGGGCTTTTCCATCTTCCTGAAAACCTCTTACCCAGATCGTTACATGGCTCGCTTCACTTCATTTAC  
TTCTCTGCTCCAGCATCACTACC  
ATAAAGAAGCATCCCACCAGCACTCTTCACAATACAGCAGGGACCATCAAAAAGCAAGAGACACAATAAAAATG  
ATAGTTGAAAAGATTAATCTCATGAC  
AGTGTACAAGATAACTTGTACGTAAACAGGCTGGATTAGACATGGATGAGTGGGGGGCCACAATTAGGAACTTTC  
ACATTCATTCAGGTGTGAATTGATAA  
TGATCCAGTTATGACCACCTTGGGCATGAGTAAAGCTACACAAACATTCTCCCAAACCAGGTCTTCCCTCGGAT  
TTGTGATGCCGGTAAATTGTTGCCA  
CCTTCCTCCATCCTCCTCTAACAGAGCCATTTCAGAACCAAAACCTTTTTAAAGAAAAGTTTGAAAAAATAAGAAG  
CACCAAAGAATTAGCAGGAAATGG  
TTAGGTTTTAGGACTAGAAACCACAATGCAGTAGGGGCTCTGCAAGAGGAAGGAGTGAGTGCTTCAAAGTGAC  
ATAGCAGCTGTGTGCTGTTGTGCTGA  
GACTCCTTCCAATCCTAGATGTTTCTAGTGGTGCTGGCAAAGCTTCTGTTGGCTGTACCACATTCTTTCCCAATCC  
AACTCCCTCTATGTAAAGTCTCA  
TAATTCTAAGGATCCTGGGTTCAAATGTAAGTTTCTTCCGTAGTAGTTACCATTTGGGTGTTTCTGCTTCTTCTAAA  
CTCCCTTCTAATTGTTTTCTC  
TTGTTTTTAAGCTAGTGATTTTTTCCCCTGTATTTTACATATGCTTATGGAAAGAACAGATATRTTTTCATGGCACT  
TAGCATGTTATAAATGCATTGTT  
TTCCATGGAGTAAGATCATTCTACCACTGAGAGCATTAAATCCATGTTTCTAAGGTAACCCATGTTTTCTAGAAA  
CAGGGGAAGAGTCATGACCAACTT  
TCACATTGTCTGCTTTGTTCTTGGGGTATCATCCATTGTGATAACCATTCCCAAGAAGTTTCCTCCCCAAAACAGA  
ATTACTTCTCTGCCTACCTTGAT  
TCAAGTGACCTTTCATTACAGATAATCTGGTTTTATAGGGAAATCTAGTAGGTTCTGCACTCAGTTGTAAAGAAA  
GACAGATGCAAATGGGTTTCAGTAA  
TTGGTACAACATTGTCACAAAGCTCTTTAGATTAGCAAGTACTTATAAACTAGGTGGAATATTTTTTATTCCATGT  
TGTTAATAAGAATGCCCTAATCCT  
CTTTACACACTTGAAAATATCCAGGGAATGCAGAGAGAGACAGACAGAGAGATAGACTCGTATGTAATCATGGA  
CTTTTGAGATAGGTAGGTATAGATGT  
TCCTGTAAACTAAGCGTTTGTAGCCTCCCAAATTTATATGTTGAAATCCCTACCATACTGTGATATGGAGCCTTT  
GGTATGAGGTCAGAGGCTTCATGC  
ATGGGATTAGTGCTCTCATAAAATAGGCCCTGGAGAGCTTCCTCACCCCTTCTGCCATTTGAAGACACAGTGAGA  
AGATGAATATCTGTGAACCAGGAAG  
CAAACCCTCACCAGACATTAAATCTGCCAGTGCTTTGATCTAGTTCTTCTCAGCCTCCAGACTGTGAGGAATAAAT  
TTCTGGTTTTTATAAGCCATCCTG  
TCTATACAGTTTGTACAGCAGCCTGAACAGACAGAGACAAGTGTAATCCCATCTTTTCTTAGCTGTATTATTAG  
GTTAGTGCAAATGTAATTGCTGTT  
TTCGCCATTTTAATGGCAAAAATCACAGCTACTTTTGCACCAACATAATTTGCTCACTTTTCTCTGTTGTAATTTT  
CTCATATGAGACTGTAGGAAAG

AATGATTTAATATATTCTATAAAGTGCCTAGTAATATAATAGGTCCAGCATATGAAAGCATTTCATAATAGGTTCTC  
 AGTAAAGAGTGGCAATGCTCGAGC  
 ATTTCTAAAGAAACGGAATTGATTATATTCCATTCTTTTTTCAGACAGTTATAAATGTTAGAAATTTCTTTCTGTTA  
 CTGCTCCAGAATCTGTGCTTTTC  
 AATTATTGGTTCTAATTCAGTATTTTTAGGCCTCAGAGAAAAAGCCTTAATGCCGTCYTCACGTGGTAGGCTACCA  
 GGTCTTTGAAGATGGCCATTTTGT  
 ATCCCACATCTTCTCCACATCTGCTCATAGAAATACAGTTTCTTCATTTGCTCCCTTATGGTTTAGGTTCTGAAGCAA  
 CTCAGAATGGAACACAACGCTAC  
 TGGTGTGTTAATTAGCAGGTGAGAATGGAACCTATCACCTTTTTTGTGTTCTTAACACAACCTTAGTTGGCATG  
 AATTTGAGGAGCAGTTATATTTCA  
 CCTTTGACTTGTCAAACCTTACACTTGACTAAAACCTACAGATGTTTTTCTTTTAATTTTTTTAATTTTTAATTTTTTR  
 TGGGTACATAGTAGGTGTATA  
 TATTTATGGGGTATGTGACATGTTTTGATACAGGTGTGCAACATGAAATAATTACATCATGAAGAATGCARTATC  
 CATCCCTTCAAGGATTTATCCATTG  
 AGTTGCAAACAATCCAATTATTACACTCTTTAAGTTATTTTAAAATGTTACAGATGTTTTTCAAACATGTCTCTGTT  
 CTACCCTATCTCTCTCATCATAT  
 ATWTGTATGGGATTTTACATTTTGACATATTAAATTTTCATCTTGTTGGAGTTTGCTCGTTTTCTTCTTACCTATGAA  
 GCTGTTTTGGGACCCTAACTATG  
 ACAGTAAGTCACTTGGCTATGCACTTAATCAGCATGTGTTTAATATCYACACTCACATTTTTATAAATATTTAACAG  
 GAAGGGAAATTTATACCTCCAAA  
 CCCACTCTAATCCCTGAACTCCAAATGCATGTAACCAACTGCCTGTAAATATCTTCACTTAGATGATTCCTGTCAG  
 TTCACACTTAACCAAACTGATTT  
 CCACCCTCAACCTTCTCCTGTAGTCTTCTCATCCTRGTCACAGTGAAGCATATCCGTCCTTCACATTGCACAGCCAA  
 AAACCTTCCTTGCAGTCATCCTT  
 GACCCATTTTGTTTCCTCCCACATCCCATATCTAATCCCTCAGCAAACCTCTACTTCACAATGCATCCTGAGTCATAC  
 CTTTCTTACTTCTTACCTTCTTG  
 CACTACCACTTTGGACCAAATCATCAGAAATAGATTCCTAACATTTTCTGTATTTACCCTTACATTCTTTAGTCTTTT  
 TAAATATTGACGCTAGAATGA  
 ACCTGTTAAGACTTAAATTATGTCACTCCTACTAAAAATACAAAAATTGGCCAGGTGTGGTGAGGCATGCCTGTA  
 GTCAGTCCCTGCTACTCGGGAGGCT  
 GAGTTGGGAGAATCACGTGAACCCAGGCAGCRGAGGTTGCAGTAAGCAGAGATTGTGCCACTGCACTCCAGCC  
 TGGGCGACAGAGTGAGACTCCCTCTCA  
 GTTAAAAAAAAAAAAAAAAAAAAAAAAATCACCTTAATGCCTTTTATCTCTGGTTACTTCAACACCATAATTACCAT  
 AAATAATTCATCTCTGTTGGAATA  
 TTTGTGTCCCATTAATAATTTCGTATGTTGAAATCTCTTCACTAATGTAATGATATGAAAATCAGCCTTTATGGGGGT  
 GATTAGGTCAAGAAGGCAGAGCCC  
 TCATGAATGGAATTACTGCTCTTATAAAAAGACAAAGGACTTTGCAGGGACGTGGATGAAGCTGGAAACCATCA  
 TTCTCAGCAAATAACCCAGGAACAG  
 AAAACCGAACACTGCATGTTCTCACTCATAAGTGGGAGTTGAACAATGAGAACACACGGACACAGGGAGGGGA  
 ACATCACACATCGGGGCCTTTTCGGGGG  
 GTGGGGGACAAGGTGAGGGATAGCATTAGGAGAAATACCTAATGTAGATGACAGGTTGATGGGTGCAGCAAA  
 CCACCATGGCACATGTATACCTATGTAA  
 CAAACCTGCACGTTCTGCACATGTATCCCAGAACTGAAAAGATAATAAAAAAAAAAATCACGCCACTCCTCTGTT  
 CAGAACTTACCAGTAGCTTTCCATC

TCACTCAGAGTCAAAGCTAAAGCCCTGACAATTGCCTGCACAACTCTACATGGTTCTGCCTCCCCTTCCCACACAT  
 GTACACAAACACATTTGCTTTTCT  
 GACCTCATTTTCTAGCCTTCTACCCCTTGCTTCCTCCTCTGCATGCACAGAGTCCTCTTTGCTGTCCCCTGAACATGC  
 CAGGCATACTCCCACCTTAGGGC  
 CTTTGCATCTACTGTTGCGCTGTCTGGAGTGCACCTTCTTCATCCCATGCACTATGAATCTTCCTTCTGCTCTCATTT  
 TTGTTTAAGGTATTATATGAGA  
 CTGTGCTACTAATAGAAATTATAGGGCTCTTTTGGCTAGTGTCAAGTATAGATTTGCGTTTTGGACTCACATAACT  
 GAATATGTTAGAATTACATCTAAA  
 GAAATAATTGCCTATAAATGGAGTATTGTGTGTGAAGAGTTAGAGTCCTGCCCCATCTCCTTTCCTATTAAATGAT  
 AGTGATTTTTCACAGTATCTGAGAT  
 AGCCCTTTGAATATCAGAGAAACATAAGAGTGTAAGATTTCAGGAAATGGTCATATTTTTCAGATGAGAAAA  
 GTTCACCAAAGATTTTACCCATTGGT  
 GCCARAAATGTAAGGCAGGCAAGAAGTTAGTAGAAGAAAGTCATCACAAGTAAACAAWTATTTTTTTTAAGTTG  
 TAGGCACTGATTTCCAAATGAGTTCC  
 ATAAACCACTGCTATTTCCATTTGTAGTATAGCAATAGGTTGTGATACTAACCTTAAAGGTCTCAAGTTCAACAGC  
 TTATTATGAGGAAAGCTGCCATTC  
 CTTTAAGAAGGTGAGTTTCTATTATCTTTTTAAAGAGAGAATTACTCCTTTATAAAAAATGTAACCTTCTTCATACTGT  
 ATTTACAGATCTCTTGCTTTTGTAG  
 TAAACTTTTTCATTGCAGTGGGTTTGATAATAATAGTGCTCACTTTTAAAAAATTCCTAAATTTTAGTTTCCCCTC  
 CAGTCACCAGCTCACTTTTCTGT  
 GATCCTCTTTAGCAGAACTCCTTAAAAGACTCATCTATACTTGCTGTTTCCTCCTTCTCTCCTGCTGTTCTCTCTATA  
 ACCCACTCCAAATAGGCTTTCA  
 TCTCGTCGCTCCACTGAAACTACTTATCAAAGTTTCCATCACCGTCATGCTACCATATTATCTTCTCCATCCTCATCT  
 TATTCCCAGCAACATTTGACAT  
 CACCTTGGAACACTTTATTTACGTGATTTCACTGAGACCTCCTGTCTGGTTTTGTTTTTCATCTTATTTATTAGATG  
 CTAGTTCTCAGTCTCATTTGAC  
 TCTTTCTCTTCTCCCAAACCCTATACATTGAAATATCCCAAGGCCTAGACCCTCTAGTCTTTGGCCTAGGTGATCTT  
 TTTTATATATGTATTTTAAGTT  
 CTAGGGTACATGTGCACAACGTGCAGGTTTGTTACATATGTATACATGTGCCATGTTGGTTTGCTGCACCCATTA  
 ACTCGTCATTTATATTAGGTGATCT  
 TATTCAATCACATGGTATATAAATTACAATATAGGCTACTATGACAGAGGCCCCAGAATGAAGGTGCCTAAACAA  
 GATGAAAGTTGATTTCTTTTCTTA  
 TAAAAGTGTGGGAGCTTGTTCCAAGACTGATAGGATAGCTTGTTAGTACTGGAGACCCACGCTCTTTCTTTCTCCTGC  
 ACTCTGCTGTTCTCAACATGCTGTC  
 TCCATCTTCTGGTCCCAGATGGCTGCTTCAGCCCCTGCATCCTGTCAGTATTCTAACCAGCAGGAAGAAGAAGAT  
 AGGAAGTGAGAACGTATCCCTTCCC  
 TTTAGGACTCAGAAGTTAAACATGTTTCATCTGCTCTCATTTTATTGGCCAGAACTGAGTCACATSATGATAACTT  
 GCTGAAGAGAATTTAGGAAATATA  
 GTCTTTAGGTAGGTAGCTATGGCTTCCAGCTAAAATGGAGCAGTTCTTTTACTAGAGGAAAAAGAGCGTGATGG  
 ATTTTGGAGGACAGGGAGCAGTTTCT  
 GCTGCACATGGCTTTAAATACTATCTATGAGCTAATGTCTCCAGATTTGTAGCTATTTCTGAGTTCAGGACTCT  
 TTCCTCAGCTGCCTCCCTAACATT  
 CTCGAACACCTGCCTGTTCAAATGAAGTCTTCACTTTTCTCCCCAACTCCTACCCATCTTAGGTCTCCCAGGTG  
 CTCATGCCAGAACTTAGAACAG

## **DEMANDES OU BREVETS VOLUMINEUX**

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS  
COMPREND PLUS D'UN TOME.**

**CECI EST LE TOME \_\_1\_\_ DE \_\_3\_\_**

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

---

## **JUMBO APPLICATIONS / PATENTS**

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE  
THAN ONE VOLUME.**

**THIS IS VOLUME \_\_1\_\_ OF \_\_3\_\_**

NOTE: For additional volumes please contact the Canadian Patent Office.

**Claims:**

1. A method for determining a human's risk for myocardial infarction (MI), the method comprising testing nucleic acid from said human for the presence or absence of a polymorphism in gene *lipoprotein(a) (LPA)* at position 101 of SEQ ID NO:78 or its complement, wherein the presence of G at position 101 of SEQ ID NO:78 or C at position 101 of its complement indicates that said human has an increased risk for MI.
2. The method of claim 1, wherein said nucleic acid is a nucleic acid extract from a biological sample from said human.
3. The method of claim 2, wherein said biological sample comprises blood, saliva, or buccal cells.
4. The method of claim 2 or 3, further comprising preparing said nucleic acid extract from said biological sample prior to said testing.
5. The method of any one of claims 1 to 4, wherein said testing comprises nucleic acid amplification.
6. The method of claim 5, wherein said nucleic acid amplification is carried out by polymerase chain reaction (PCR).
7. The method of any one of claims 1 to 6, further comprising correlating the presence of said G or said C with an increased risk for MI, or correlating the absence of said G or said C with no increased risk for MI.
8. The method of claim 7, wherein said correlating is performed by computer software.
9. The method of any one of claims 1 to 8, wherein said testing is performed using sequencing, 5' nuclease digestion, molecular beacon assay, oligonucleotide ligation assay (OLA), size analysis, single-stranded conformation polymorphism (SSCP) analysis, or denaturing gradient gel electrophoresis (DGGE).
10. The method of any one of claims 1 to 8, wherein said testing is performed using an allele-specific method.

11. The method of any one of claims 1 to 10, wherein said human is homozygous for said G or said C.

12. The method of any one of claims 1 to 10, wherein said human is heterozygous for said G or said C.

13. The method of claim 10, wherein said allele-specific method is allele-specific probe hybridization, allele-specific primer extension, or allele-specific amplification.

14. The method of claim 10 or 13, wherein said allele-specific method is carried out using an allele-specific primer comprising SEQ ID NO:90 or SEQ ID NO:91.

15. An allele-specific polynucleotide for use in a method as defined in any one of claims 1 to 12, wherein said polynucleotide comprises a segment of SEQ ID NO: 78 or its complement at least 16 nucleotides in length that includes said position 101.

16. An allele-specific polynucleotide for use in a method as defined in any one of claims 1 to 12, wherein the allele-specific polynucleotide is a primer that comprises SEQ ID NO:90 or SEQ ID NO:91.

17. The allele-specific polynucleotide of claim 15 or 16, wherein said polynucleotide is detectably labeled.

18. The allele-specific polynucleotide of claim 17, wherein said polynucleotide is labeled with a florescent dye.

19. A kit for use in a method as defined in any one of claims 1 to 12, wherein said kit comprises at least one allele-specific polynucleotide as defined in any one of claims 15 to 18 and at least one further component, wherein the at least one further component is a buffer, deoxynucleotide triphosphates (dNTPs), an amplification primer pair, an enzyme or any combination thereof.

20. The kit of claim 19, wherein said enzyme is a polymerase or a ligase.

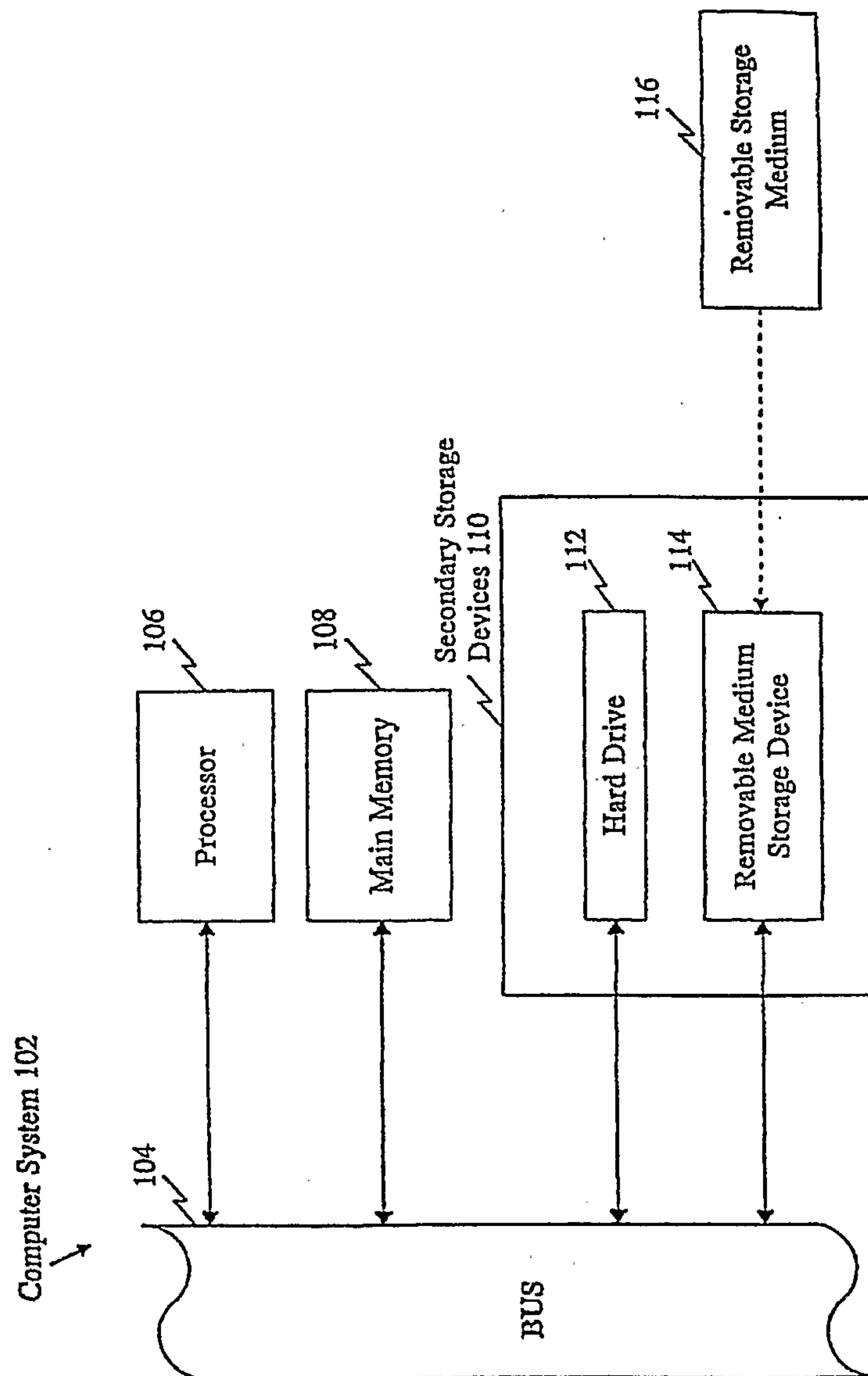


FIGURE 1

Computer System 102

