



US 20120003644A1

(19) **United States**(12) **Patent Application Publication**
Skorecki et al.(10) **Pub. No.: US 2012/0003644 A1**(43) **Pub. Date: Jan. 5, 2012**(54) **METHODS AND KITS FOR DETERMINING
PREDISPOSITION TO DEVELOP KIDNEY
DISEASES**(75) Inventors: **Karl Skorecki**, Haifa (IL); **Shay
Tzur**, Haifa (IL); **Saharon Rosset**,
Ramat-HaShavim (IL); **Walter
Gabriel Wasser**, Jerusalem (IL);
Doron M. Behar, Hof HaCarmel
(IL); **Revital Shemer**, Haifa (IL)(73) Assignee: **Rappaport Family Institute for
Research in the Medical Sciences**,
Haifa (IL)(21) Appl. No.: **13/153,569**(22) Filed: **Jun. 6, 2011****Related U.S. Application Data**(60) Provisional application No. 61/351,960, filed on Jun.
7, 2010.**Publication Classification**(51) **Int. Cl.**
C12Q 1/68 (2006.01)
G01N 33/566 (2006.01)
C07H 21/00 (2006.01)
C07K 16/18 (2006.01)
(52) **U.S. Cl.** **435/6.11**; 530/387.9; 436/501;
435/7.92; 435/7.1; 536/23.1(57) **ABSTRACT**

Provided are methods and kits for determining predisposition of a subject to develop a kidney disease, by identifying in a sample of the subject at least one APOL1 polypeptide variant which is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions; or at least one APOL1 nucleotide mutation in the APOL1 genomic sequence set forth in SEQ ID NO:3, wherein the at least one nucleotide mutation or polypeptide variant being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, wherein presence of the APOL1 polypeptide variant indicates increased predisposition of the subject to develop the kidney disease.

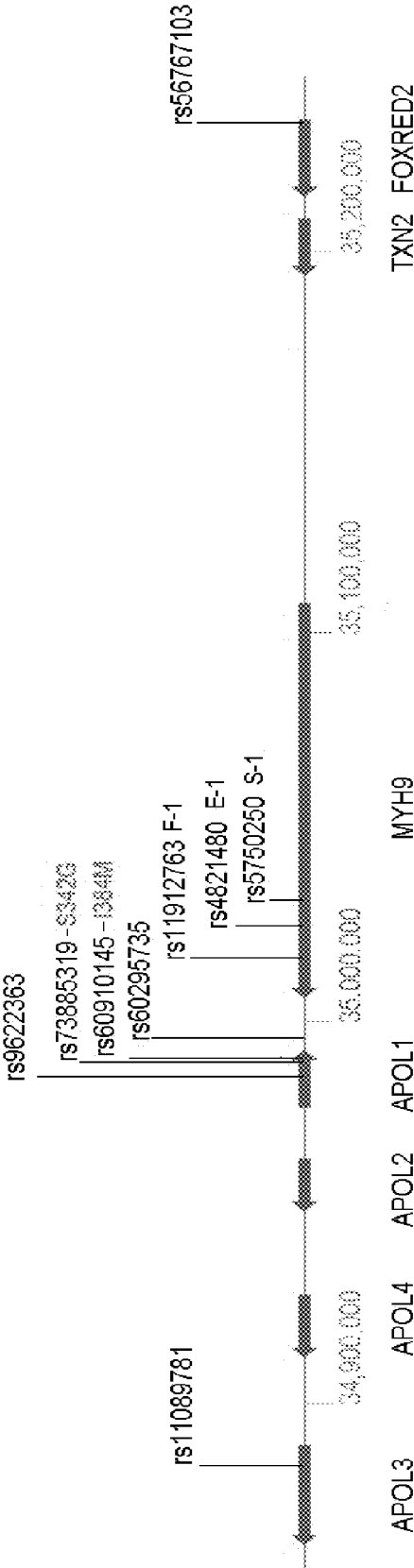


Fig. 1

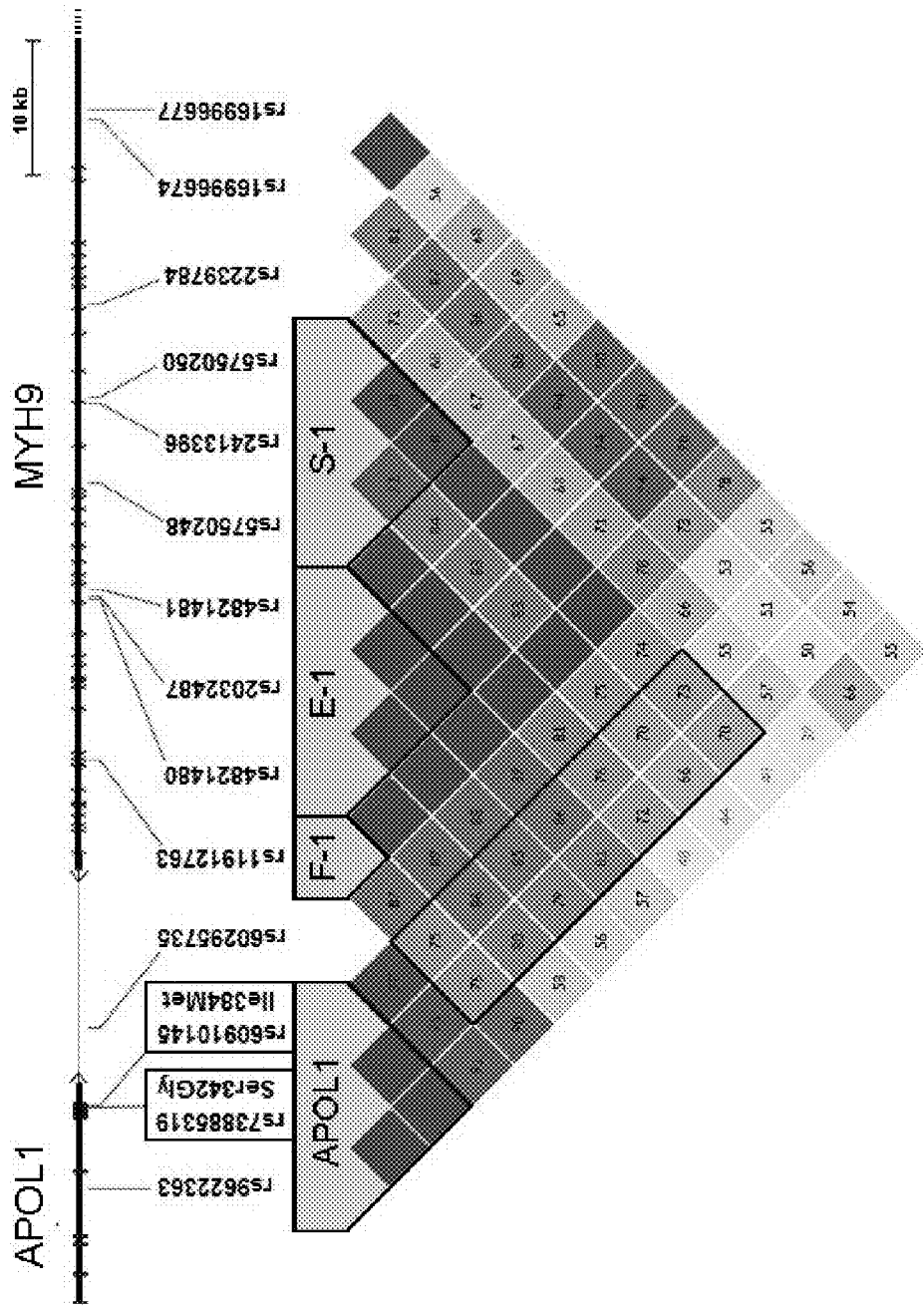


Fig. 2

Fig. 3A

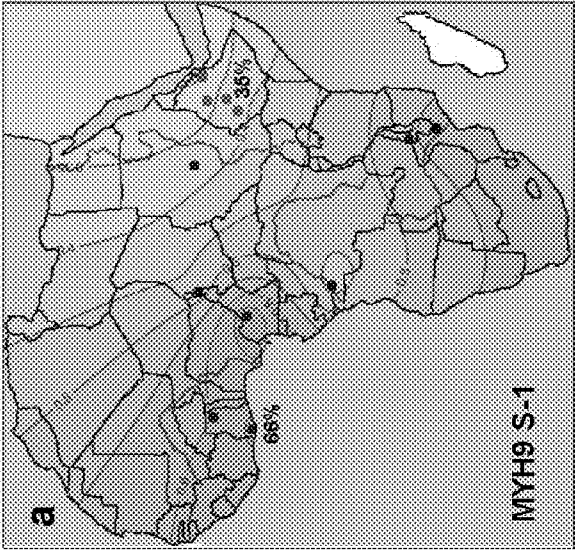


Fig. 3B

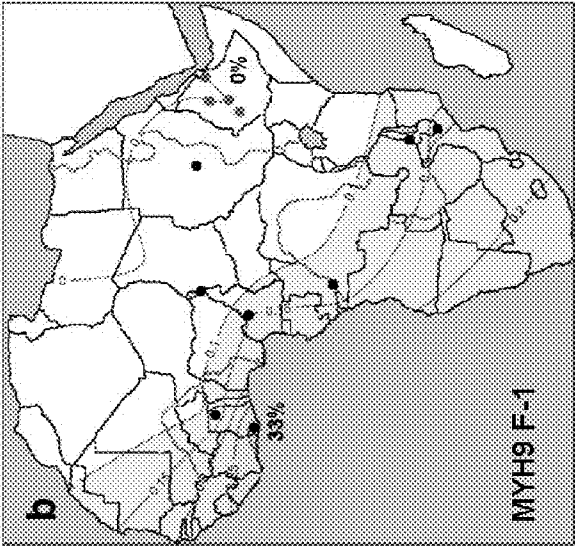


Fig. 3C

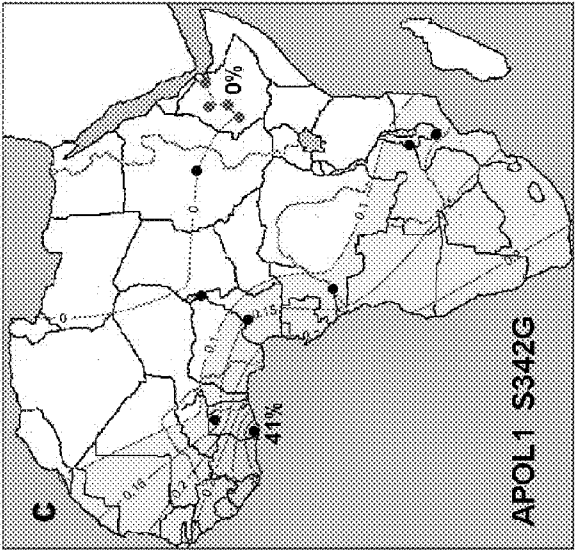


Fig. 4A

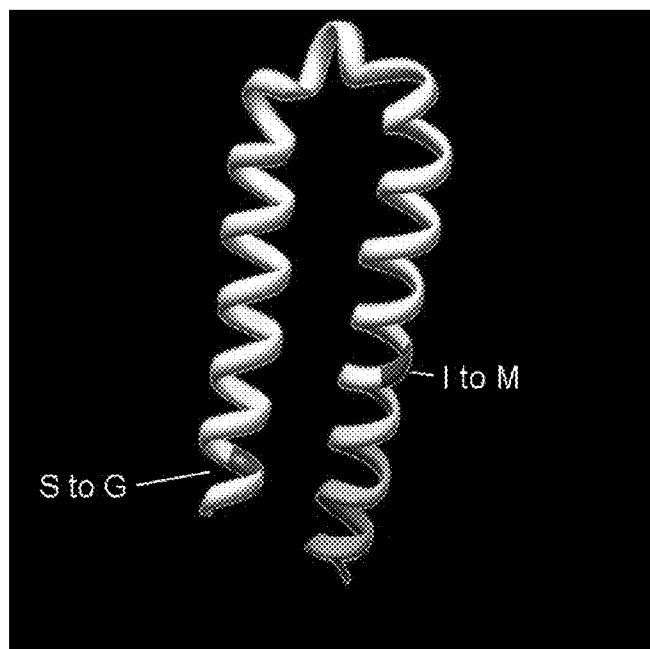


Fig. 4B

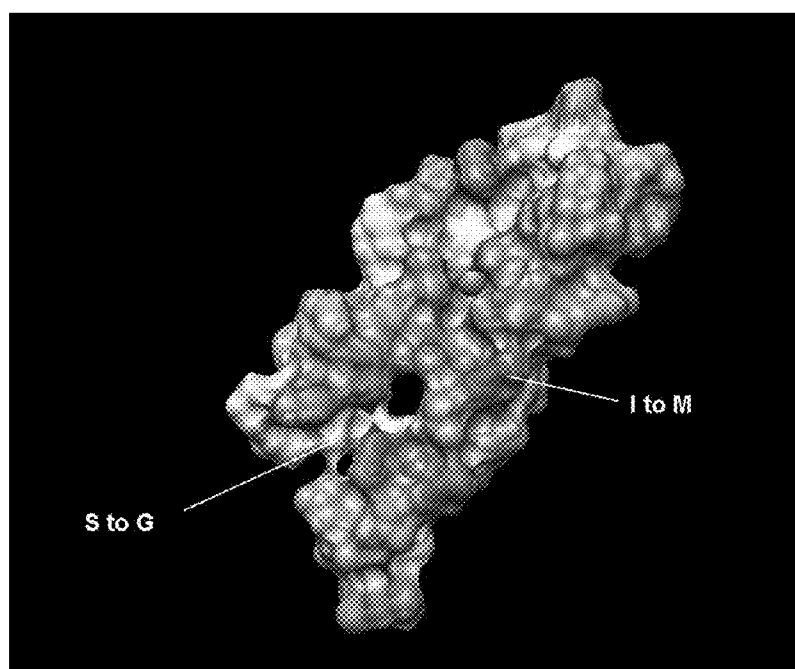
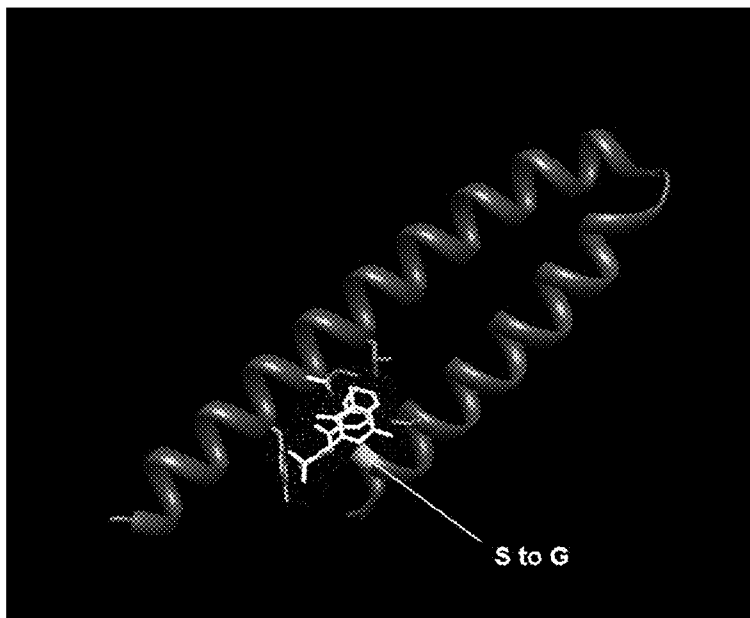
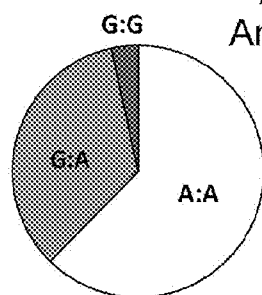


Fig. 4C



Figs. 5A

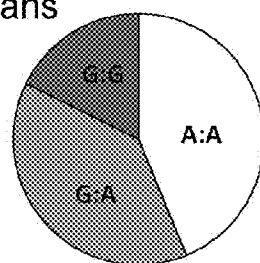
Controls



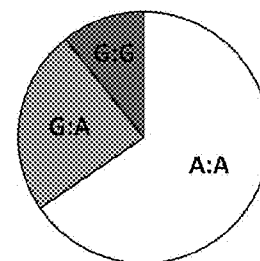
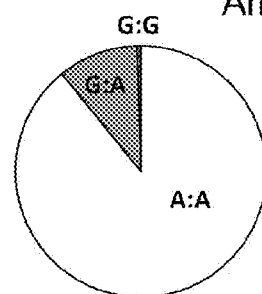
African
Americans

Figs. 5B

Cases



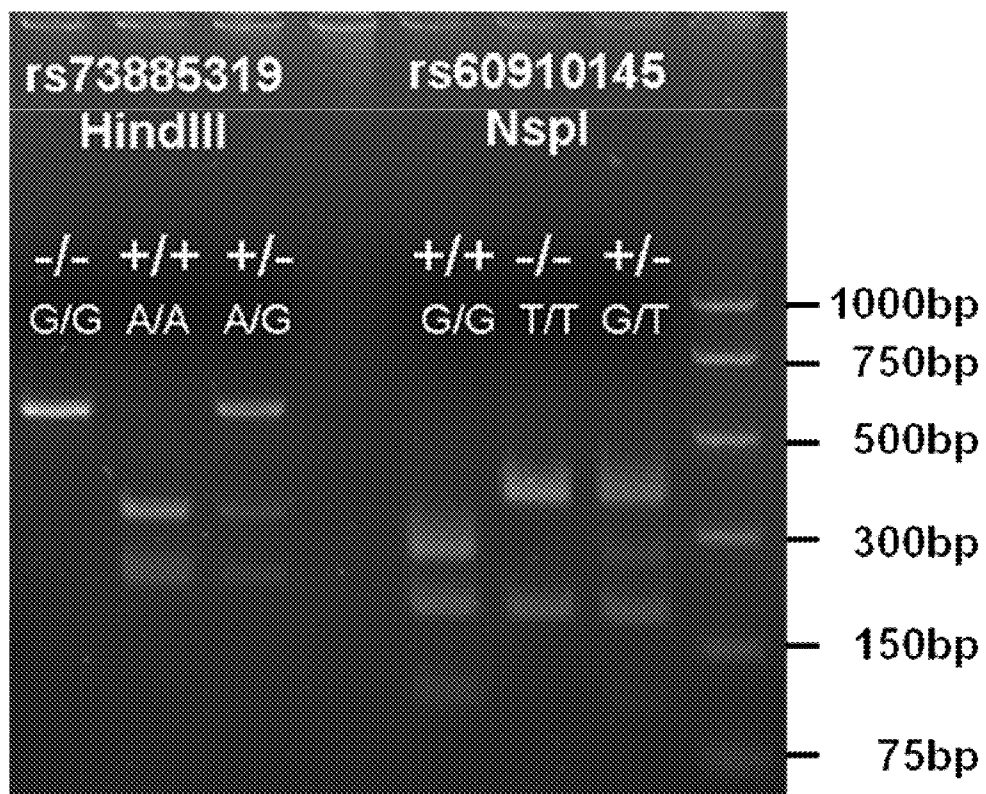
Hispanic
Americans



Figs. 5C

Figs. 5D

Fig. 6



METHODS AND KITS FOR DETERMINING PREDISPOSITION TO DEVELOP KIDNEY DISEASES

RELATED APPLICATION

[0001] This application claims the benefit of priority under 35 USC 119(e) of U.S. Provisional Patent Application No. 61/351,960 filed Jun. 7, 2010, the contents of which are incorporated herein by reference in their entirety.

FIELD AND BACKGROUND OF THE INVENTION

[0002] The present invention, in some embodiments thereof, relates to methods and kits for determining predisposition to develop kidney diseases, and, more particularly, but not exclusively, to methods of designing life style change and treatment regimens for a subject predisposed to develop a kidney disease.

[0003] Chronic Kidney Disease (CKD) is a powerful independent risk factor for cardiovascular disease and death at all stages. Most patients with CKD will succumb to cardiovascular complications rather than reach end-stage kidney disease (ESKD). Recent reports estimate that as many as 600 million people worldwide have CKD, and hence are at greatly increased risk of cardiovascular disease including hypertension [Hypertext Transfer Protocol (dot) world wide web (dot) kidney (dot) org (dot) au/NewsEvents/WorldKidneyDay/tabid/655/Default (dot) aspx].

[0004] Based on information from those countries where renal replacement therapy is available and registries are maintained, there are close to one million people with ESKD receiving dialysis treatment or a kidney transplant. In the United States, an estimated 15.5 million have stage 3 and 4 CKD, and 584,000 have ESKD. African Americans have five times greater risk for kidney disease than European Americans. Correspondingly, African-Americans also have a higher risk of mortality from cardiovascular disease than Americans of European ancestry, including death, non-fatal myocardial infarction, and stroke. Finally, it is now appreciated that hypertension, once thought to be a leading cause of CKD leading to ESKD, is the consequence of primary renal glomerular disease in certain risk populations.

[0005] Mapping by admixture linkage disequilibrium (MALD) localized an interval on chromosome 22, in a region that includes the MYH9 gene, which was shown to contain African ancestry risk variants associated with certain forms of ESKD (Kao et al. 2008; Kopp et al. 2008). This led to the new designation of MYH9 associated nephropathies. Subsequent studies identified clusters of single nucleotide polymorphisms (SNPs) within MYH9 with the largest odds ratios (OR) reported to date for the association of common variants with common disease risk (Bostrom and Freedman 2010). These MYH9 association studies were extended to earlier stage and related kidney disease phenotypes, and to population groups with varying degrees of recent African ancestry admixture (Behar et al. 2010; Nelson et al. 2010). Thus, the MYH9 has been proposed as a major genetic risk locus for a spectrum of non-diabetic ESKD. However, despite intensive efforts including re-sequencing of the MYH9 gene no suggested functional mutation has been identified (Nelson et al. 2010).

[0006] U.S. Patent Application No. 20100297660 to Winkler Cheryl et al. ("SINGLE NUCLEOTIDE POLYMOR-

PHISMS ASSOCIATED WITH RENAL DISEASE") discloses methods for determining the genetic predisposition of a human subject to developing renal disease, such as focal segmental glomerulosclerosis (FSGS) or end-stage kidney disease by detection of one or more haplotype blocks comprising at least two tag single nucleotide polymorphisms (SNPs) in a non-coding region of a MYH9 gene or detecting the presence of at least one tag SNP in a non-coding region of a MYH9 gene.

[0007] Additional background art includes Tzur S, et al. (Missense mutations in the APOL1 gene are highly associated with end stage kidney disease risk previously attributed to the MYH9 gene) Hum. Genet. 2010, 128:345-50; Rosset S, et al. (The population genetics of chronic kidney disease: insights from the MYH9-APOL1 locus) Nat. Rev. Nephrol. May 3, 2011. [Epub ahead of print]; Genovese G, et al. (Association of trypanolytic ApoL1 variants with kidney disease in African Americans) Science 2010; 329:841-5; Behar D M, et al. (African ancestry allelic variation at the MYH9 gene contributes to increased susceptibility to non-diabetic end-stage kidney disease in Hispanic Americans) Hum. Mol. Genet. 2010 19(9):1816-27; Shlush L I, et al. (Admixture mapping of end stage kidney disease genetic susceptibility using estimated mutual information ancestry informative markers) BMC Med Genomics. 2010 3:47; and U.S. Patent Application No. 20110030078.

SUMMARY OF THE INVENTION

[0008] According to an aspect of some embodiments of the present invention there is provided a method of determining predisposition of a subject to develop a kidney disease, comprising: identifying in a sample of the subject at least one APOL1 polypeptide variant which is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions, wherein presence of the APOL1 polypeptide variant indicates increased predisposition of the subject to develop the kidney disease, thereby determining the predisposition of the subject to develop the kidney disease.

[0009] According to an aspect of some embodiments of the present invention there is provided a method of determining predisposition of a subject to develop a kidney disease, comprising: identifying in a sample of the subject at least one APOL1 nucleotide mutation in the APOL1 genomic sequence set forth in SEQ ID NO:3 or at least one APOL1 polypeptide variant, wherein the at least one nucleotide mutation or polypeptide variant being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, wherein presence of the APOL1 nucleotide mutation or the APOL1 polypeptide variant indicates increased predisposition of the subject to develop the kidney disease, thereby determining the predisposition of the subject to develop the kidney disease.

[0010] According to an aspect of some embodiments of the present invention there is provided a method of designing a life style change to a subject with the risk of kidney disease, comprising: (a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation, (i) wherein the variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the

Trypanosoma brucei rhodesiense under identical assay conditions; (ii) wherein the APOL1 nucleotide mutation is included in the APOL1 genomic sequence set forth in SEQ ID NO:3, and wherein the at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, wherein presence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation indicates increased predisposition of the subject to develop the kidney disease, and; (b) designing the life style change based on presence or absence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation, thereby designing life style change to a subject with the risk of kidney disease.

[0011] According to an aspect of some embodiments of the present invention there is provided a method of designing a life style change to a subject with the risk of kidney disease, comprising: (a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation, (i) wherein the variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions; (ii) wherein the APOL1 nucleotide mutation is included in the APOL1 genomic sequence set forth in SEQ ID NO:3, and wherein the at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, wherein presence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation indicates increased predisposition of the subject to develop the kidney disease, and; (b) designing the life style change based on presence or absence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation, thereby designing life style change to a subject with the risk of kidney disease.

[0012] According to an aspect of some embodiments of the present invention there is provided a method of designing a nephrotoxic anti-retroviral treatment regimen to a subject infected with HIV, comprising: (a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation, (i) wherein the variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions; (ii) wherein the APOL1 nucleotide mutation is included in the APOL1 genomic sequence set forth in SEQ ID NO:3, and wherein the at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, wherein presence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation indicates increased predisposition of the subject to develop the kidney disease, and; (b) designing the nephrotoxic anti-retroviral treatment regimen based on presence or absence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation, thereby designing a nephrotoxic anti-retroviral treatment regimen to the subject infected with HIV.

[0013] According to an aspect of some embodiments of the present invention there is provided a method of determining if a subject is suitable for donating a kidney for transplantation, comprising: (a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation, (i) wherein the variant is characterized by a

higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions; (ii) wherein the APOL1 nucleotide mutation is included in the APOL1 genomic sequence set forth in SEQ ID NO:3, and wherein the at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, wherein presence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation indicates increased predisposition of the subject to develop a kidney disease, and; wherein increased predisposition of the subject to develop the kidney disease indicates that the subject is not suitable for donating the kidney for transplantation, thereby determining if the subject is suitable for donating a kidney for transplantation.

[0014] According to an aspect of some embodiments of the present invention there is provided a kit for determining predisposition to a kidney disease, comprising a reagent capable of specifically detecting at least one APOL1 nucleotide mutation in the APOL1 genomic sequence set forth in SEQ ID NO:3, wherein the at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1.

[0015] According to some embodiments of the invention, the APOL1 polypeptide variant comprises a mutation in the C-terminal helix of the APOL1 polypeptide.

[0016] According to some embodiments of the invention, the APOL1 polypeptide variant has a reduced binding ability to serum resistance-associated protein (SRA) expressed by *Trypanosoma brucei rhodesiense* as compared to the binding ability of the APOL1 wild type polypeptide under identical assay conditions.

[0017] According to some embodiments of the invention, the APOL1 polypeptide variant comprises the amino acid sequence set forth in SEQ ID NO:1 with a mutation selected from the group consisting of G342, M384, N388-del and Y389-del.

[0018] According to some embodiments of the invention, the LD between the APOL1 nucleotide mutation and the S342G mutation is characterized by a Lewontin correlation coefficient (D') higher than 0.5.

[0019] According to some embodiments of the invention, the significance of the LD is characterized by $LOD \geq 2$.

[0020] According to some embodiments of the invention, the APOL1 nucleotide mutation is selected from the group consisting of the guanine-containing allele of single nucleotide polymorphism (SNP) rs73885319 (SEQ ID NO:6), the guanine-containing allele of SNP rs60910145 (SEQ ID NO:7), the guanine-containing allele of SNP rs9622363 (SEQ ID NO:8), the adenine-containing allele of SNP rs60295735 (SEQ ID NO:9) and the cytosine-containing allele of SNP rs58384577 (SEQ ID NO:10).

[0021] According to some embodiments of the invention, the kidney disease comprises end stage kidney disease.

[0022] According to some embodiments of the invention, the kidney disease comprises HIV-associated nephropathy (HIVAN).

[0023] According to some embodiments of the invention, the subject is infected with HIV.

[0024] According to some embodiments of the invention, the APOL1 nucleotide mutation is selected from the group consisting of the guanine-containing allele of single nucleotide polymorphism (SNP) rs73885319 (SEQ ID NO:6), the

guanine-containing allele of SNP rs60910145 (SEQ ID NO:7), the guanine-containing allele of SNP rs9622363 (SEQ ID NO:8), the adenine-containing allele of SNP rs60295735 (SEQ ID NO:9) and the cytosine-containing allele of SNP rs58384577 (SEQ ID NO:10).

[0025] According to some embodiments of the invention, the reagent comprises a polynucleotide capable of specifically detecting the APOL1 nucleotide mutation.

[0026] According to some embodiments of the invention, the reagent comprises an antibody capable of specifically binding an APOL1 variant comprises the nucleotide mutation and not to the APOL1 wild type polypeptide.

[0027] According to some embodiments of the invention, the method further comprising informing the subject on the state of the predisposition to develop the kidney disease.

[0028] According to some embodiments of the invention, the nucleotide mutation creates an APOL1 protein variant.

[0029] According to some embodiments of the invention, the nucleotide mutation is detected by a DNA detection method.

[0030] According to some embodiments of the invention, the APOL1 protein variant is detected by a protein detection method.

[0031] Unless otherwise defined, all technical and/or scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of the invention, exemplary methods and/or materials are described below. In case of conflict, the patent specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be necessarily limiting.

BRIEF DESCRIPTION OF THE DRAWINGS

[0032] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0033] Some embodiments of the invention are herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of embodiments of the invention. In this regard, the description taken with the drawings makes apparent to those skilled in the art how embodiments of the invention may be practiced.

[0034] In the drawings:

[0035] FIG. 1 is a schematic view of the chromosomal region encompassing the SNPs according to some embodiments of the present study.

[0036] FIG. 2 is a linkage disequilibrium (LD) plot of the non-diabetic ESKD associated SNPs in the APOL1 and MYH9 region with their physical locations on chromosome 22. LD was calculated based on the African American control samples (n=140) and the LD plot was generated using the program HaploView (Barrett et al. 2005). The color scheme represents the pairwise linkage disequilibrium value (D' /LOD) for the 4 new SNPs outside MYH9 (2 of which are missense mutations in APOL1) and for the previously published 10 MYH9 SNPs described in Behar et al. 2010. Bright

red squares presents SNPs with linkage $LOD \geq 2$ and $D'=1$. D' values lower than "1" are expressed in percentages;

[0037] FIGS. 3A-C are contour map depicting spatial allele frequency distributions in Africa of the ESKD risk variants. FIG. 3A—MYH9 S-1 (SNP rs5750250); FIG. 3B—MYH9 F-1 SNP (rs11912763); FIG. 3C—APOL1 S342G missense mutation (rs73885319). Maps were generated based on genotyping of 12 African populations (n=676) (Table 3, in the Examples section which follows), using Surfer V.9 (Golden Software). Populations locations are marked (red circles for Ethiopia). Risk allele frequencies in Ethiopia and in South-Ghana are indicated;

[0038] FIGS. 4A-C are predicted peptide structures of the C-terminus domain of the APOL1 gene product (amino acid positions 339-398 in SEQ ID NO:1) that contains the missense mutations S342G and I384M. All predictions were generated using the program I-TASSER (Zhang 2008, 2009), structures were edited with the program CHIMERA (Pettersen et al. 2004). All suggested predicted structures exhibit T_m values >0.5 . FIG. 4A—Predicted structure and location of amino acid changes. C-terminus domain is predicted to have a bent alpha-helix structure. The mutation I384M is located on the external surface of the predicted alpha-helix, while the S342G is buried inside. FIG. 4B—Hydrophobicity of the predicted peptide surface (RED-hydrophobic amino-acids, BLUE-polar amino-acids). Hydrophobic core is predicted to stabilize the bent C-terminus helical structure. FIG. 4C—Identified binding site in the predicted structure of the APOL1 C-terminus domain (Zhang 2008, 2009), based on similarity to an analogous known binding site (Billas et al. 2003). S342G is involved in the predicted binding site domain, and is predicted to modify its binding ability.

[0039] FIGS. 5A-D are schematic pie charts depicting allele frequencies for the APOL1 SNP rs73885319 (S342G) in African Americans and Hispanic Americans ESKD cases versus controls. "G" refers to the risk allele and "A" refers to the "protective" allele; FIG. 5A—Controls African Americans; FIG. 5B (Cases African Americans); FIG. 5C (Controls, Hispanic Americans); FIG. 5D (Cases, Hispanic Americans). Note the high prevalence of the risk allele (G) among African American controls as compared to Hispanic American controls, and the significant increase in frequency of the risk allele among both African Americans and Hispanic American controls.

[0040] FIG. 6 is a gel image depicting results of restriction fragments length polymorphism (RFLP) analysis for APOL1 missense mutations SNP rs73885319 (S342G) and SNP rs60910145 (I384M). Shown are the restriction fragments resulted from digestion of DNA with the HindIII and NSPI restriction enzymes. SNP APOL1 rs73885319 A/G: The allele containing G eliminates a recognition site of endonuclease HindIII. APOL1 SNP rs60910145 T/G: The allele containing T generates a new recognition site of endonuclease NspI. The sign '+' indicates an additional cutting site in the DNA fragment due to the presence of the mutation. The sign '-' indicates no additional cutting site in the DNA fragment.

DESCRIPTION OF SPECIFIC EMBODIMENTS OF THE INVENTION

[0041] The present invention, in some embodiments thereof, relates to methods and kits for determining predisposition to develop a kidney disease, and, more particularly,

but not exclusively, to methods of designing a life-style change and/or a treatment regimen in predisposed subjects.

[0042] Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not necessarily limited in its application to the details set forth in the following description or exemplified by the Examples. The invention is capable of other embodiments or of being practiced or carried out in various ways.

[0043] The present inventors have uncovered that SNPs in the APOL1 gene are highly associated with renal disease and therefore identification of the APOL1 risk alleles can be used to determine predisposition to renal diseases.

[0044] Thus, as described in the Examples section which follows, the present inventors have uncovered that two Western African specific missense mutations (S342G and I384M) in the APOL1 gene/polypeptide are more strongly associated with ESKD than previously reported MYH9 variants (Example 1, Tables 1 and 3, FIGS. 1, 2 and 5A-D). In addition, the present inventors have uncovered that the distribution of these risk variants in African populations is consistent with the pattern of African ancestry ESKD risk previously attributed to MYH9 (Example 2, Table 3, FIGS. 3A-C). These results demonstrate that the functional polymorphisms in the APOL1 protein account for the increased susceptibility of the African American population to develop renal disease such as ESKD, and suggest the use of these markers as diagnostic and prognostic markers for diagnosing renal diseases and for designing life style change and treatment regimens to subjects at risk of renal disease based on the presence or absence of the predisposition to develop the renal disease.

[0045] Thus, according to an aspect of some embodiments of the invention there is provided a method of determining predisposition of a subject to develop a kidney disease. The method comprising identifying in a sample of the subject at least one APOL1 polypeptide variant which is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions, wherein presence of the APOL1 polypeptide variant indicates increased predisposition of the subject to develop the kidney disease, thereby determining the predisposition of the subject to develop the kidney disease.

[0046] As used herein the term “predisposition” refers to the tendency of a subject to develop a certain condition (e.g., a kidney disease).

[0047] According to some embodiments of the invention, the predisposition which is determined by the method of some embodiments of the invention is a genetic predisposition, i.e., predisposition to develop a condition due to a certain genetic background as compared to a subject devoid of such genetic background.

[0048] It should be noted that a subject who is predisposed to develop a disease (e.g., kidney disease) is more likely to develop the disease than a non-predisposed subject.

[0049] The predisposition to develop the kidney disease can be quantified by generating and using genotype relative risk (GRR) values. The GRR is the increased chance of an individual with a particular genotype to develop the disease. Thus, the GRR of the risk genotype G, with respect to the protective genotype G_0 , is the ratio between the risk of an individual carrying genotype G to develop the disease, and the risk of an individual carrying genotype G_0 to develop the disease. The GRR used herein is represented in terms of an

appropriate odds ratio (OR) of G versus G_0 in cases and controls. Moreover, computation of GRR of genotypes or haplotypes is based on a multiplicative model in which the GRR of an homozygote individual is the square of the GRR of an heterozygote individual. For further details see Risch and Merikangas, 1996 [The future of genetic studies of complex human diseases. Science 273: 1516-1517], which is incorporated herein by reference in its entirety.

[0050] Once calculated, the GRR can reflect the increased predisposition risk on an individual with a specific APOL1 genotype to develop the kidney disease.

[0051] As used herein the phrase “APOL1 polypeptide variant” refers to an APOL1 amino acid sequence which comprises at least one amino acid change with respect to the wild type APOL1 polypeptide set forth by SEQ ID NO:1. The amino acid change can be an amino acid(s) substitution, deletion or insertion. The amino acid substitution can be a conservative substitution, i.e., with a similar amino acid (e.g., a positively charged amino acid is replaced by another positively charged amino acid; a hydrophobic amino acid is replaced by another hydrophobic amino acid; a negatively-charged amino acid is replaced by another negatively charged amino acid), or it can be a non-conservative substitution, i.e., with an amino acid having different characteristics (e.g., a positively charged amino acid is replaced by a negatively charged amino acid; a hydrophobic amino acid is replaced by a positively-charged amino acid; a negatively-charged amino acid is replaced by a hydrophobic amino acid, etc.).

[0052] It should be noted that the amino acid change can include one or more amino acids of the APOL1 polypeptide set forth by SEQ ID NO:1.

[0053] It should be noted that the “wild type APOL1 polypeptide” comprises “S” (Serine) at amino acid position 342 in SEQ ID NO:1, “I” (Isoleucine) at amino acid position 384 in SEQ ID NO:1, “N” (Asparagine) at amino acid position 388 in SEQ ID NO:1 and “Y” (Tyrosine) at amino acid position 389 in SEQ ID NO:1.

[0054] According to some embodiments of the invention, the APOL1 variant comprises only one amino acid change with respect to the wild type APOL1 polypeptide set forth by SEQ ID NO:1.

[0055] According to some embodiments of the invention, the APOL1 variant comprises two amino acid changes with respect to the wild type APOL1 polypeptide set forth by SEQ ID NO:1.

[0056] According to some embodiments of the invention, the APOL1 variant comprises at least one amino acid change and no more than about 60 amino acid changes with respect to the wild type APOL1 polypeptide set forth by SEQ ID NO:1.

[0057] According to some embodiments of the invention, the APOL1 variant comprises about 2-20 (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20) amino acid changes with respect to the wild type APOL1 polypeptide set forth by SEQ ID NO:1.

[0058] According to some embodiments of the invention, the APOL1 polypeptide variant comprises a mutation (missense, nonsense, deletion or insertion) in the C-terminal helix of the APOL1 polypeptide. According to some embodiments of the invention, the amino acid change(s) in APOL1 variant occurs in the SRA-interacting domain in the C-terminal end of the APOL1 polypeptide, between amino acids 339-398 of SEQ ID NO:1 (FIG. 1 of Lecordier et al., 2009, PLoS Pathog. December 2009; 5(12):e1000685).

[0059] According to some embodiments of the invention, the amino acid change(s) in APOL1 variant does not occur within the pore forming domain (amino acids 60-238 of SEQ ID NO:1) and/or within the membrane-addressing domain (amino acids 238-304 of SEQ ID NO:1).

[0060] As described, the APOL1 polypeptide variant which is identified by the method of some embodiments of the invention is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions.

[0061] According to some embodiments of the invention, the *Trypanosoma brucei rhodesiense* is resistant to lysis by the apolipoprotein present in normal human serum.

[0062] According to some embodiments of the invention, the *Trypanosoma brucei rhodesiense* expresses serum resistance-associated protein (SRA), also referred to as SRA-positive.

[0063] It should be noted that wild type APOL1 fails to lyse the SRA+ (positive) *Trypanosoma brucei rhodesiense* (See for example, FIG. 3C in Genovese G., et al. "Association of Trypanolytic APOL1 variants with kidney disease in African-Americans", Science 2010, 329: 841-845, which is fully incorporated herein by reference).

[0064] As used herein the phrase "higher trypanolytic activity" refers to at least about 10%, at least about 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% greater trypanolytic activity of the APOL1 polypeptide variant as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1.

[0065] According to some embodiments of the invention, the APOL1 polypeptide variant which exhibits a higher trypanolytic activity as compared to the APOL1 wild type polypeptide set forth by SEQ ID NO:1 is selected from the group G342, M384, N388-del and Y389-del (positions refer to the APOL1 polypeptide sequence set forth in SEQ ID NO:1).

[0066] According to some embodiments of the invention, the APOL1 polypeptide variant is APOL1 N388-del/Y389-del (numbers relate to the APOL1 polypeptide sequence set forth in SEQ ID NO:1).

[0067] The sample used according to the method of this aspect of the invention can be any biological sample (obtained from an individual, e.g., human) which comprises the APOL1 polypeptide, or a sample of a recombinant or synthetic APOL1 polypeptide. Examples of biological samples include, but are not limited to whole blood or fractions thereof, such as serum, lipoproteins [High-density lipoprotein (HDL)], body fluids and excretions, tissue samples, and the like. The sample can be at least partially isolated from the subject, and further purified, or can be a crude sample.

[0068] According to some embodiments of the invention the sample is a serum sample.

[0069] According to some embodiments of the invention the sample is plasma HDL.

[0070] The trypanolytic activity can be measured as the percentage of survival of the trypanosome in the presence of a sample containing the APOL1 polypeptide. Thus, for example, while in the presence of the wild type APOL1 polypeptide set forth in SEQ ID NO:1 the percentage of survival can be about 100%, in the presence of an APOL1 variant the percentage of survival of the trypanosome can be

reduced to 80%, 60%, 40%, 20% and 0% survival, which is referred to as 20%, 40%, 60%, 80% and 100% trypanolytic activity of the APOL1 polypeptide, respectively.

[0071] Assays which can be used to measure the trypanolytic activity of a protein on *Trypanosoma brucei rhodesiense* are known in the art and described, for example in Lecordier L, et al. "C-terminal mutants of apolipoprotein L-I efficiently kill both *Trypanosoma brucei brucei* and *Trypanosoma brucei rhodesiense*." PLoS Pathog. December 2009; 5(12):e1000685. Epub Dec. 4, 2009; U.S. Patent Application No. 20110030078; Genovese G., et al. "Association of Trypanolytic APOL1 variants with kidney disease in African-Americans", Science 2010, 329: 841-845; and in Tomlinson, et al., "High-density-lipoprotein-independent Killing of *Trypanosoma brucei brucei* by Human Serum," Mol. Biochem. Parasitol. 70:131-138 (1995), each of which is hereby incorporated by reference in its entirety.

[0072] Following is a non-limiting description of an in vitro assays which can be used to determine the trypanolytic activity of the APOL1 variant. Samples containing the APOL1 polypeptide can be for example, serum samples, plasma HDL or recombinant APOL1 polypeptides. Serum samples can be obtained by simple blood drawing and separation of serum from the blood. Plasma HDLs can be prepared by density gradient centrifugation and gel filtration. Recombinant APOL1 polypeptides [wild type APOL1 (SEQ ID NO:1) or APOL1 polypeptide variants (e.g., the G342, M384, N388-del and/or Y389-del APOL1 polypeptide variant)] can be expressed from a nucleic acid construct comprising the coding sequence of APOL1 which is expressed in prokaryotic (e.g., *Escherichia coli*) or eukaryotic (e.g., 293T) cells essentially as described in Lecordier L, et al. 2009 (Supra).

[0073] The trypanolytic activity is determined in *Trypanosoma brucei rhodesiense*. There are two major types of *Trypanosoma brucei rhodesiense* clones: the normal human serum-resistant (SRA+) and the normal human serum-sensitive (SRA-) *T. b. rhodesiense* ETat 1.2 clones can be used. ETat 1.2R is resistant to normal human serum, and ETat 1.2S is sensitive to normal human serum. It should be noted that while the SRA- *T. b. rhodesiense* (sensitive) clone is used as a positive control for the trypanolytic activity by all forms of APOL1 polypeptides (wild type and variant APOL1 polypeptides), the SRA+ *T. b. rhodesiense* (resistance) is used to differentiate between the APOL1 variants which are capable of lysing the *Trypanosoma brucei rhodesiense* and the wild type APOL1 polypeptide set forth in SEQ ID NO:1 which is incapable of lysing the *Trypanosoma brucei rhodesiense* as can be determined by a survival assay as described below. The trypanosomes can be diluted into pre-warmed DMEM containing 10% fetal bovine serum (FBS). In 96-well plates, serum samples are diluted into DMEM+10% fetal calf serum (FCS) and trypanosomes are added to a final volume of 0.2 ml, containing about 2.5×10^5 /ml parasites. Killing is allowed to proceed for about 17 hours at 37° C. in a CO₂ equilibrated incubator and living trypanosomes are counted using a hemacytometer. HDL killing assays are performed for 150 minutes at 37° C. in DMEM, 0.2% BSA using aliquots of sized fractionated lipoproteins incubated with parasites (trypanosome) diluted at a final concentration of about 5×10^5 cells/ml. Parasite lysis can be determined using a calcein-AM fluorescence-based assay (Tomlinson et al., "High-density-lipoprotein-independent Killing of *Trypanosoma brucei* by Human Serum," Mol Biochem Parasitol 70: 131-8 (1995), which is hereby incorporated by reference in its entirety).

Recombinant APOL1 polypeptides (e.g., about 0.1-20 µg/ml, e.g., about 0.6-10 µg/ml, e.g., about 0.6-5 µg/ml) can be incubated with the *trypanosoma* overnight and the survival of *trypanosoma* can be determined and expressed as percentage (%) survival as compared to control (e.g., FCS)

[0074] According to some embodiments of the invention, the APOL1 polypeptide variant has a reduced binding ability to serum resistance-associated protein (SRA) expressed by *Trypanosoma brucei rhodesiense* as compared to the binding ability of the APOL1 wild type polypeptide under identical assay conditions.

[0075] Assays of measuring the binding ability of the APOL1 polypeptide to *Trypanosoma brucei rhodesiense* SRA are known in the art and are described for example, in Genovese G., et al. "Association of Trypanolytic APOL1 variants with kidney disease in African-Americans", Science 2010, 329: 841-845, which is hereby incorporated herein by its entirety.

[0076] Determination of predisposition of a subject to develop a kidney disease can be also performed by single nucleotide polymorphism detection methods.

[0077] According to an aspect of some embodiments of the invention there is provided a method of determining predisposition of a subject to develop a kidney disease, comprising: identifying in a sample of the subject at least one APOL1 nucleotide mutation in the APOL1 genomic sequence set forth in SEQ ID NO:3 (chr22:36,622,000-36,677,000 in NCBI37.1/hg19 assembly) or at least one APOL1 polypeptide variant, wherein the at least one nucleotide mutation or polypeptide variant being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, wherein presence of the APOL1 nucleotide mutation or the APOL1 polypeptide variant indicates increased predisposition of the subject to develop the kidney disease, thereby determining the predisposition of the subject to develop the kidney disease.

[0078] The sample according to some embodiments of the invention is a DNA or protein sample.

[0079] The DNA sample can be obtained from any source of cells of the individual, including, but not limited to, peripheral blood cells (obtained using a syringe), skin cells (obtained from a skin biopsy), saliva or mouth epithelial cells (obtained from a mouth wash), and body secretions such as urine and tears, and from biopsies, etc. Alternatively, nucleic acid tests can be performed on dry samples (e.g. hair or skin). The sample may contain genomic DNA, cDNA or RNA. Methods of preparing genomic DNA or cDNA and RNA are well known in the art.

[0080] According to some embodiments of the invention, the DNA sample is obtained from a peripheral blood sample. Methods of extracting DNA from blood samples are well known in the art.

[0081] Once obtained, the DNA sample is preferably characterized for the presence or absence of at least one or more of the APOL1 nucleotide mutations in a homozygous or a heterozygous form in the sample of the subject.

[0082] The term "absence" as used herein in regard to the nucleotide or protein mutation describes the negative result of a specific genotype determination test. For example, if the genotype determination test is suitable for the identification of guanine nucleotide—containing allele of SNP rs73885319 (encoding APOL1 G342), and the individual on which the test is performed is homozygote for the adenosine nucleotide—

containing allele of SNP rs73885319 (encoding APOL1 S342), then the result of the test will be "absence of nucleotide or protein mutation".

[0083] The terms "homozygous" or "heterozygous" refer to two identical or two different alleles, respectively, of a certain mutation or polymorphism.

[0084] The phrases "APOL1 nucleotide mutation" or "APOL1 single nucleotide polymorphism (SNP)", which are interchangeably used herein, refer to a substitution, deletion, insertion or inversion of a nucleotide sequence in the APOL1 genomic sequence with respect to the wild type genomic sequence set forth in SEQ ID NO:3.

[0085] It should be noted that the APOL1 mutation can be a substitution, deletion, insertion or inversion of a nucleotide sequence in the APOL1 coding sequence with respect to the wild type APOL1 mRNA sequence set forth in SEQ ID NO:2 (GenBank Accession No. NM_001136540.1); or it can be a missense, nonsense, deletion or insertion resulting in a mutant APOL1 polypeptide with respect to the wild type polypeptide sequence set forth in SEQ ID NO:1 (GenBank Accession No. NP_003652.2).

[0086] It should be noted that "wild type APOL1 mRNA sequence" comprises "A" (adenosine nucleotide) at nucleotide position 1231 in SEQ ID NO:2, "T" (thymidine nucleotide) at position 1359 in SEQ ID NO:2, "ATT" at nucleotide positions 1369-1371 in SEQ ID NO:2, "ATA" at nucleotide positions 1372-1374 in SEQ ID NO:2, and "T" (thymidine nucleotide) at nucleotide position 2538 in SEQ ID NO:2.

[0087] It should be noted that the "wild type APOL1 genomic sequence" comprises "A" (adenosine nucleotide) at nucleotide position 34556 in SEQ ID NO:3, "A" (adenosine nucleotide) at nucleotide position 39907 in SEQ ID NO:3, "T" (thymidine nucleotide) at nucleotide position 40035 in SEQ ID NO:3, "AAT" at nucleotide positions 40045-40047 in SEQ ID NO:3, "TAT" at nucleotide positions 40048-40050 in SEQ ID NO:3, "T" (thymidine nucleotide) at nucleotide position 41214 in SEQ ID NO:3, and "G" (guanine nucleotide) at nucleotide position 45155 in SEQ ID NO:3.

[0088] The APOL1 mutation can be a missense mutation (i.e., a mutation which changes an amino acid residue in the protein with another amino acid residue), a nonsense mutation (i.e., a mutation which introduces a stop codon in a protein), a frameshift mutation (i.e., a mutation, usually, deletion or insertion of nucleic acids which changes the reading frame of the protein, and may result in an early termination or in a longer amino acid sequence), a readthrough mutation (i.e., a mutation which results in an elongated protein due to a change in a coding frame or a modified stop codon), a promoter mutation (i.e., a mutation in a promoter sequence, usually 5' to the transcription start site of a gene, which result in up-regulation or down-regulation of a specific gene product), a regulatory mutation (i.e., a mutation in a region upstream or downstream, or within a gene, which affects the expression of the gene product), a deletion (i.e., a mutation which deletes coding or non-coding nucleic acids in a gene sequence), an insertion (i.e., a mutation which inserts coding or non-coding nucleic acids into a gene sequence), an inversion (i.e., a mutation which results in an inverted coding or non-coding sequence), and a duplication (i.e., a mutation which results in a duplicated coding or non-coding sequence).

[0089] According to some embodiments of the invention, the at least one mutation comprises no more than about 100, no more than about 90, no more than about 80, no more than about 70, no more than about 60, no more than about 50, no

more than about 40, no more than about 30, no more than about 20, no more than about 15, no more than about 10, e.g., no more than 9, 8, 7, 6, 5, 4, 3, 2 nucleotide mutations in the APOL1 genomic sequence with respect to the wild type genomic sequence set forth in SEQ ID NO:3.

[0090] According to some embodiments of the invention, the at least one mutation comprises no more than about 100, no more than about 90, no more than about 80, no more than about 70, no more than about 60, no more than about 50, no more than about 40, no more than about 30, no more than about 20, no more than about 15, no more than about 10, e.g., no more than 9, 8, 7, 6, 5, 4, 3, 2 nucleotide mutations in the APOL1 coding sequence with respect to the wild type coding sequence set forth in SEQ ID NO:2.

[0091] According to some embodiments of the invention, the at least one mutation comprises no more than about 100, no more than about 90, no more than about 80, no more than about 70, no more than about 60, no more than about 50, no more than about 40, no more than about 30, no more than about 20, no more than about 15, no more than about 10, e.g., no more than 9, 8, 7, 6, 5, 4, 3, 2 amino acid mutations in the APOL1 genomic sequence with respect to the wild type polypeptide sequence set forth in SEQ ID NO:1.

[0092] As described, the nucleotide mutation is in linkage disequilibrium with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, which is encoded by the A→G (adenosine to guanine) substitution at position 1231 of the APOL1 mRNA (coding sequence) set forth by SEQ ID NO:2.

[0093] The phrase “linkage disequilibrium” (LD) is used to describe the statistical correlation between two neighboring polymorphic genotypes. Typically, LD refers to the correlation between the alleles of a random gamete at the two loci, assuming Hardy-Weinberg equilibrium (statistical independence) between gametes. LD is quantified with either Lewontin’s parameter of association (D') or with Pearson correlation coefficient (r) [Devlin B, Risch N. (1995). A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*. 29: 311-322.]. Two loci with a LD value of 1 are the to be in complete LD. At the other extreme, two loci with a LD value of 0 are termed to be in linkage equilibrium. Linkage disequilibrium is calculated following the application of the expectation maximization algorithm (EM) for the estimation of haplotype frequencies [Slatkin M, Excoffier L. (1996). Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity*. 76: 377-83.]. Preferably, LD values according to the present invention for neighboring genotypes/loci are selected above 0.1, preferably, above 0.2, more preferably above 0.5, more preferably, above 0.6, still more preferably, above 0.7, preferably, above 0.8, more preferably above 0.9, ideally about 1.0.

[0094] In the D-Prime Plot, each diagonal represents a different SNP, with each square representing a pairwise comparison between two SNPs. SNPs are numbered sequentially, 5' to 3', and their relative location is indicated along the top. Red squares indicate statistically significant ($LOD > 2$) allelic association (linkage disequilibrium, LD) between the pair of SNPs, as measured by the D' statistic; darker colors of red indicate higher values of D' , up to a maximum of 1. White squares indicate pairwise D' values of < 1 with no statistically significant evidence of LD. Blue squares indicate pairwise D' values of 1 but without statistical significance.

[0095] According to some embodiments of the invention, the LD between the APOL1 nucleotide mutation and the S342G mutation is characterized by a Lewontin correlation coefficient (D') higher than about 0.5, e.g., higher than about 0.6, higher than about 0.7, higher than about 0.8, higher than about 0.9, higher than about 0.91, higher than about 0.92, higher than about 0.93, higher than about 0.94, higher than about 0.95, higher than about 0.96, higher than about 0.97, higher than about 0.98, higher than about 0.99, e.g., about 1.0.

[0096] According to some embodiments of the invention, the significance of the LD is characterized by $LOD \geq 2$, e.g., $LOD \geq 3$, $LOD \geq 4$, $LOD \geq 5$, $LOD \geq 6$, $LOD \geq 7$.

[0097] As shown in FIG. 2, and described in Example 1 of the Examples section which follows, and based on the 1000-genomes-project (Hypertext Transfer Protocol://World Wide Web (dot)1000genomes(dot)org/) the present inventors have uncovered that the rs9622363, rs58384577, rs73885319 and rs60910145 SNPs are in complete LD ($D'/LOD=100$), and that the rs60295735 SNP is in tight linkage disequilibrium with each of them [i.e., D'/LOD (in percentages)=91 (with rs9622363), 95 (with rs73885319) and 97 (with rs60910145)].

[0098] According to some embodiments of the invention, the APOL1 nucleotide mutation is selected from the group consisting of the guanine-containing allele of single nucleotide polymorphism (SNP) rs73885319 (SEQ ID NO:6), the guanine-containing allele of SNP rs60910145 (SEQ ID NO:7), the guanine-containing allele of SNP rs9622363 (SEQ ID NO:8), the adenine-containing allele of SNP rs60295735 (SEQ ID NO:9) and the cytosine-containing allele of SNP rs58384577 (SEQ ID NO:10).

[0099] According to some embodiments of the invention, the nucleotide mutation creates an APOL1 protein variant.

[0100] The SNPs of some embodiments of the invention can be identified using a variety of approaches suitable for identifying sequence alterations. One option is to determine the entire gene sequence of a PCR reaction product. Alternatively, a given segment of nucleic acid may be characterized on several other levels. At the lowest resolution, the size of the molecule can be determined by electrophoresis by comparison to a known standard run on the same gel. A more detailed picture of the molecule may be achieved by cleavage with combinations of restriction enzymes prior to electrophoresis, to allow construction of an ordered map. The presence of specific sequences within the fragment can be detected by hybridization of a labeled probe, or the precise nucleotide sequence can be determined by partial chemical degradation or by primer extension in the presence of chain-terminating nucleotide analogs.

[0101] Following is a non-limiting list of SNPs detection methods which can be used to identify one or more of the SNPs of some embodiments of the invention.

[0102] Restriction fragment length polymorphism (RFLP): This method uses a change in a single nucleotide (the SNP nucleotide) which modifies a recognition site for a restriction enzyme resulting in the creation or destruction of an RFLP. Single nucleotide mismatches in DNA heteroduplexes are also recognized and cleaved by some chemicals, providing an alternative strategy to detect single base substitutions, generically named the “Mismatch Chemical Cleavage” (MCC) (Gogos et al., *Nucl. Acids Res.*, 18:6807-6817, 1990). However, this method requires the use of osmium tetroxide and piperidine, two highly noxious chemicals which are not suited for use in a clinical laboratory.

[0103] Allele specific oligonucleotide (ASO): In this method, an allele-specific oligonucleotide (ASO) is designed to hybridize in proximity to the polymorphic nucleotide, such that a primer extension or ligation event can be used as the indicator of a match or a mis-match. Hybridization with radioactively labeled allelic specific oligonucleotides (ASO) also has been applied to the detection of specific SNPs (Conner et al., *Proc. Natl. Acad. Sci.*, 80:278-282, 1983). The method is based on the differences in the melting temperature of short DNA fragments differing by a single nucleotide. Stringent hybridization and washing conditions can differentiate between mutant and wild-type alleles.

[0104] Denaturing/Temperature Gradient Gel Electrophoresis (DGGE/TGGE): Two other methods rely on detecting changes in electrophoretic mobility in response to minor sequence changes. One of these methods, termed "Denaturing Gradient Gel Electrophoresis" (DGGE) is based on the observation that slightly different sequences will display different patterns of local melting when electrophoretically resolved on a gradient gel. In this manner, variants can be distinguished, as differences in melting properties of homoduplexes versus heteroduplexes differing in a single nucleotide can detect the presence of SNPs in the target sequences because of the corresponding changes in their electrophoretic mobilities. The fragments to be analyzed, usually PCR products, are "clamped" at one end by a long stretch of G-C base pairs (30-80) to allow complete denaturation of the sequence of interest without complete dissociation of the strands. The attachment of a GC "clamp" to the DNA fragments increases the fraction of mutations that can be recognized by DGGE (Abrams et al., *Genomics* 7:463-475, 1990). Attaching a GC clamp to one primer is critical to ensure that the amplified sequence has a low dissociation temperature (Sheffield et al., *Proc. Natl. Acad. Sci.*, 86:232-236, 1989; and Lerman and Silverstein, *Meth. Enzymol.*, 155:482-501, 1987). Modifications of the technique have been developed, using temperature gradients (Wartell et al., *Nucl. Acids Res.*, 18:2699-2701, 1990), and the method can be also applied to RNA:RNA duplexes (Smith et al., *Genomics* 3:217-223, 1988).

[0105] Limitations on the utility of DGGE include the requirement that the denaturing conditions must be optimized for each type of DNA to be tested. Furthermore, the method requires specialized equipment to prepare the gels and maintain the needed high temperatures during electrophoresis. The expense associated with the synthesis of the clamping tail on one oligonucleotide for each sequence to be tested is also a major consideration. In addition, long running times are required for DGGE. The long running time of DGGE was shortened in a modification of DGGE called constant denaturant gel electrophoresis (CDGE) (Borresen et al., *Proc. Natl. Acad. Sci. USA* 88:8405, 1991). CDGE requires that gels be performed under different denaturant conditions in order to reach high efficiency for the detection of SNPs.

[0106] A technique analogous to DGGE, termed temperature gradient gel electrophoresis (TGGE), uses a thermal gradient rather than a chemical denaturant gradient (Scholz, et al., *Hum. Mol. Genet.* 2:2155, 1993). TGGE requires the use of specialized equipment which can generate a temperature gradient perpendicularly oriented relative to the electrical field. TGGE can detect mutations in relatively small fragments of DNA therefore scanning of large gene segments requires the use of multiple PCR products prior to running the gel.

[0107] Single-Strand Conformation Polymorphism (SSCP): Another common method, called "Single-Strand Conformation Polymorphism" (SSCP) was developed by Hayashi, Sekya and colleagues (reviewed by Hayashi, *PCR Meth. Appl.*, 1:34-38, 1991) and is based on the observation that single strands of nucleic acid can take on characteristic conformations in non-denaturing conditions, and these conformations influence electrophoretic mobility. The complementary strands assume sufficiently different structures that one strand may be resolved from the other. Changes in sequences within the fragment will also change the conformation, consequently altering the mobility and allowing this to be used as an assay for sequence variations (Orita, et al., *Genomics* 5:874-879, 1989).

[0108] The SSCP process involves denaturing a DNA segment (e.g., a PCR product) that is labeled on both strands, followed by slow electrophoretic separation on a non-denaturing polyacrylamide gel, so that intra-molecular interactions can form and not be disturbed during the run. This technique is extremely sensitive to variations in gel composition and temperature. A serious limitation of this method is the relative difficulty encountered in comparing data generated in different laboratories, under apparently similar conditions.

[0109] Dideoxy fingerprinting (ddF): The dideoxy fingerprinting (ddF) is another technique developed to scan genes for the presence of mutations (Liu and Sommer, *PCR Methods Appl.*, 4:97, 1994). The ddF technique combines components of Sanger dideoxy sequencing with SSCP. A dideoxy sequencing reaction is performed using one dideoxy terminator and then the reaction products are electrophoresed on nondenaturing polyacrylamide gels to detect alterations in mobility of the termination segments as in SSCP analysis. While ddF is an improvement over SSCP in terms of increased sensitivity, ddF requires the use of expensive dideoxynucleotides and this technique is still limited to the analysis of fragments of the size suitable for SSCP (i.e., fragments of 200-300 bases for optimal detection of mutations).

[0110] In addition to the above limitations, all of these methods are limited as to the size of the nucleic acid fragment that can be analyzed. For the direct sequencing approach, sequences of greater than 600 base pairs require cloning, with the consequent delays and expense of either deletion subcloning or primer walking, in order to cover the entire fragment. SSCP and DGGE have even more severe size limitations. Because of reduced sensitivity to sequence changes, these methods are not considered suitable for larger fragments. Although SSCP is reportedly able to detect 90% of single-base substitutions within a 200 base-pair fragment, the detection drops to less than 50% for 400 base pair fragments. Similarly, the sensitivity of DGGE decreases as the length of the fragment reaches 500 base-pairs. The ddF technique, as a combination of direct sequencing and SSCP, is also limited by the relatively small size of the DNA that can be screened.

[0111] Pyrosequencing™ analysis (Pyrosequencing, Inc. Westborough, Mass., USA): This technique is based on the hybridization of a sequencing primer to a single stranded, PCR-amplified, DNA template in the presence of DNA polymerase, ATP sulfurylase, luciferase and apyrase enzymes and the adenosine 5' phosphosulfate (APS) and luciferin substrates. In the second step the first of four deoxynucleotide triphosphates (dNTP) is added to the reaction and the DNA polymerase catalyzes the incorporation of the deoxynucle-

otide triphosphate into the DNA strand, if it is complementary to the base in the template strand. Each incorporation event is accompanied by release of pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide. In the last step the ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5' phosphosulfate. This ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) camera and seen as a peak in a pyrogramTM. Each light signal is proportional to the number of nucleotides incorporated.

[0112] AcycloprimeTM analysis (Perkin Elmer, Boston, Mass., USA): This technique is based on fluorescent polarization (FP) detection. Following PCR amplification of the sequence containing the SNP of interest, excess primer and dNTPs are removed through incubation with shrimp alkaline phosphatase (SAP) and exonuclease I. Once the enzymes are heat inactivated, the Acycloprime-FP process uses a thermostable polymerase to add one of two fluorescent terminators to a primer that ends immediately upstream of the SNP site. The terminator(s) added are identified by their increased FP and represent the allele(s) present in the original DNA sample. The Acycloprime process uses AcycloPolTM, a novel mutant thermostable polymerase from the Archeon family, and a pair of AcycloTerminatorsTM labeled with R110 and TAMRA, representing the possible alleles for the SNP of interest. AcycloTerminatorTM non-nucleotide analogs are biologically active with a variety of DNA polymerases. Similarly to 2',3'-dideoxynucleotide-5'-triphosphates, the acyclic analogs function as chain terminators. The analog is incorporated by the DNA polymerase in a base-specific manner onto the 3'-end of the DNA chain, and since there is no 3'-hydroxyl, is unable to function in further chain elongation. It has been found that AcycloPol has a higher affinity and specificity for derivatized AcycloTerminators than various Taq mutant have for derivatized 2',3'-dideoxynucleotide terminators.

[0113] Reverse dot blot: This technique uses labeled sequence specific oligonucleotide probes and unlabeled nucleic acid samples. Activated primary amine-conjugated oligonucleotides are covalently attached to carboxylated nylon membranes. After hybridization and washing, the labeled probe, or a labeled fragment of the probe, can be released using oligomer restriction, i.e., the digestion of the duplex hybrid with a restriction enzyme. Circular spots or lines are visualized colorimetrically after hybridization through the use of streptavidin horseradish peroxidase incubation followed by development using tetramethylbenzidine and hydrogen peroxide, or via chemiluminescence after incubation with avidin alkaline phosphatase conjugate and a luminous substrate susceptible to enzyme activation, such as CSPD, followed by exposure to x-ray film.

[0114] It will be appreciated that advances in the field of SNP detection have provided additional accurate, easy, and inexpensive large-scale SNP genotyping techniques, such as dynamic allele-specific hybridization (DASH, Howell, W. M. et al., 1999. Dynamic allele-specific hybridization (DASH). Nat. Biotechnol. 17: 87-8), microplate array diagonal gel electrophoresis [MADGE, Day, I. N. et al., 1995. High-throughput genotyping using horizontal polyacrylamide gels with wells arranged for microplate array diagonal gel electrophoresis (MADGE). Biotechniques. 19: 830-5], the Taq-Man system (Holland, P. M. et al., 1991. Detection of specific

polymerase chain reaction product by utilizing the 5'→3' exonuclease activity of *Thermus aquaticus* DNA polymerase. Proc Natl Acad Sci USA. 88: 7276-80), as well as various DNA "chip" technologies such as the GeneChip microarrays (e.g., Affymetrix SNP chips) which are disclosed in U.S. Pat. No. 6,300,063 to Lipshutz, et al. 2001, which is fully incorporated herein by reference, Genetic Bit Analysis (GBATM) which is described by Goelet, P. et al. (PCT Appl. No. 92/15712), peptide nucleic acid (PNA, Ren B, et al., 2004. Nucleic Acids Res. 32: e42) and locked nucleic acids (LNA, Latorra D, et al., 2003. Hum. Mutat. 22: 79-85) probes, Molecular Beacons (Abravaya K, et al., 2003. Clin Chem Lab Med. 41: 468-74), intercalating dye [Germer, S. and Higuchi, R. Single-tube genotyping without oligonucleotide probes. Genome Res. 9:72-78 (1999)], FRET primers (Solinas A et al., 2001. Nucleic Acids Res. 29: E96), AlphaScreen (Beaudet L, et al., Genome Res. 2001, 11(4): 600-8), SNPstream (Bell P A, et al., 2002. Biotechniques. Suppl.: 70-2, 74, 76-7), Multiplex minisequencing (Curcio M, et al., 2002. Electrophoresis. 23: 1467-72), SnaPshot (Turner D, et al., 2002. Hum Immunol. 63: 508-13), MassEXTEND (Cashman JR, et al., 2001. Drug Metab Dispos. 29: 1629-37), GOOD assay (Sauer S, and Gut I G. 2003. Rapid Commun. Mass. Spectrom. 17: 1265-72), Microarray minisequencing (Liljedahl U, et al., 2003. Pharmacogenetics. 13: 7-17), arrayed primer extension (APEX) (Tonisson N, et al., 2000. Clin. Chem. Lab. Med. 38: 165-70), Microarray primer extension (O'Meara D, et al., 2002. Nucleic Acids Res. 30: e75), Tag arrays (Fan J B, et al., 2000. Genome Res. 10: 853-60), Template-directed incorporation (TDI) (Akula N, et al., 2002. Biotechniques. 32: 1072-8), fluorescence polarization (Hsu T M, et al., 2001. Biotechniques. 31: 560, 562, 564-8), Colorimetric oligonucleotide ligation assay (OLA, Nickerson D A, et al., 1990. Proc. Natl. Acad. Sci. USA. 87: 8923-7), Sequence-coded OLA (Gasparini P, et al., 1999. J. Med. Screen. 6: 67-9), Microarray ligation, Ligase chain reaction, Padlock probes, Rolling circle amplification, Invader assay (reviewed in Shi M M. 2001. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. Clin Chem. 47: 164-72), coded microspheres (Rao K V et al., 2003. Nucleic Acids Res. 31: e66) and MassArray (Leushner J, Chiu N H, 2000. Mol Diagn. 5: 341-80).

[0115] For example, as described under "GENERAL MATERIALS AND EXPERIMENTAL METHODS" in the Examples section which follows, determination of the SNPs was performed using the KasPar methodology (Petkov et al. 2004), as well as by PCR-RFLP [for SNP RS73885319 (S342G) and SNP RS60910145 (I384M)].

[0116] The protein sample can be obtained from any source of cells, tissues or body fluids of the individual. According to some embodiments of the invention, the protein sample is obtained from a blood sample (serum or plasma) of the individual. Methods of extracting proteins from blood samples are well known in the art. Once extracted, determination of the APOL1 polypeptide variants can be accomplished directly, by analyzing the protein gene products of the APOL1 gene, or portions thereof. Such a direct analysis is often accomplished using an immunological detection method.

[0117] Immunological detection methods: The immunological detection methods used in context of the present invention are fully explained in, for example, "Using Antibodies: A Laboratory Manual" [Ed Harlow, David Lane eds., Cold Spring Harbor Laboratory Press (1999)] and those familiar with the art will be capable of implementing the

various techniques summarized hereinbelow as part of the present invention. All of the immunological techniques require antibodies specific to at least one of the two APOL1 alleles (the wild type allele and the APOL1 variant allele). Immunological detection methods suited for use as part of the present invention include, but are not limited to, radio-immunoassay (RIA), enzyme linked immunosorbent assay (ELISA), western blot, immunohistochemical analysis, and fluorescence activated cell sorting (FACS).

[0118] Radio-immunoassay (RIA): In one version, this method involves precipitation of the desired substrate, APOL1 in this case and in the methods detailed hereinbelow, with a specific antibody and radiolabelled antibody binding protein (e.g., protein A labeled with I^{125}) immobilized on a precipitable carrier such as agarose beads. The number of counts in the precipitated pellet is proportional to the amount of substrate.

[0119] In an alternate version of the RIA, a labeled substrate and an unlabelled antibody binding protein are employed. A sample containing an unknown amount of substrate is added in varying amounts. The decrease in precipitated counts from the labeled substrate is proportional to the amount of substrate in the added sample.

[0120] Enzyme linked immunosorbent assay (ELISA): This method involves fixation of a sample (e.g., fixed cells or a proteinaceous solution) containing a protein substrate to a surface such as a well of a microtiter plate. A substrate specific antibody coupled to an enzyme is applied and allowed to bind to the substrate. Presence of the antibody is then detected and quantitated by a colorimetric reaction employing the enzyme coupled to the antibody. Enzymes commonly employed in this method include horseradish peroxidase and alkaline phosphatase. If well calibrated and within the linear range of response, the amount of substrate present in the sample is proportional to the amount of color produced. A substrate standard is generally employed to improve quantitative accuracy.

[0121] Western blot: This method involves separation of a substrate from other protein by means of an acrylamide gel followed by transfer of the substrate to a membrane (e.g., nylon or PVDF). Presence of the substrate is then detected by antibodies specific to the substrate, which are in turn detected by antibody binding reagents. Antibody binding reagents may be, for example, protein A, or other antibodies. Antibody binding reagents may be radiolabelled or enzyme linked as described hereinabove. Detection may be by autoradiography, colorimetric reaction or chemiluminescence. This method allows both quantitation of an amount of substrate and determination of its identity by a relative position on the membrane which is indicative of a migration distance in the acrylamide gel during electrophoresis.

[0122] Immunohistochemical analysis: This method involves detection of a substrate in situ in fixed cells by substrate specific antibodies. The substrate specific antibodies may be enzyme linked or linked to fluorophores. Detection is by microscopy and subjective evaluation. If enzyme linked antibodies are employed, a colorimetric reaction may be required.

[0123] Fluorescence activated cell sorting (FACS): This method involves detection of a substrate in situ in cells by substrate specific antibodies. The substrate specific antibodies are linked to fluorophores. Detection is by means of a cell sorting machine which reads the wavelength of light emitted

from each cell as it passes through a light beam. This method may employ two or more antibodies simultaneously.

[0124] The antibody used in the method of the present invention is selected differentially interactable with at least one form of a APOL1 protein encoded by an APOL1 allele having an SNP rs73885319 (S342G, i.e., the 5342 polymorph or the G342 polymorph) and SNP rs60910145 (I384M, i.e., the I384 polymorph or the M384 polymorph) and can differentiate between polymorphs of the APOL1 protein via differential antibody interaction. Antibodies useful in context of this embodiment of the invention can be prepared using methods of antibody preparation well known to one of ordinary skills in the art, using, for example, synthetic peptides derived from the two different forms of the APOL1 protein for vaccination of antibody producing animals and subsequent isolation of antibodies therefrom. Monoclonal antibodies specific to each of the APOL1 variants can also be prepared as is described, for example, in "Current Protocols in Immunology" Volumes I-III Coligan J. E., Ed. (1994); Stites et al. (Eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, Conn. (1994); Mishell and Shiigi (Eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980).

[0125] The term "antibody" as used in the present invention includes intact molecules as well as functional fragments thereof, such as Fab, $F(ab')_2$, and Fv that are capable of binding to macrophages. These functional antibody fragments are defined as follows: Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule, can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; Fab', the fragment of an antibody molecule that can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule; $(Fab')_2$, the fragment of the antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; $F(ab')_2$ is a dimer of two Fab' fragments held together by two disulfide bonds; Fv, defined as a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and single chain antibody ("SCA"), a genetically engineered molecule containing the variable region of the light chain and the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule.

[0126] Methods of making these fragments are known in the art. See for example, Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York, 1988, incorporated herein by reference.

[0127] Antibody fragments according to the present invention can be prepared by proteolytic hydrolysis of the antibody or by expression in *E. coli* or mammalian cells (e.g. Chinese hamster ovary cell culture or other protein expression systems) of DNA encoding the fragment.

[0128] Antibody fragments can be obtained by pepsin or papain digestion of whole antibodies by conventional methods. For example, antibody fragments can be produced by enzymatic cleavage of antibodies with pepsin to provide a 5S fragment denoted $F(ab')_2$. This fragment can be further cleaved using a thiol reducing agent, and optionally a blocking group for the sulfhydryl groups resulting from cleavage of disulfide linkages, to produce 3.5S Fab' monovalent fragments. Alternatively, an enzymatic cleavage using pepsin

produces two monovalent Fab' fragments and an Fc fragment directly. These methods are described, for example, by Goldenberg, U.S. Pat. Nos. 4,036,945 and 4,331,647, and references contained therein, which patents are hereby incorporated by reference in their entirety. See also Porter, R. R., *Biochem. J.*, 73: 119-126, 1959. Other methods of cleaving antibodies, such as separation of heavy chains to form monovalent light-heavy chain fragments, further cleavage of fragments, or other enzymatic, chemical, or genetic techniques may also be used, so long as the fragments bind to the antigen that is recognized by the intact antibody.

[0129] Fv fragments comprise an association of V_H and V_L chains. This association may be noncovalent, as described in Inbar et al., *Proc. Nat'l Acad. Sci. USA* 69:2659-62, 1972. Alternatively, the variable chains can be linked by an intermolecular disulfide bond or cross-linked by chemicals such as glutaraldehyde. Preferably, the Fv fragments comprise V_H and V_L chains connected by a peptide linker. These single-chain antigen binding proteins (sFv) are prepared by constructing a structural gene comprising DNA sequences encoding the V_H and V_L domains connected by an oligonucleotide. The structural gene is inserted into an expression vector, which is subsequently introduced into a host cell such as *E. coli*. The recombinant host cells synthesize a single polypeptide chain with a linker peptide bridging the two V domains. Methods for producing sFvs are described, for example, by Whitlow and Filpula, *Methods*, 2: 97-105, 1991; Bird et al., *Science* 242:423-426, 1988; Pack et al., *Bio/Technology* 11:1271-77, 1993; and Ladner et al., U.S. Pat. No. 4,946,778, which is hereby incorporated by reference in its entirety.

[0130] Another form of an antibody fragment is a peptide coding for a single complementarity-determining region (CDR). CDR peptides ("minimal recognition units") can be obtained by constructing genes encoding the CDR of an antibody of interest. Such genes are prepared, for example, by using the polymerase chain reaction to synthesize the variable region from RNA of antibody-producing cells. See, for example, Larrick and Fry, *Methods*, 2: 106-10, 1991.

[0131] Once the predisposition to develop the kidney disease has been determined, the information regarding predisposition can be presented to the individual and/or to the treating physician. This is particularly important in case of a positive prediction of kidney disease in a subject (i.e., when there is predisposition to develop kidney disease according to the methods described hereinabove), in order to take actions, which might prevent or delay the development of the kidney disease before its onset.

[0132] It should be noted that upon the onset (or occurrence) of the kidney disease, the subject's general health can deteriorate, and the subject can become severely ill, without cure, leading to death.

[0133] On the other hand, absence of predisposition to develop a kidney disease may be of a great relief to a subject being at risk thereof based on family history and/or presence of other disease(s).

[0134] Thus, according to some embodiments of the invention, the method further comprising informing the subject on the state of the predisposition to develop the kidney disease.

[0135] As used herein, the term "subject" includes mammals, preferably human beings at any age which suffer from a pathology (kidney disease) or is at risk of developing a pathology (kidney disease).

[0136] According to some embodiments of the invention, the kidney disease comprises an end stage kidney disease.

[0137] According to some embodiments of the invention, the kidney disease is not a diabetic-associated or diabetic-related kidney disease.

[0138] It should be noted that determination of the predisposition of the subject to develop a kidney disease can affect the subject's life style, such that the subject is aware of the risk to develop the disease and takes action(s) in order to prevent the kidney disease, which actions were otherwise not been taken.

[0139] For example, a subject who is predisposed to develop a kidney disease may be put on a restrictive diet for preventing overload of protein or salt on the kidney, and/or be treated with suitable medications which prevent or delay onset of the kidney disease.

[0140] According to an aspect of some embodiments of the invention there is provided a method of designing a life style change to a subject with the risk of kidney disease, comprising:

[0141] (a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation,

[0142] (i) wherein the variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions;

[0143] (ii) wherein the APOL1 nucleotide mutation is included in the APOL1 genomic sequence set forth in SEQ ID NO:3, and wherein the at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1,

[0144] wherein presence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation indicates increased predisposition of the subject to develop the kidney disease, and;

[0145] (b) designing the life style change based on presence or absence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation,

[0146] thereby designing life style change to a subject with the risk of kidney disease.

[0147] As mentioned, the knowledge of presence or absence or the predisposition to develop the kidney disease can be used to design a suitable treatment regimen to a subject having a certain disease and being at risk of developing also a kidney disease. Non-limiting examples of such subjects include those infected with HIV which causes AIDS, patients who have a single kidney, a past episode of acute kidney injury, have donated a kidney or otherwise lost a kidney, recurrent urinary infections, childhood reflux disease, exposure of kidney toxic medications, patients with hypertensive nephrosclerosis, non-monogenic focal segmental glomerulosclerosis, a patient with sickle-cell kidney disease and the like.

[0148] It should be noted that HIV-infected subjects are at risk to develop HIV-associated nephropathy (HIVAN). However, the threshold for introducing an antiretroviral therapy in patients with HIV does not currently take into account the patient's risk of HIV-related kidney disease. Such risk could potentially be accounted for by a management algorithm that includes the effects of APOL1 genotype. Thus, determining the etiology of kidney dysfunction in a patient with AIDS who is receiving potentially nephrotoxic antiretroviral therapy may be assisted by APOL1 genotyping.

[0149] According to some embodiments of the invention, the subject is infected with HIV.

[0150] According to some embodiments of the invention, the kidney disease comprises HIV-associated nephropathy (HIVAN).

[0151] Thus, according to an aspect of some embodiments of the invention there is provided a method of designing a nephrotoxic anti-retroviral treatment regimen to a subject infected with HIV, comprising:

[0152] (a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation,

[0153] (i) wherein the variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions;

[0154] (ii) wherein the APOL1 nucleotide mutation is included in the APOL1 genomic sequence set forth in SEQ ID NO:3, and wherein the at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1,

[0155] wherein presence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation indicates increased predisposition of the subject to develop the kidney disease, and;

[0156] (b) designing the nephrotoxic anti-retroviral treatment regimen based on presence or absence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation,

[0157] thereby designing a nephrotoxic anti-retroviral treatment regimen to the subject infected with HIV.

[0158] For example, the clinician might decide to press on with therapy to alleviate the kidney injury in an HIV-infected patient carrying APOL1 risk alleles, or might choose not to implement such therapy in an HIV-infected patient who does not carry APOL1 risk alleles.

[0159] Similarly, determination of the predisposition to develop a kidney disease may have important considerations for both living kidney donors and recipients. For example, subjects having increased predisposition to develop a kidney disease may not be eligible to donate a kidney. Similarly, a subject having an increased predisposition to develop a kidney disease may not benefit from a kidney transplantation. Thus, the APOL1 genotyping may well be added to the battery of tests undertaken prior to living kidney donation, for the benefit of both donor and recipient.

[0160] Thus, according to an aspect of some embodiments of the invention, there is provided a method of determining if a subject is suitable for donating a kidney for transplantation, comprising:

[0161] (a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation,

[0162] (i) wherein the variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on the *Trypanosoma brucei rhodesiense* under identical assay conditions;

[0163] (ii) wherein the APOL1 nucleotide mutation is included in the APOL1 genomic sequence set forth in SEQ ID NO:3, and wherein the at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1,

[0164] wherein presence of the APOL1 polypeptide variant or the APOL1 nucleotide mutation indicates increased predisposition of the subject to develop a kidney disease, and;

[0165] wherein increased predisposition of the subject to develop the kidney disease indicates that the subject is not suitable for donating the kidney for transplantation,

[0166] thereby determining if the subject is suitable for donating a kidney for transplantation.

[0167] Current guidelines for the management of hypertension suggest that African Americans as a group do not enjoy renal benefit from antihypertensive therapy, and moreover the recommendations for use of specific antihypertensive drug categories in African American patients are influenced by which African subpopulation they belong to. The presence or absence of the APOL1 risk alleles for predisposition to a kidney disease may be also used to improve the individualization of therapy in the African American group of patients, such that those without APOL1 risk alleles might be expected to enjoy a substantial renal benefit from antihypertensive regimens, similar to those that have proven effective in patients of European ancestry.

[0168] It will be appreciated that the reagents utilized by the methods for determining predisposition to develop kidney disease according to some embodiments of the invention and which are described hereinabove can form a part of a kit.

[0169] Such a kit includes at least one reagent for determining a presence or absence in a homozygous or heterozygous form, of at least one APOL1 nucleotide mutation or APOL1 protein variant.

[0170] According to an aspect of some embodiments of the invention there is provided a kit for determining predisposition to a kidney disease, comprising a reagent capable of specifically detecting at least one APOL1 nucleotide mutation in the APOL1 genomic sequence set forth in SEQ ID NO:3, and/or a reagent capable of specifically detecting at least one APOL1 nucleotide mutation in the APOL1 coding sequence set forth in SEQ ID NO:2, wherein the at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, and/or a reagent capable of specifically detecting at least one APOL1 polypeptide variant, wherein said APOL1 polypeptide variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on said *Trypanosoma brucei rhodesiense* under identical assay conditions.

[0171] According to some embodiments of the invention, the APOL1 nucleotide mutation is selected from the group consisting of the guanine-containing allele of single nucleotide polymorphism (SNP) rs73885319 (SEQ ID NO:6), the guanine-containing allele of SNP rs60910145 (SEQ ID NO:7), the guanine-containing allele of SNP rs9622363 (SEQ ID NO:8), the adenine-containing allele of SNP rs60295735 (SEQ ID NO:9) and the cytosine-containing allele of SNP rs58384577 (SEQ ID NO:10).

[0172] According to some embodiments of the invention, the reagent comprises a polynucleotide capable of specifically detecting the APOL1 nucleotide mutation by any of the mutation and/or SNP detection methods described hereinabove.

[0173] According to some embodiments of the invention, the reagent comprises an antibody capable of specifically binding an APOL1 variant comprises the nucleotide mutation and not to the APOL1 wild type polypeptide.

[0174] According to preferred embodiments the APOL1 polypeptide variant which is detected by the reagent comprised in the kit has a mutation selected from the group consisting of G342, M384, N388-del and Y389-del with respect to the wild type amino acid sequence set forth in SEQ ID NO:1.

[0175] According to preferred embodiments the kit further comprising packaging material packaging at least one reagent and a notification in or on the packaging material. Such a notification identifies the kit for use in determining if an individual is predisposed to develop a kidney disease, and/or suitable to donate a kidney for transplantation, and/or suitable for receiving kidney transplantation, and/or for designing a life style change and/or for designing a treatment regimen for a subject.

[0176] The kit also includes the appropriate instructions for use and labels indicating FDA approval for use in vitro.

[0177] As used herein the term "about" refers to $\pm 10\%$.

[0178] The terms "comprises", "comprising", "includes", "including", "having" and their conjugates mean "including but not limited to".

[0179] The term "consisting of" means "including and limited to".

[0180] The term "consisting essentially of" means that the composition, method or structure may include additional ingredients, steps and/or parts, but only if the additional ingredients, steps and/or parts do not materially alter the basic and novel characteristics of the claimed composition, method or structure.

[0181] As used herein, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a compound" or "at least one compound" may include a plurality of compounds, including mixtures thereof. Throughout this application, various embodiments of this invention may be presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

[0182] Whenever a numerical range is indicated herein, it is meant to include any cited numeral (fractional or integral) within the indicated range. The phrases "ranging/ranges between" a first indicate number and a second indicate number and "ranging/ranges from" a first indicate number "to" a second indicate number are used herein interchangeably and are meant to include the first and second indicated numbers and all the fractional and integral numerals therebetween.

[0183] As used herein the term "method" refers to manners, means, techniques and procedures for accomplishing a given task including, but not limited to, those manners, means, techniques and procedures either known to, or readily developed from known manners, means, techniques and procedures by practitioners of the chemical, pharmacological, biological, biochemical and medical arts.

[0184] It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a

single embodiment, may also be provided separately or in any suitable subcombination or as suitable in any other described embodiment of the invention. Certain features described in the context of various embodiments are not to be considered essential features of those embodiments, unless the embodiment is inoperative without those elements.

[0185] Various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below find experimental support in the following examples.

EXAMPLES

[0186] Reference is now made to the following examples, which together with the above descriptions illustrate some embodiments of the invention in a non limiting fashion.

[0187] Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical, microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A Laboratory Manual" Sambrook et al., (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel et al., "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Md. (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson et al., "Recombinant DNA", Scientific American Books, New York; Birren et al. (eds) "Genome Analysis: A Laboratory Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A Laboratory Handbook", Volumes I-III Cellis, J. E., ed. (1994); "Current Protocols in Immunology" Volumes I-III Coligan J. E., ed. (1994); Stites et al. (eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, Conn. (1994); Mishell and Shiigi (eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980); available immunoassays are extensively described in the patent and scientific literature, see, for example, U.S. Pat. Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521; "Oligonucleotide Synthesis" Gait, M. J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., Eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning" Perbal, B., (1984) and "Methods in Enzymology" Vol. 1-317, Academic Press; "PCR Protocols: A Guide To Methods And Applications", Academic Press, San Diego, Calif. (1990); Marshak et al., "Strategies for Protein Purification and Characterization—A Laboratory Course Manual" CSHL Press (1996); all of which are incorporated by reference as if fully set forth herein. Other general references are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

General Materials and Experimental Methods

[0188] Sample sets—The non-diabetic ESKD sample set is a sub-set of a larger ESKD cohort previously reported (Behar et al. 2010). The designation of MYH9 associated nephropathies included hypertension affiliated chronic kidney disease,

HIV-associated nephropathy (HIVAN), and non-monogenic forms of focal segmental glomerulosclerosis (Bostrom and Freedman 2010). The sample includes 430 non-diabetic ESKD cases and 525 controls, all of whom are self-identified as either African American or Hispanic American. The present inventors have already reported the African, European and Native American ancestry admixture proportions for these samples, based on a set of 40 genome wide Ancestry Informative Markers, and also the association and risk parameters for a set of 42 SNP polymorphic markers in the MYH9 genetic locus in the entire sample set (Behar et al. 2010).

[0189] The African populations sample set consists of 676 samples from 12 African populations, including Cameroon (2 ethnic groups), Congo, Ethiopia (4 ethnic groups), Ghana (2 groups), Malawi, Mozambique, Sudan with details provided in Table 2 hereinbelow. Whole Genome Amplification of extracted DNA was carried out as previously reported (van Eijk et al. 2010).

[0190] All sampling and testing was conducted with institutional review board approval for human genetic studies on anonymized samples, with informed consent.

[0191] SNP Genotyping—Genotyping of the novel six SNPs reported herein and in Tzur et al. 2010, i.e., SNPs rs73885319, rs60910145, rs9622363 [APOL1], rs60295735 [near APOL1], rs56767103 [FOXRED2], rs11089781 [APOL3], was performed with the KasPar methodology (Petkov et al. 2004). In addition PCR and RFLP reactions were designed for the APOL1 SNPs as follows. DNA fragments containing the both APOL1 SNPs [SNP rs73885319 (S342G) and SNP rs60910145 (I384M)] were generated by PCR amplification of genomic DNA using the following primers: forward primer: 5'-ACA AGC CCA AGC CCA CGA CC-3' (SEQ ID NO:4) and the reverse primer: 5'-CCT GGC CCC TGC CAG GCA TA-3' (SEQ ID NO:5). PCR reaction mix included 30 ng template DNA with 7.5 pmol of each primer and the Red Load Taq Master (Larova). PCR conditions were 95° C. for 3 minutes followed by 40 cycles of denaturation at 95° C. for 30 seconds, annealing 65° C. for 20 seconds, and elongation 72° C. for 1 minute. The resulting amplicon was digested with both endonucleases HindIII and NspI and was run on a 2% agarose gel. SNP APOL1 rs73885319 A/G: The allele containing G eliminates a recognition site of endonuclease HindIII. APOL1 SNP rs60910145 T/G: The allele containing T generates a new recognition site of endonuclease NspI. An example for RFLP genotyping of APOL1 rs73885319 A/G and APOL1 SNP rs60910145 T/G is shown in FIG. 6. SNP validation was performed by Sanger sequencing.

Example 1

Search for Sequence Variants Associated with ESKD

[0192] In order to detect functional mutations which are significantly more associated with ESKD than all previously reported SNPs in MYH9, the present inventors have re-examined the MALD interval surrounding MYH9, and detected novel missense mutations with predicted functional effects in the neighboring loci such as the APOL1 gene.

[0193] Identification of candidate non-synonymous SNPs related to ESKD—To search for variants possibly associated with ESKD, the present inventors analyzed 119 whole genome sequences recently released by the 1000 Genomes Project [world wide web (dot) 1000 genomes (dot) org] and examined in these genomes a 1.55 Mbp interval surrounding

MYH9, spanning nucleotide positions 34,000,000 to 35,550,000 (NCBI36 assembly). Of the 119 whole genome sequences, 60 are of European origin (HapMap CEU cohort), and 59 are of West-African origin (HapMap YRI cohort) yielding a total of 7,479 SNPs in the 1.55 Mbp chromosome 22 interval. Filtering criteria were applied to identify candidates for further consideration and analysis, based on: (1) low allele frequency in CEU but not in YRI, and (2) LD patterns with the previously identified leading MYH9 risk variants. From the 7,479 SNPs, the present inventors have selected for further examination all SNPs which complied with the following conditions:

[0194] (1) Minor allele frequency in the CEU cohort not exceeding 7.5% (9/120 chromosomes). This minor allele was designated as a putative “risk state” for this SNP.

[0195] (2) Risk state in the YRI cohort at allele frequency exceeding 17.5% (21/118 chromosomes).

[0196] (3) A minimal level of LD with the MYH9 S-1 SNP rs5750250 (FIG. 2). The criterion used was a chi-square test p-value not exceeding 0.15 for the 3*3 genotype table comparing each candidate SNP to the S-1 SNP in the YRI cohort.

[0197] All candidate SNPs which passed these requirements (total 250) were inspected in order to identify non-synonymous exonic SNPs, indicating a possible functional role. These were again examined for consistency with association patterns in the leading MYH9 risk variants S-1 (rs5750250) and F-1 (rs11912763) (Nelson et al. 2010) to confirm higher prevalence of the African “risk” state in the YRI cohort, and rarity of the risk state in the S-1 protective state (G:G), given the high attributable risk of S-1, especially for HIVAN (Winkler et al. 2010), indicating that the causative variant should be extremely rare in the presence of the S-1 protective state. This yielded four candidate non-synonymous exonic SNPs for genotyping: The first two (rs73885319 and rs60910145) are missense mutations in the last exon of the APOL1 gene (S342G and I384M) which is the neighboring gene, located 14 kbp 3' downstream from MYH9. A third SNP (rs11089781) is a nonsense mutation (Q58X) in the APOL3 gene located 110 kbp further 3' downstream. The fourth SNP (rs56767103) is a missense mutation (R71C) in the gene FOXRED2 located 100 kbp upstream to the 5' side of MYH9 (FIG. 1). Of note the two variants located 128 bp apart in APOL1 are in almost perfect LD (237 out of 238 chromosomes from the 1000 Genomes Project).

[0198] These four variants were genotyped in a previously reported sample set of African American and Hispanic American cases and controls (n=955) (Behar et al. 2010). Associations of these candidates with ESKD phenotypes previously attributed to MYH9, were determined using logistic regression, with correction for global and local ancestry, and considering three major modes of inheritance as previously reported (Behar et al. 2010).

[0199] Association analysis—For determining the association of each SNP with ESKD in the dataset, the present inventors performed logistic regression with ESKD status (Cases/Control) as the response, and included covariates for local and global ancestry as well as for cohort (African American/Hispanic American). Ancestry estimates were calculated as previously described (Behar et al. 2010). In addition to the four exonic SNPs described above, two additional SNPs in the APOL1 region were chosen for genotyping, including SNP rs9622363 located in intron 3, and SNP rs60295735 located in the inter-genic region between the genes APOL1 and MYH9. These SNPs were chosen using the same criteria as those applied to the exonic SNPs, as described above.

TABLE 1

Association with non-diabetic ESKD of non-synonymous SNPs in APOL1, APOL3 and FOXRED2 in the MALD peak and comparison with leading MYH9 SNPs									
rs number	Gene	Type	Chr22 Location ^a	Alleles ^b	YRI risk frequency ^c	CEU risk frequency	Mode ^d	OR	P value
rs73885319 ^e	APOL1	exon 5	36661906	A/G	46%	0%	Recessive	6.7	2.71E-06
		S342G					Additive	2.22	2.38E-08
		missense					Dominant	2.23	8.11E-06
rs60910145	APOL1	exon 5	36662034	T/G	45%	0%	Recessive	6.74	9.89E-06
		1384M					Additive	2.28	3.00E-08
		missense					Dominant	2.32	4.75E-06
rs11089781	APOL3	exon 1	36556768	G/A	31%	0%	Recessive	6.62	2.82E-03
		Q58X					Additive	2.18	3.79E-06
		nonsense					Dominant	2.22	3.23E-05
rs56767103	FOXRED2	exon 1	36902259	G/A	18%	0%	Recessive	1.33	6.83E-01
		R71C					Additive	1.52	5.19E-02
		missense					Dominant	1.66	3.64E-02
rs11912763	MYH9	intron 33	36684722	G/A	48%	0%	Recessive	2.38	2.86E-02
		F-1					Additive	1.96	4.05E-05
		designation					Dominant	2.28	4.20E-05
rs5750250	MYH9	intron13	36708483	A/G	66%	6%	Recessive	2.48	4.29E-05
		S-1					Additive	1.78	6.68E-05
		designation					Dominant	1.55	4.97E-02

^aLocation on Chromosome 22 in NCBI37.1/hg19 assembly.^bAfrican ESKD "risk" state in bold.^cFrequencies according to available 1000genome data.^dAssociation results were derived using logistic regression, correcting for global and local African ancestry, and combining the Hispanic and African American cohorts.^eSee FIGS. 5A-D for allele frequency pie-charts in cases versus controls.

[0200] The APOL1 Missense Variants (rs73885319 and rs60910145) are More Strongly Associated with ESKD Risk than the Leading MYH9 Risk Variants

[0201] Table 1 hereinabove shows the association results for the combined dataset of the African American and Hispanic American cohorts. For comparison, the present inventors include the results from the two previously reported leading MYH9 risk variants (S-1 rs5750250, F-1 rs11912763) (Nelson et al. 2010). The APOL1 missense variants (rs73885319 and rs60910145) are more strongly associated with ESKD risk than the leading MYH9 risk variants, both in terms of OR and p values (Table 1 above). The lower allele frequency and OR, with a higher p value for the APOL3 nonsense variant, and the weak association for the FOXRED2 missense mutation, render these variants unlikely candidates to explain the risk attributed to this genomic region. The results for combined and meta-analysis of the two separate cohort-based results are congruent (Table 2, below).

[0202] For determining whether the APOL1 SNP rs73885319 explains the association of ESKD with MYH9 SNPs, the present inventors performed a logistic regression with the same covariates, but included both the APOL1 SNP and an MYH9 SNP in the analysis. To avoid committing to a

specific mode of inheritance, the present inventors included the SNPs as categorical variables with three values (three possible genotypes), and then performed an analysis of deviance on the results (Hastie and Pregibon 1992), first adding the APOL1 SNP, and then the MYH9 SNP, and performing a chi-square test for the null hypothesis that the MYH9 SNP does not add to the ability to explain ESKD status.

[0203] Analysis of deviance of the combined logistic regression indicates that LD with APOL1 SNP rs73885319 accounted for much or all the statistical association previously attributed to the leading MYH9 variants with ESKD. In this regard, the present inventors also examined two non-coding variants in the APOL1 region which are in high LD with the APOL1 missense mutations, and as expected, both showed significant disease risk association (FIG. 2 and Table 2, below).

[0204] Three major modes of association (recessive, additive, dominant) were tested through definition of appropriate dummy variables in the regression. In addition to this combined analysis, the regression analysis was performed in each cohort separately, and the resulting p-values were combined using Fisher's meta-analysis. Results are shown in Table 2 below.

TABLE 2

Association of the examined SNPs in the MALD peak with non-diabetic ESKD in African and Hispanic Americans									
Chr22					YRI risk	CEU risk	Hispanic American		
rs number	Location	Gene	Type	Alleles	freq.	freq.	Mode	OR	lower
rs73885319	36661906	APOL1	exon 5 S342G missense	A/G	0.457	0	Recessive	15.48	3.99
							Additive	3.59	2.21
							Dominant	3.47	1.95

TABLE 2-continued

rs60910145	36662034	APOL1	exon 5 1384M missense	T/G	0.449	0	Recessive Additive	12.80 3.54	3.28 2.17
rs60295735	36667154	Intergenic	Intergenic (APOL1- MYH9)	G/A	0.432	0	Dominant Recessive Additive	3.56 12.79 3.43	2.00 3.28 2.10
rs9622363	36656555	APOL1	intronic	A/G	0.711	0	Dominant Recessive Additive	3.32 5.11 2.80	1.87 2.36 1.79
rs56767103	36902259	FOXRED2	exon 1 R71C missense	G/A	0.177	0	Dominant Recessive Additive	2.03 Inf 2.75	1.15 0.00 1.40
rs11089781	36556768	APOL3	exon 1 Q58X nonsense	G/A	0.305	0	Dominant Recessive Additive	2.69 2.30 2.57	1.33 0.42 1.55
rs4821480	36695247	MYH9	intron23 E-1 designation	T/G	0.763	0.058	Dominant Recessive Additive	2.87 3.37 1.67	1.66 1.56 1.06
rs5750250	36708483	MYH9	intron31 S-1 designation	A/G	0.661	0.058	Dominant Recessive Additive	1.14 3.82 1.50	0.66 1.67 0.96
rs11912763	36684722	MYH9	intron33 F-1 designation	G/A	0.48	0	Dominant Recessive Additive	1.02 4.31 3.02	0.60 1.21 1.80
							Dominant	3.64	1.97

Association of the examined SNPs in the MALD peak with non-diabetic ESKD in African and Hispanic Americans

rs number	Hispanic American		African American				Meta- analysis	Combined analysis	
	upper	p-value	OR	lower	upper	p-value	p-value	OR	p-value
rs73885319	60.00	8.8E-04	4.86	2.35	10.06	3.5E-04	4.9E-06	6.70	2.7E-06
	5.83	1.5E-05	1.90	1.46	2.48	5.9E-05	2.0E-08	2.22	2.4E-08
	6.16	3.7E-04	1.89	1.34	2.67	2.2E-03	1.2E-05	2.23	8.1E-06
rs60910145	49.94	2.1E-03	11.13	5.05	2.29	7.4E-04	2.2E-05	6.74	9.9E-06
	5.78	2.3E-05	1.94	1.48	2.56	7.3E-05	3.5E-08	2.28	3.0E-08
	6.33	3.0E-04	1.95	1.37	2.76	1.8E-03	8.1E-06	2.32	4.8E-06
rs60295735	49.92	2.1E-03	2.99	1.56	5.73	5.8E-03	1.5E-04	4.27	1.2E-04
	5.60	3.6E-05	1.76	1.35	2.31	5.6E-04	3.8E-07	2.10	4.6E-07
	5.91	5.9E-04	1.87	1.31	2.66	3.5E-03	2.9E-05	2.22	1.5E-05
rs9622363	11.06	5.2E-04	3.68	2.46	5.52	1.2E-07	1.5E-09	3.92	6.3E-10
	4.53	4.2E-04	2.34	1.79	3.06	1.8E-07	1.8E-09	2.49	2.6E-10
	3.60	4.1E-02	2.26	1.41	3.61	4.3E-03	1.7E-03	2.20	3.2E-04
rs56767103	Inf	9.9E-01	1.02	0.33	3.17	9.8E-01	1.0E+00	1.33	6.8E-01
	5.41	1.4E-02	1.23	0.84	1.82	3.7E-01	3.3E-02	1.52	5.2E-02
	5.46	2.1E-02	1.33	0.85	2.10	3.0E-01	3.9E-02	1.66	3.6E-02
rs11089781	12.44	4.2E-01	13.06	2.43	70.14	1.2E-02	3.1E-02	6.62	2.8E-03
	4.24	2.0E-03	2.01	1.45	2.78	4.2E-04	1.3E-05	2.18	3.8E-06
	4.96	1.5E-03	1.93	1.32	2.82	4.3E-03	8.2E-05	2.22	3.2E-05
rs4821480	7.29	9.6E-03	1.82	1.24	2.69	1.1E-02	1.1E-03	2.05	6.6E-04
	2.63	6.4E-02	1.64	1.223	2.19	4.4E-03	2.6E-03	1.65	6.5E-04
	1.99	6.9E-01	1.92	1.10	3.36	5.6E-02	1.6E-01	1.47	1.0E-01
rs5750250	8.73	7.5E-03	2.29	1.54	3.43	6.7E-01	6.7E-05	2.48	4.3E-05
	2.34	1.4E-01	1.92	1.45	2.54	1.3E-04	2.1E-04	1.78	6.7E-05
	1.74	9.5E-01	2.28	1.37	3.81	7.8E-03	4.4E-02	1.55	5.0E-02
rs11912763	15.34	5.9E-02	1.95	0.95	3.99	1.3E-01	4.4E-02	2.38	2.9E-02
	5.08	4.7E-04	1.67	1.23	2.27	5.8E-03	3.8E-05	1.96	4.1E-05
	6.75	5.7E-04	1.90	1.29	2.79	6.2E-03	4.8E-05	2.28	4.2E-05

Location on Chromosome 22 in NCBI37.1/hg19 assembly.

Provided are the results of the association studies performed on SNPs in the APOL1, MYH9, FOXRED2 and APOL3 genes.

The Table also demonstrates similarity between the p values obtained from combining the p values from the separate cohort based analyses (African American, Hispanic American) in a meta analysis, and the p values obtained directly from a combined analysis of both cohorts including an indicator for cohort. This demonstrates the robustness of the statistical conclusions to variations in methodology.

[0205] Residual associations of MYH9 SNPs beyond LD with APOL1 missense mutations—As described above, the present inventors performed analysis of deviance tests of the hypothesis that the APOL1 missense mutation rs73885319 can account for the statistical association between MYH9 SNPs and ESKD status. Analysis of deviance shows that the associations of the E-1 and F-1 haplotype SNPs are satisfactorily explained by this mutation ($p > 0.5$ for F-1, $p > 0.1$ for E-1). For the S-1 SNP rs5750250, the present inventors obtained a borderline p -value of 0.01, indicating the possibility that the S-1 SNP carries an association signal beyond its LD with rs73885319.

Example 2

Distribution and Frequencies of the APOL1 Risk Allele in African Populations

[0206] HIV-associated nephropathy (HIVAN) has been considered as the most prominent non-diabetic form of kidney disease within what has been termed the MYH9-associated nephropathies (Kopp et al. 2008). Behar et al. 2006, reported absence of HIVAN in HIV infected Ethiopian.

[0207] The APOL1 S342G missense mutation is absent in certain African populations—To test whether genomic factors are associated with the absence of HIVAN in Ethiopian population, the present inventors have examined the allele frequencies of the APOL1 missense mutations, as well as the leading MYH9 risk variants in 676 individuals from 12 African populations, including 304 Ethiopians and the results are presented in Table 3, hereinbelow.

TABLE 3

Distribution of the rs73885319 (S342G) risk allele among various African populations					
Country	Population	Sample Size	Latitude	Longitude	rs73885319 risk allele frequency
Ghana	Bulsa	22	10.7	-1.3	11%
Ghana	Asante	35	5.8	-2.8	41%
Cameroon	Somie	65	6.45	11.45	16%
Congo	COG	55	-4.25	15.28	11%
Malawi	MWI	50	-13.95	33.7	12%
Mozambique	Sena	51	-17.45	35	12%
Sudan	Kordofan	30	13.08	30.35	0%
Cameroon	Far-North-CMR/Chad	64	12.5	14.5	1%
Ethiopia	Afar	76	12	41.5	0%
Ethiopia	Amhara	76	11.5	38.5	0%
Ethiopia	Oromo	76	9	38.7	0%
Ethiopia	Maale	76	7.6	37.2	0%

Table 3: Frequency of the risk allele for rs73885319 (S342G) in APOL1 among the African populations sample set (total $n = 676$), according to the location of each population. The samples analyzed and presented in this Table form part of the collection of DNA maintained by The Centre for Genetic Anthropology at University College London. Buccal cells were collected with informed consent and institutional ethics approval from anonymous donors unrelated at the paternal grandfather level, classified by self declared ethnic identity.

[0208] A reduced frequency of the APOL1 missense mutations (Table 3), and of the MYH9 risk variants is shown in North-East Africa populations (e.g., Sudan, Chad and Ethiopia), with the complete absence of the APOL1 missense mutations (as well as the F-1 risk variant) in Ethiopia (FIGS. 3A-C).

[0209] Analysis and Discussion

[0210] The present inventors have uncovered that the two missense mutations S342G and I384M in the APOL1 protein can explain the increased risk for kidney disease in African

Americans. The kidney diseases include: 1) Primary non-monogenic focal segmental glomerulosclerosis (FSGS), 2) Hypertension-attributed kidney disease (hypertensive nephrosclerosis), and 3) HW-associated nephropathy (HIVAN).

[0211] The associations of these diseases with APOL1 mutations have an odds-ratio (OR) between 7-10, the highest OR ever reported for complex disease. These allelic variants are common in African continental heritage populations in the United States, such as the African and Hispanic-American populations, as well as West African immigrant communities in New York City. It is estimated that ~13% of the African American population have a homozygous or compound heterozygous state of the APOL1 risk alleles that dramatically increase their risk for CKD and attendant cardiovascular and cerebrovascular disease [Tzur S, et al. Hum Genet 2010 128: 345-50; Rosset S, et al. Nat Rev Nephrol. May 3, 2011. [Epub ahead of print]; and Genovese G, et al., Science 2010; 329: 841-5].

[0212] Functional role of the APOL1 S342G missense mutation in *Trypanosoma brucei rhodesiense* infection—The APOL1 gene encodes apolipoprotein L-1, whose known activities include powerful trypanosomolysis (Lecordier et al. 2009). *Trypanosoma brucei rhodesiense* transmitted by tsetse flies prevalent in central and western Africa is a cause of human African trypanosomiasis, and shows resistance to this trypanosomolytic effect, by virtue of expressing serum resistance-associated protein (SRA), which interacts with the C-terminal domain of apolipoprotein L-1 (Lecordier et al. 2009). The S342G variant is predicted to modify the binding site of the C-terminal domain of the APOL1 gene product (FIGS. 4A-C).

[0213] With respect to kidney disease risk, APOL1 is also prominently involved in autophagic pathways (Zhaorigetu et al. 2008), and a recent study has provided compelling evidence for the role of well preserved autophagy in the integrity of renal glomerular podocytes (Hartleben et al. 2010). Moreover, apolipoproteins have also been identified as circulating inhibitors of glomerular proteinuria (Candiano et al. 2001).

[0214] Functional studies, combined with clinical and genetic evidence—point to a “gain-of-injury” model, as opposed to a “loss-of-function” model for the mechanism of APOL1 mediated risk. Moreover, preliminary results indicate that the mechanistic distinction between the risk (African risk variant states) and non-risk (non-African protective states) of the gene, relate to access to its target of injury.

[0215] The APOL1 genetic markers can be useful in the clinical setting. For example, the threshold for introducing antiretroviral therapy in patients with HIV does not currently take into account the patient’s risk of HW-related kidney disease. Such risk could potentially be accounted for by a management algorithm that includes the effects of APOL1 genotype. So too, determining the etiology of kidney dysfunction in a patient with AIDS who is receiving potentially nephrotoxic antiretroviral therapy may be assisted by APOL1 genotyping. For example, the clinician might decide to press on with therapy to alleviate the kidney injury in a patient carrying APOL1 risk alleles, or might choose not to implement such therapy in a patient who does not carry APOL1 risk alleles.

[0216] With regard to transplantation, APOL1 genotype may have important considerations for both living kidney donors and recipients. African American kidney donors have poorer long-term renal function than European American

donors. If this observation turns out to be restricted to the subgroup of donors who are in the APOL1 risk category, APOL1 genotyping may well be added to the battery of tests undertaken prior to living kidney donation, for the benefit of both donor and recipient.

[0217] New insights may also be forthcoming with regard to the recurrence of focal segmental glomerulosclerosis following kidney transplantation. One these studies, implicated apolipoproteins among the circulating factors that govern recurrent proteinuria following transplantation in focal segmental glomerulosclerosis. Thus advance knowledge of genotype might guide preparatory therapy (e.g. plasmapheresis) for this possible outcome, or suggest new preventive or therapeutic approached.

[0218] Current guidelines for the management of hypertension suggest that African Americans as a group do not enjoy renal benefit from antihypertensive therapy, and moreover the recommendations for use of specific antihypertensive drug categories in African American patients are influenced by which African subpopulation they belong to. APOL1 genotyping may improve the individualization of therapy in this very large group of patients, such that those without APOL1 risk alleles might be expected to enjoy a substantial renal benefit from antihypertensive regimens, similar to those that have proven effective in patients of European ancestry.

[0219] Indeed, APOL1 seems to offer a particularly enlightening example of a gene with DNA sequence variants that are common in the population, and which confer an advantage under one set of environmental conditions (trypanosomiasis epidemics) but may have deleterious effects in the face of a different set of environmental conditions (such as HIV infection and increased life expectancy). This interplay is part of the emerging discipline of evolutionary medicine and validates in the clinical setting the maxim of the eminent geneticist Theodosius Dobzhansky, "Nothing in biology makes sense except in the light of evolution".

[0220] Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims.

[0221] All publications, patents and patent applications mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention. To the extent that section headings are used, they should not be construed as necessarily limiting.

REFERENCES

(Additional References are Cited in Text)

[0222] Barrett J C, Fry B, Maller J, Daly M J (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263-5;

[0223] Behar D M, Rosset S, Tzur S, Selig S, Yudkovsky G, Bercovici S, Kopp J B, Winkler C A, Nelson G W, Wasser W G, Skorecki K (2010) African ancestry allelic variation at the

MYH9 gene contributes to increased susceptibility to non-diabetic end-stage kidney disease in Hispanic Americans. *Hum Mol Genet* 19: 1816-27;

[0224] Behar D M, Shlush L I, Maor C, Lorber M, Skorecki K (2006) Absence of HIV-associated nephropathy in Ethiopians. *Am J Kidney Dis* 47: 88-94;

[0225] Billas I M, Iwema T, Gamier J M, Mitschler A, Rochel N, Moras D (2003) Structural adaptability in the ligand-binding pocket of the ecdysone hormone receptor. *Nature* 426: 91-6;

[0226] Bostrom M A, Freedman B I (2010) The Spectrum of MYH9-Associated Nephropathy. *Clin J Am Soc Nephrol*;

[0227] Candiano G, Musante L, Carraro M, Faccini L, Campanacci L, Zennaro C, Artero M, Ginevri F, Perfumo F, Gusmano R, Ghiggeri G M (2001) Apolipoproteins prevent glomerular albumin permeability induced in vitro by serum from patients with focal segmental glomerulosclerosis. *J Am Soc Nephrol* 12: 143-50;

[0228] Hartleben B, Godel M, Meyer-Schwesinger C, Liu S, Ulrich T, Kobler S, Wiech T, Grahmmer F, Arnold S J, Lindenmeyer M T, Cohen C D, Pavenstadt H, Kerjaschki D, Mizushima N, Shaw A S, Walz G, Huber T B (2010) Autophagy influences glomerular disease susceptibility and maintains podocyte homeostasis in aging mice. *J Clin Invest* 120: 1084-96;

[0229] Hastie T J, Pregibon D (1992) In: Chambers J M, Hastie T J (eds) *Statistical Models in S*. Wadsworth & Brooks;

[0230] Kao W H, Klag M J, Meoni L A, Reich D, Berthier-Schaad Y, Li M, Coresh J, Patterson N, Tandon A, Powe N R, Fink N E, Sadler J H, Weir M R, Abboud H E, Adler S G, Divers J, Iyengar S K, Freedman B I, Kimmel P L, Knowler W C, Kohn O F, Kramp K, Leehey D J, Nicholas S B, Pahl M V, Schelling J R, Sedor J R, Thornley-Brown D, Winkler C A, Smith M W, Parekh R S (2008) MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nat Genet* 40: 1185-92;

[0231] Kopp J B, Smith M W, Nelson G W, Johnson R C, Freedman B I, Bowden D W, Oleksyk T, McKenzie L M, Kajiyama H, Ahuja T S, Berns J S, Briggs W, Cho M E, Dart R A, Kimmel P L, Korbet S M, Michel D M, Mokrzycki M H, Schelling J R, Simon E, Trachtman H, Vlahov D, Winkler C A (2008) MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nat Genet* 40: 1175-84;

[0232] Lecordier L, Vanhollebeke B, Poelvoorde P, Tebabi P, Paturiaux-Hanocq F, Andris F, Lins L, Pays E (2009) C-terminal mutants of apolipoprotein L-I efficiently kill both *Trypanosoma brucei brucei* and *Trypanosoma brucei rhodesiense*. *PLoS Pathog* 5: e1000685;

[0233] Nelson G W, Freedman B I, Bowden D W, Langefeld C D, An P, Hicks P J, Bostrom M A, Johnson R C, Kopp J B, Winkler C A (2010) Dense mapping of MYH9 localizes the strongest kidney disease associations to the region of introns 13 to 15. *Hum Mol Genet* 19: 1805-15;

[0234] Pettersen E F, Goddard T D, Huang C C, Couch G S, Greenblatt D M, Meng E C, Ferrin T E (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-12;

[0235] Petkov P M, Ding Y, Cassell M A, Zhang W, Wagner G, Sargent E E, Asquith S, Crew V, Johnson K A, Robinson P, Scott V E, Wiles M V (2004) An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res* 14: 1806-11;

[0236] van Eijk R, van Puijenbroek M, Chhatta A R, Gupta N, Vossen R H, Lips E H, Cleton-Jansen A M, Morreau H, van

Wezel T (2010) Sensitive and specific KRAS somatic mutation analysis on whole-genome amplified DNA from archival tissues. *J Mol Diagn* 12: 27-34;

[0237] Winkler C A, Nelson G, Oleksyk T K, Nava M B, Kopp J B (2010) Genetics of focal segmental glomerulosclerosis and human immunodeficiency virus-associated collapsing glomerulopathy: the role of MYH9 genetic variation. *Semin Nephrol* 30: 111-25;

[0238] Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9: 40;

[0239] Zhang Y (2009) I-TASSER: fully automated protein structure prediction in CASP8. *Proteins* 77 Suppl 9: 100-13;

[0240] Zhaorigetu S, Wan G, Kaini R, Jiang Z, Hu C A (2008) ApoL1, a BH3-only lipid-binding protein, induces autophagic cell death. *Autophagy* 4: 1079-82;

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 10

<210> SEQ ID NO 1
 <211> LENGTH: 398
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (342)..(342)
 <223> OTHER INFORMATION: Can be S or G
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (384)..(384)
 <223> OTHER INFORMATION: Can be I or M
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (388)..(388)
 <223> OTHER INFORMATION: May be missing
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (389)..(389)
 <223> OTHER INFORMATION: May be missing

<400> SEQUENCE: 1

```

Met Glu Gly Ala Ala Leu Leu Arg Val Ser Val Leu Cys Ile Trp Met
1           5           10          15

Ser Ala Leu Phe Leu Gly Val Gly Val Arg Ala Glu Glu Ala Gly Ala
20          25          30

Arg Val Gln Gln Asn Val Pro Ser Gly Thr Asp Thr Gly Asp Pro Gln
35          40          45

Ser Lys Pro Leu Gly Asp Trp Ala Ala Gly Thr Met Asp Pro Glu Ser
50          55          60

Ser Ile Phe Ile Glu Asp Ala Ile Lys Tyr Phe Lys Glu Lys Val Ser
65          70          75          80

Thr Gln Asn Leu Leu Leu Leu Thr Asp Asn Glu Ala Trp Asn Gly
85          90          95

Phe Val Ala Ala Ala Glu Leu Pro Arg Asn Glu Ala Asp Glu Leu Arg
100         105         110

Lys Ala Leu Asp Asn Leu Ala Arg Gln Met Ile Met Lys Asp Lys Asn
115         120         125

Trp His Asp Lys Gly Gln Gln Tyr Arg Asn Trp Phe Leu Lys Glu Phe
130         135         140

Pro Arg Leu Lys Ser Glu Leu Glu Asp Asn Ile Arg Arg Leu Arg Ala
145         150         155         160

Leu Ala Asp Gly Val Gln Lys Val His Lys Lys Thr Thr Ile Ala Asn
165         170         175

Val Val Ser Gly Ser Leu Ser Ile Ser Ser Gly Ile Leu Thr Leu Val
180         185         190

Gly Met Gly Leu Ala Pro Phe Thr Glu Gly Gly Ser Leu Val Leu Leu

```

-continued

195	200	205
Glu Pro Gly Met Glu Leu Gly Ile Thr Ala Ala Leu Thr Gly Ile Thr 210 215 220		
Ser Ser Thr Met Asp Tyr Gly Lys Lys Trp Trp Thr Gln Ala Gln Ala 225 230 235 240		
His Asp Leu Val Ile Lys Ser Leu Asp Lys Leu Lys Glu Val Arg Glu 245 250 255		
Phe Leu Gly Glu Asn Ile Ser Asn Phe Leu Ser Leu Ala Gly Asn Thr 260 265 270		
Tyr Gln Leu Thr Arg Gly Ile Gly Lys Asp Ile Arg Ala Leu Arg Arg 275 280 285		
Ala Arg Ala Asn Leu Gln Ser Val Pro His Ala Ser Ala Ser Arg Pro 290 295 300		
Arg Val Thr Glu Pro Ile Ser Ala Glu Ser Gly Glu Gln Val Glu Arg 305 310 315 320		
Val Asn Glu Pro Ser Ile Leu Glu Met Ser Arg Gly Val Lys Leu Thr 325 330 335		
Asp Val Ala Pro Val Xaa Phe Phe Leu Val Leu Asp Val Val Tyr Leu 340 345 350		
Val Tyr Glu Ser Lys His Leu His Glu Gly Ala Lys Ser Glu Thr Ala 355 360 365		
Glu Glu Leu Lys Lys Val Ala Gln Glu Leu Glu Glu Lys Leu Asn Xaa 370 375 380		
Leu Asn Asn Asn Tyr Lys Ile Leu Gln Ala Asp Gln Glu Leu 385 390 395		

<210> SEQ ID NO 2
 <211> LENGTH: 2924
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1231)..(1231)
 <223> OTHER INFORMATION: can be A or G
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1359)..(1359)
 <223> OTHER INFORMATION: can be T or G
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1369)..(1371)
 <223> OTHER INFORMATION: May be missing
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1372)..(1374)
 <223> OTHER INFORMATION: May be missing
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (2538)..(2538)
 <223> OTHER INFORMATION: can be T or C

<400> SEQUENCE: 2

gactttcact ttccctttcg aattcctcgg tatactcttg ggactggagg acctgtcttg	60
ttattataca gacgcataac tggaggtggg atccacacag ctcagaacag ctggatcttg	120
ctcagtctct gccaggggaa gattccttgg aggagcacac tgtctcaacc cctcttttcc	180
tgctcaagga ggaggccctg cagcgacatg gagggagctg ctttgctgag agtctctgtc	240
ctctgcatct ggatgagtgc acttttctt ggtgtgggag tgagggcaga ggaagctgga	300

-continued

gcgaggggtgc aacaaaaacgt tccaagtggg acagatactg gagatcctca aagtaagccc	360
ctcggtgact gggctgctgg caccatggac ccagagagca gtatctttat tgaggatgcc	420
attaagtatt tcaagaaaa agtgagcaca cagaatctgc tactcctgct gactgataat	480
gaggcctgga acggattcgt ggctgctgct gaactgcccc ggaatgaggc agatgagctc	540
cgtaaaagctc tggacaacct tgcaagacaa atgatcatga aagacaaaaa ctggcacgat	600
aaaggccagc agtacagaaa ctgggtttctg aaagagtttc ctcggttgaa aagtgaagctt	660
gaggataaca taagaaggct ccgtgccctt gcagatgggg ttcagaaggt ccacaaaggc	720
accaccatcg ccaatgtggt gtctggctct ctcagcattt cctctggcat cctgacctc	780
gtcggcatgg gtctggcacc cttcacagag ggaggcagcc ttgtactctt ggaacctggg	840
atggagttgg gaatcacagc cgctttgacc gggattacca gcagtaccat ggactacgga	900
aagaagtggg ggacacaagc ccaagccac gacctggtca tcaaaagcct tgacaaattg	960
aaggaggtga gggagttttt gggtgagaac atatccaact ttctttcctt agctggcaat	1020
acttaccac tcacacaggg cattgggaag gacatccgtg ccctcagacg agccagagcc	1080
aatcttcagt cagtaccgca tgcctcagcc tcacgcccc gggtcactga gccaatctca	1140
gctgaaaagc gtgaacaggt ggagaggggt aatgaacca gcctcctgga aatgagcaga	1200
ggagtcaagc tcacggatgt ggccctgtg ngcttcttct ttgtgctgga tgtagtctac	1260
ctcgtgtacg aatcaaagca cttacatgag ggggcaaagt cagagacagc tgaggagctg	1320
aagaaggtgg ctcaggagct ggaggagaag ctaaacatnc tcaacaataa ttataagatt	1380
ctgcaggcgg accaagaact gtgaccacag ggcagggcag ccaccaggag agatatgcct	1440
ggcaggggccc aggacaaaat gcaaactttt tttttttct gagacagagt cttgctctgt	1500
cgccaagtgt gagtgcattg gtgcgatctc agctcactgc aagctctgcc tcccggttc	1560
aagcgattct cctgccttgg cctcccaagt agctgggact acaggcgcc accaccatgc	1620
ccagctaatt tttgtatttt taatagagat ggggtttcac catgttgcc aggatggtct	1680
cgatctcctg acctcttgat ctgccacct tggcctccca aagtgtggg attacaggcg	1740
tgagccatcg cttttgaccc aaatgcaaac attttattag ggggataaag aggggtgaggt	1800
aaagtttatg gaactgagtg ttagggactt tggcatttcc atagctgagc acagcagggg	1860
aggggttaat gcagatggca gtgcagcaag gagaaggcag gaacattgga gctgcaata	1920
agggaaaaat gggaaactgga gagtgtgggg aatgggaaga agcagtttac tttagactaa	1980
agaatatatt ggggggcccgt gtgtagtggc tcatgcctgt aatccagca ctttgggagg	2040
ccaaggcggg cggatcacga ggtcaggaga tcgagaccat cctggctaac acagtgaaac	2100
cccgctctca ctaaaaatac aaaaaattag ccgggcatgg tggcgggcgc ctgtagttcc	2160
agctaactgg gcggctgagg caggagaatg gcgtgaacct gggaggtgga gcttgcaagt	2220
agccgagata tcgccactgc actccagcct gggtgacaga gcgagactcc atctcaaaaa	2280
aaaaaaaaaa aagaatatat tgacggaaga atagagagga ggcttgaagg aaccagcaat	2340
gagaaggcca ggaagaaaa gagctgaaaa tggagaaaagc ccaagagtta gaacagttgg	2400
atacaggaga agaaacagcg gctccactac agaccagcc ccagggtcaa tgtcctccga	2460
agaatgaagt ctttccctgg tgatggctcc ctgcctgtc tttccagcat ccactctccc	2520
ttgtcctcct gggggcanat ctcagtcagg cagcggcttc ctgatgatgg tcattggggg	2580

-continued

ggttgatcatg tgatgggtcc cctccaggtt actaaagggt gcatgtcccc tgettgaaca	2640
ctgaaggcca ggtgggtgggc catggccatg gtccccagct gaggagcagg tgtccctgag	2700
aacccaaact tcccagagag tatgtgagaa ccaaccaatg aaaacagtcc catcgctctt	2760
acccggtaag taaacagtca gaaaattagc atgaaagcag tttagcattg ggaggaagct	2820
cagatctcta gagctgtctt gtcgcgcccc aggattgacc tgtgtgtaag tcccaataaa	2880
ctcacctact catcaagctg gaaaaaaaaa aaaaaaaaaa aaaa	2924

<210> SEQ ID NO 3
 <211> LENGTH: 51060
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (34556)..(34556)
 <223> OTHER INFORMATION: can be A or G
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (39907)..(39907)
 <223> OTHER INFORMATION: can be A or G
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (40035)..(40035)
 <223> OTHER INFORMATION: can be G or T
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (40045)..(40047)
 <223> OTHER INFORMATION: May be missing
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (40048)..(40050)
 <223> OTHER INFORMATION: May be missing
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (41214)..(41214)
 <223> OTHER INFORMATION: can be C or T
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (45155)..(45155)
 <223> OTHER INFORMATION: can be A or G

<400> SEQUENCE: 3

tcttggcatc tcctttcatt tttcttttct tcccaaatta acctctttcc aacacagcac	60
ttcaagttcc cctggaactt catcggtcca tcggtgtccc ctggtctgga agagcaccga	120
ccacttgcca aggccccac cactctgctg ccatcacaac caccatcacc accattacga	180
gaaaaaacttg agactgatgg ccccaaaagc ttagaaggag ctgagagacc aaagaatgac	240
tcgaacaagt ccagcttggt gagtaggtga gtttattggg acttacacac aggtcaatcc	300
tgggcagcga caggacagct ctagagatct gtgcttctc ccaatgctaa actgctttca	360
tgctaattct ctgactgttt acttacggc taagagcgat gggactgttt tcattgggtg	420
gttctcagat actctctggg aagtttgggt tctcaggac acctgctcct cagctgggga	480
ttgtggccat ggccaccac ctgcctctca gggttcaagc aggggacatg caccctttag	540
taacctggag gggaccgcgc acatgacaac caccacaacc accatcatga agaagccact	600
ggctgactgt gatacaccac cagaaggaca agggagagtg gatgctggaa agacagagcg	660
agagaccatc accagggaaa gactttattc ttggaaggac atcaaacctg ggggggggtc	720
ggtagtgag ctgctgtttc ttctcctgta tccaacagtt ctaactctgg ttttctccat	780
tttcagctct ttcttttctc ggccttctca ttgctggttc ctgcaagctc cccctctatt	840

-continued

cttcgccccaa tatattcttt agtccaaagt aaactacttc ttcccatc ccacactctc	900
cagtctcttc tctccctta ttgcaggctc cagtgttctt gccttctctt tgggtgactt	960
gccatctgca ttaacccctc ccttgctgtg ctcagctaca caaatgccaa agtcactaac	1020
actcagttcc ataagcttta ccttgccctc cttatccccc taataaaatg tctgcatttt	1080
ggccaggcgc ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggtgggcaga	1140
tcacgaggtc aggagatgga gaccatcgtg gctaacacag tgaaccccg tctccagtga	1200
aaatacaaaa aaattagccg ggcgtggtgg caggtgcttg tagtccagc tactcgggag	1260
actgaggccg gagaatggtg tgaacccggg aggcggagtt tgcagtgagc cgagattgag	1320
ccactgcact ccatcctggg cgatagagcg agactccatc tcaaaaaaaaa aaaaaaaaaa	1380
aaaaaaaaagt ctgcattttg tcttgccctg tgcccgcat ttctgcccctg gtggctgcac	1440
tgctctgggg tcatgtgtct tggcctggct gcagcatctc atggatcttg gtgagaaagt	1500
tgagcttccc ctccagctcc tgagcccgct tcttcagctc ctcagctgac tctgactttg	1560
ccccctcaag caagtgtttt gactcatatg caaggctgac cacatccagc agaagcaaga	1620
tgctccagt ggtgcaccc acgatcatgg ttctctgct cattgcccctg gcggggcctt	1680
caacaacctt ctcaacctgt tcaccgcctt cagctgagat tcgcccattg atatgcgggg	1740
gtggggcata cgctcctaac tgagggttggt ctctggctcg tctgatggca cggatgttcc	1800
tcccaatccc ttgtgtgact tggtaaccaat tgtcaactaa ggtaagaaca ttgggtgtgt	1860
tcccacccac aaactccttc atcacctttg ctacattggt gccgcttttg tccaagtgc	1920
gggcttgggc tctgtcccgc aatttgttta ctagtctac cacactgcag gtaatcccag	1980
ccacagcagc tctgtctccc agaccatgc cagtgtccaa gagcacaaa ctgattcctt	2040
ctgtgaaggg tgccagaccc aggcgagga gggtcaggat gccagaggta gtgccaacag	2100
agttggacac cacattggca atggtggtgc ctctgtggac ctgctcaacc tctctgcaa	2160
gggcacggag ctcccttatg tgatcctcaa gctccctttt caaccgagga aactccttca	2220
aaaaccactg cctgtgctgc tggcttttat cgtggcggtt tttgtccttc atgacctgt	2280
gacttgcaag cttgttcaga gctttacgga gctcatctgc ctcacccctg cacaaggaaa	2340
gggttaaagga ttaagaaatg cagtttcctt atctgtgaaa tgagctccat gcgatctcta	2400
tctgtgtaa attgcagagc ttctactatc aatcaaatag aaacaaattt taaaaatgga	2460
attaattatc tttaaaactc aaacacattt tgttcacatc agatattcct cagtacttat	2520
tccatatgga aataaatatc ttaaaggcac ctcagatgct ctttcattgt aggtgtttga	2580
taaatgatgc tctcactca agggcaggtc aacaccgtga ctgccaaagt gtttgagag	2640
ggaaaactct gcagatgaac aaagagcagt gagagtgaga caaagagacc ttattaccag	2700
gatggcagga aagtgaagca tgaactttac ccaaagagag gccttcctt tgtcctcagc	2760
tctgggaag tgatctccag gccctggaa tgtcctgctt ggtaggaaca tctttctttg	2820
cctggtgatt tggccacggg acagtatagc gatgtgatgg atgatagggt ttggggctat	2880
ttggtgtctt tctgacttc cagaggactt ggggactaaa ggtgtgagat ctcagggaag	2940
ggctggagac tctagccatg agggcagctt gtgatccagc cccagcagat actctggaca	3000
ctaaggcatg ggttgagaaa gaaggacttt ctccttgacc aactccagct gggctcttca	3060
gagcttcttc tccactaggc gccgacctg gccttcaca ctcaccact gctttctgga	3120

-continued

cttgaccac	cttggctcag	caagaatctc	cccaaatacag	tttagggagg	gtcctccccg	3180
cttgctgtct	gagcaccgtt	tggtatctga	tcattgctcct	cacccccacc	actgggtggt	3240
gagggctctga	cctgcctttg	gcgagaatcc	tgtaggcca	gttttagcagg	aacccccac	3300
tctggcgtct	cctgtcagtc	ctgttccatc	tcccgcctcc	accctggtgt	tggtctgtgac	3360
ttccctctctg	tctttactat	atttgagct	gggctcagtc	tcctccccct	gtaacaacaa	3420
tctgaataa	aggcttcatt	tctggtttaa	tgtgattcac	aggaaatctt	ccttaactgg	3480
gtgagcttcc	ctggttggca	acaccgcatg	tgttttgtcc	cacattcaca	caggagaaca	3540
cagcctcctg	aggacaatga	agattcacac	ctgaaatcct	cccagactct	gccctgtgtg	3600
tctcttcctt	aggctgattt	taattctgtg	tcttccctg	tgataaacca	caatcgtgca	3660
atagcagttt	tcagtgtggt	acgtgagtc	ttctagttaa	ttataaaatc	taaggatgat	3720
tttaggaatc	cccatagatt	tgaggtggt	gtcaaaggta	agggtgccct	cagacattgc	3780
agtctggcta	tctctgagta	ctttgaattc	aaacaaaaac	aaaacctcag	aaagacagga	3840
gatcactttg	cctggcatct	gtcccttcc	aatgcctcat	gccctttcac	ctggctgaag	3900
gctactgggt	gcggccactg	ccccagctga	ctggagtggg	agatctggct	attcattctg	3960
gtccacatg	gacgtccag	ccaccttact	gcctcctgac	ctaccacctg	cacccccaac	4020
cttcagggtc	tactcgcag	gtgtcaccag	cctaactgat	gccttagcag	tgacagcttc	4080
acctgtgtgc	acagcaaggc	catgtgcaat	tttatggaac	cactagaaga	ctcaagacac	4140
tttcagaacc	atctatataa	attgctaatt	ctgaaaagac	ctgaatttgg	tcaagatggt	4200
agaggaagtg	tttctcctct	cggttgggac	tttgtgagtc	tgctcagtag	cctgaaatt	4260
acacaaattg	gcagcaactc	attctacttt	ggaaatggtc	actaaggcat	gccatggtaa	4320
ggagcttctc	tgtgtagaag	agaagaagg	ctgaaagggt	agtccagaac	tcaccagttc	4380
caaggcatgg	ggtgagagac	acaggtctac	atgacctgt	atgccatga	gtgttgggat	4440
catccagaca	gggaagaaat	atccctctcc	cactcacatg	tcctctgcaa	gttcattgac	4500
tccttaagtt	tcaagccctt	ggtgcctctc	tttccacacc	ctagtcccag	cctgggtgcc	4560
ccagccttcc	tgcttgcctc	accaaaccac	ggctgactct	tactccatcc	tgactccagg	4620
ctctgacact	ggtcccggtc	tctctgggct	tttgagcagg	atcttagtcc	ttgggttccc	4680
acctttgacc	tggagtgtgc	tctaaggcct	tcaccacact	cctatgacat	tgaattgcta	4740
ttctatttgg	acagaagaca	gttccaagtc	tcctatgacat	aagcccagag	cccaggaaat	4800
aggcctctca	ggatcctgga	ccagcagaag	ggctcctgctg	tggttgggga	attcttcaga	4860
tgagaaaaaa	ggccccagcc	agcatcccca	ggaacacata	ttgtttagg	taggattttt	4920
ctgacttttc	ccaaagggaag	tcctgctggg	ctccatctt	caagaagaag	atccttaacc	4980
ctggctctct	tcaccaagc	ttgtaagatg	ttaacgttaa	ctgtgcacct	actatgtgtc	5040
agtcattatg	ctaagcactt	tacatgcatg	atctcatttt	tctcataata	aaactccatt	5100
gatgaagaaa	ggctgaaagt	gagaaagtaa	ctggcaaggt	tcctatcacc	agcagatggc	5160
agaaatgacc	ctgttctctg	gggcttccca	tcctctcagc	ctgaggctac	tcactgccag	5220
cgaaggacat	ggaggaaata	ttcctcctgg	agtgcgaga	aggctcctct	caagcccagc	5280
agagggggct	gcctggagga	ggtgtgctg	tcagggacaa	ccagcccagg	ggagatctga	5340
gtggcatcgc	tggggcgccc	catggagtta	accccatgga	gcttacctgg	gcagttcagc	5400

-continued

agcagccacg aatccattcc aggccttcac atcagtcagc agttgtagca gattctctct	5460
gctcacttgg tcttggaat acttaaggta atcctcaata aagatactgc tctctagttg	5520
gaaagaagaa aggataaggt tggaggatag ttaggggag cataacagcc attgcacatt	5580
aggtgggttac atgttaattt ttcctggaca actgtgatgt gggacatgtt tgatggagtc	5640
atcagtgcata tagagggtt gtgtgcccc aaaattcata tgcagaagcc cttaaactga	5700
gtgactatat ttgaagattg ggcctttaag gaagcaatta agactgaatg agatcacaag	5760
ggtagggttc taatcctact agagggttag tgcctattg taagaagaga aagagaaaca	5820
ataaggccag gtaagcacac agcgcaaagg cagctgccgc aagccagcaa gagagtctc	5880
ccaaggctcc aacccttaa ttaataaac catcaaagac agaaaccca aaccctactg	5940
ggcccaactg gctctctga gtccacctgc acttccatcc ctgagcctgt gcttctgctt	6000
tgcagctgat ttgtttttt tttgactcaa atttttctc tcagtggagt caaaaactg	6060
gacaaggcta aggtctggtc tctctggat tcggagaact tcctaaaccc accagacca	6120
atgtgaccag gaattttcca aaaaggtaaa tcaaacacag cagttttgta actgcaacca	6180
atcaaataat gtccttattg cagtttcata tttactgat aaatactgt ctctgacact	6240
ttttcatcat aacatgaagc ctcttcagat ttggttttct caattcatga attccttctt	6300
attcaataaa atgttactgt gcctcagatt tttctttgac aaggggtatc tgagatcatg	6360
ggaatgaggc cctgcctgag gcagtcaccc ctggtcagca ggtgcttggt atgggcagtg	6420
cagggcggtt gccctaacc agatgagggg actgggtggc ttccacctgg aatcaacagg	6480
ggatagatag aggagataga ggggatgcaa actaattctt agggaggccc tactgtgtgg	6540
caggcttttg ccaggccagg tacccttata atatttaac ttcgtgacaa catggggaaa	6600
tattattttc atggtacacg agaggggagg taaggagctc caggtcacac agaagggact	6660
tgaaaaaggc cagaccgggg tccatttaag ttggaagctg ggtcaagcat gataattcc	6720
catagggccca ctcaaacag aggggctggt gcagggcaag agctggaggg aggggaccgg	6780
cctgtgctgg aaaaaacctt gggtcagcag ggagaagagc ggtagataag agatcagggc	6840
tgcaggteaa cctcctctcc caccctccc ctctcctgg accctgctct tccccactc	6900
ccctctcttg cttgttccac ttcctcatgg ctgtcctgcc ccacttgagt catctgctgg	6960
gtgacagggt agggaggaag tgggcacctg ggcgtgtca cccgaggag aggtcaggat	7020
gcagatgacc cccagacccc tgggtcgggg gactccactg gacctctct gctgctggg	7080
gcagactcat tggcctccta gtccatctgg gttcttctgg ggctcactca gctgtggagg	7140
tgccaccctc cattctaggt gcgagtagga actagccagg aaagaggaag cagacctacc	7200
tgggttcacg gtgccagcgg ctgggttacc gaggggcttt ccttgagct ctccagtcac	7260
tgtccagact ggaaggtttt ccttaacctt tgagaacgag aaaacaaatt gtgggatcga	7320
aggtgagcct cctgtgtaca gagggaggct ggaggggctg ggtctttgca gatccctctg	7380
tggtagcagc aaatgtccag agtgtttatc tctggggccc ctgccagcta gtatttacag	7440
aaactcacct cgttccagct tctctctccc tcaactctac accaaggcag ggtcctcttg	7500
tctgtagggt gtatgagaag aagataatca gaggggggag agccatctga tcattacaga	7560
aactacagag tctatacaca ggaatagaga tgggagggaa cattctgaca atgacctggg	7620
cctgggaaca tgtatgataa ctagttgatg atagccagtg tgtatttgtt gtctgtgatg	7680

-continued

tgcatggcct ctgacatgta tctcattcga tgatcacaca caggacagga gatgggtacc	7740
actttcccat ttgaacacag aggaatccac aaaagttagg gaacttatgg gtgttttcac	7800
aatagtgcta gagccaggac ttctgcttca atcccatgtc ttcaattact gtgctctact	7860
cccttaaata tacaatatgt agtcattgat acacgctcat ggtaaataac tctcatttgc	7920
aaatccctgt ctcacactct gtgctcaact cttcgtcttg gtgaactcac tgggtgtgca	7980
gtggacagcc tgtgtcccg ttgcttctct aagattcttt gtcccacaaa gacctggatc	8040
ctgctagatg ccggcccgag ccctgctgtg gctgccgagc cctcctgatg cttccccgtc	8100
ccctctgctg tctcccccga gctgatggac cttagggtgt ttctaattct tcatataaat	8160
tatgtacaaa agaatacct ccctcacgta aaactgctca catttgagat tatttcctta	8220
agatactttt aagagtagaa tgactaggtc aaaagctatt aacatctcaa aagtttttat	8280
acctagagcc actctgtatg aaaactgtac aaattcacac tcaactaac aggatatgaa	8340
gatgtccatt ctttcagtct tactggctct cagtagtctt ttcttcagga aaacaaaaca	8400
aaacaaaaca ttattccttg agacaaggtc tcaactcctag tatgcagccc aggtgtgtg	8460
caaaatcctg gccttagggg atcctccac cctgccctct aaagtgtgtg attacaaaaca	8520
tgagccacga ccccgccct gaataaatgc tgatctttat atagaggaaa gttcttcaca	8580
ttctgtcctt gattatctta ccctgggagg tgagggaaaa tagcaaatgt taatcatatt	8640
aatcatcttt aggcactgtt ttaagaagac agaagagcat ttatcaacct ccctcctga	8700
gatcaaaacta cagccatagt caaggcactt acaggtataa tggcacttaa tgttaatgtg	8760
ggcaaacatt ggtggatttt gcaccactaa ttgctctaag atgttataaa tagtaactca	8820
cagaactctt ccaagaacct aggaagtgtg taccattggt atttcatttt tcaggtaaga	8880
aaactgaaac agatgttcag tgacttgccc aagctcagaa tctactacgt ggtaaagact	8940
gggttccac ccagacagac ggactcaaga actcacgcac tgctctgca ccctctgtg	9000
ccaatgaaaa tttaaatgag ggcaacagga gatcagagat gccaagagat gttggaaaaac	9060
agaactggct gaagaaatga caactgaata aagaggggtg aagggagctg tgaaccacag	9120
caatgccaa agggcatcca gccaggcaga gtggtctact aggcagagcc tcagagagg	9180
gaggagtta ccaggccac gtgccatag ggattcatgc cagccttgag tctacaccag	9240
tgaaaaaat cagagagttc atgactgaag caatatgccc cagaaataat atctttgtat	9300
tctgacactt cagtgggttc ccattcaatc accctacagg gaagactgcc agtagacctg	9360
ccccccccc cgttcaatat gaatagatct aggatcagaa attggagaaa ggtgaaagt	9420
gggaaacaga ccaagatgac caaatacaaa aggagtccag aacaaacaag taatttttta	9480
aatctcaaac ccattgcttc agggagatta aaaaatattt tattcataaa cttaaaacaa	9540
gatatataga aaagaagaaa aatcaatgaa acaaacaaaa aaagagattt tctagaaact	9600
aaaaggtgat gatcaataa acaagtctag agagaccatg tgaagaaaag tcacggaaat	9660
cctccagaaa ataaacaaaa aagatgacag agggaaaagg gctatatggg gaatagaaaa	9720
gttaccaggc acagggtcac aggggtgtct aagaaagcta cttgagaatg agccaccggg	9780
aggaagaaag ggtgggtggg tgggtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtatgtg	9840
ttgtgcaccc atggagggca ggtggggata accaagaaag aaaaagacat gggatccaga	9900
aaacaggaaa atcaaccag agagaaggga atctgaggat cacagtggga catgcagtag	9960

-continued

gcctacaaaa	cgaccagtcc	aggatggaat	aaaaagggaa	tggttctgtg	ggaaaaagg	10020
actcaataga	atacctcaaa	ggattcacag	tttgagagaa	actgaggaaa	tgaagaaggc	10080
aagcagtaca	agggaggaaa	aaaaacaaag	gcaattagaa	acatcaggaa	aaagtaaagc	10140
tgtgccacaa	aacatgtgca	aacatggtgt	acaactcagc	tccacagtaa	atagtattta	10200
aaacatcaca	aaatgtaaac	actcttgatt	gatttttaac	tttgagaaca	gaacagagca	10260
tataaatatg	agaactaaat	atacttgta	ccgattagta	caaacttaaa	gcacaaaggt	10320
agcagttggg	atgtgaagca	ctgaacgaag	gagaaggagg	attaatatgt	tcatctttta	10380
aagaggggag	tcaagaaata	ctaggtaaaa	ctgagagtat	aagaaagaaa	gttgtaagta	10440
aattgtttac	agctacaaag	ttgctaattg	aggacataaa	actggtgata	caatattagg	10500
aaaaggtgga	gaggagagat	agagggagtg	tacaaacaag	atgaaatga	ctaataatga	10560
cactataagt	aataatgata	ataccgattc	aggcccatcc	cagttagggg	cgctccaaa	10620
atcaccattg	ttctttttct	accttggtgc	tctgcgcact	ccccggcaca	cccacttcca	10680
gcagctggtc	ttgttttgca	gacttttatt	taacagctca	gtggttcccg	ggcatcctcc	10740
agctcttata	caaaagcccc	ctcctcaaaa	gccactcgcc	tccctctttc	caettcccac	10800
ctctgtttca	gggtcgaggc	atccagatat	ctgtcttctg	gggcccctac	gcagcaagac	10860
taaaaggggg	accagtcct	gggcagcagt	tcaccagggg	agtatgcaga	ggggcatctg	10920
gacacttctt	gctgtaaggg	acgggggtgac	gggagggggg	cttcccttcc	ccttcttcct	10980
gctttgttca	ctctgtgaca	ccctaggcca	ggggagctgt	ctcaacgcca	tctcatgagc	11040
catgagaacc	aggaggacca	ggaggagaaa	gaggagcagg	agggcagaca	gagctggggc	11100
ctcggaacctg	cccaataact	gtgaggagct	gcatccagga	tcccatcctc	ctgggtcatt	11160
gttggcctgg	ctcccacctt	tgtctggagc	ctctgctggg	ggttgaggct	ggatactgcc	11220
cctcaggctt	gactaggaca	cagtgtgtac	tccatgggtg	agcctgcagg	ctgggtcaatt	11280
ccggtgtcca	actgagcgac	ctggaagagg	agccctccct	cccacctcag	ctctgagcca	11340
agctctccag	atgcagagga	cagggactct	cagcaaagca	tctccctcca	agtggttgtg	11400
gggcacctct	cttgggcagg	gaaagaggag	tcgaaatgat	ttcctgagaa	atcgtcattt	11460
tcattttatt	tttgagaaac	tgtcttgctc	tgtcaccag	tctggagtgc	tttggggcta	11520
tcacagcttg	ctgtagccta	gagccctggg	gctcaagtga	tccttctgcc	tcagccttcc	11580
aggtagctgg	gactacaggc	acacggcacc	atgcaggact	aatttgtaa	tttttcattt	11640
cagagagaca	gcatctcctg	tgttgcccaa	gctgtcttga	gctcctagca	tcatgtgatt	11700
ctccgcctc	tgcctccac	agcactgaat	tacaggctc	agccgtggca	ctgcacctt	11760
ctgaggaatc	ttcagtga	agccacggtc	aagttcaaac	cccatagcag	ggcacacagg	11820
ccctctgatg	ctctggaccc	caccttctc	accgtccct	gccaccacag	tcaccttagg	11880
tccaggcacc	cactgacgcc	ctgcttcggg	gcctcctgga	gcccactctac	tcctcccaact	11940
cccatcccca	gcctcaggcc	ttcactcagc	tgtcctctcg	ccctggaatg	tcctcccccac	12000
tgtcccccac	atgtgcaagt	cacctgtcct	cagagaccca	cccaaagctg	cctgggtccat	12060
gcgggggccc	ccaacctccc	ccaggctccc	tgagcaactct	gctttccaga	gggagctcta	12120
cctgtatgtg	gctctttggg	cctctgtctc	ccctttagac	tctgaacatt	caatggggac	12180
actgttgtgt	tcccagctgc	acccctacg	actgacctgg	atgccctccg	ccctccctgg	12240

-continued

cacagcctgg	ctgaaatccc	aggtgcagga	ggcattactt	gaactaaaag	caaccagcct	12300
ctggatgaca	tggcccagct	tcagtttcac	tttcagttga	gaaattgaat	ttaatttttc	12360
ttttatctct	gatgggcac	tcttttttgt	gaagaacaaa	taagaccatg	actcatattt	12420
cacagaaatt	acccccctga	tttcatggac	ttttggcagc	tctatcaagt	tttttagagg	12480
ggaaaaatct	taggctggga	ggggagggtga	cttgcccaga	gtcacacagc	tcttgagggg	12540
cagaggggag	gccaacatct	gcccttgtag	ctctgatatt	gaagcaaaaa	tgggtgattt	12600
taataatcca	acagtgtcca	ggtttgatcc	tgcttctaca	aacacagaga	cctaagctct	12660
cactcctgta	gttatactca	caattccaca	gtgtggcttt	tgaatctcat	gattgcaaag	12720
ctggcatcat	gccacaggta	ccattctgca	atgcgacctt	ctcactcccc	tgtctcagcc	12780
atgtttccca	ttagggacac	agcacagcct	gtttcttctc	tatcagctca	gaggtctgaa	12840
cagagggaca	gaggtgaggg	ccagagacgt	gaagtcctat	gatctgcctg	ctcagcctcc	12900
tgacacggga	ccctggcaag	tcattcccct	cccagggcct	cagtttcccc	acagggcctg	12960
gcacaagtag	gtgctcagac	tcagggtgag	tgagacaacc	tctgagttcc	ttgtgtctca	13020
aaagaaatct	gttgttccca	ttgggcttca	gcagcagcac	cctggggcca	gcctgggtgt	13080
ggcccggcca	caaggacatg	ccgggctcca	ggtcacctcc	cctcctccac	cttctttgat	13140
tccttcaaat	ttcccaaaga	gatgggcacc	cccaacacct	acctttcttc	tggggtcagc	13200
tggccgtgtc	tgagggtgtc	agaaccagag	ctgagcccg	ggagctttgt	gaaccatct	13260
aaagtgtttt	ccccacctct	ctctacagtt	taccgcacca	gcaggaggga	gggagcaatc	13320
agactcaagc	ctgggtgcaa	atcccggctc	tgccactgct	ttcctgtctg	atctgaacca	13380
gctacctaac	ctctctgagc	ttatctacaa	aagctgaatg	atccttccct	tatagagcta	13440
ttgcaagaat	aaggacatgg	gggtagatca	cactatcccc	aacttaccaa	gggatcttcc	13500
tctgacagag	actgagcaag	atccaactgt	tctgagctgt	gtggatccca	cgtccagctg	13560
tgcatctgtg	taataaccag	acacgtctcc	cggcctccca	gatataccct	ggaattcgaa	13620
agggaaaagt	aaagtacaaa	cttcccagca	gctcatgacc	aagcacagta	aacacgctgc	13680
ttcccagcac	cacctgcagt	ccaggcccag	cctccttget	gctgcactta	gaggagcagc	13740
cgagaccag	acatccgggt	ccctctcatc	caaccacact	gcctcacatc	ctcaggattc	13800
aaggctctgt	gtgtcagctg	cccatgtctc	aggcagtcag	aaccagagtt	gagcctgggg	13860
agttttttga	acctatctga	gccgcttgcc	ccccaacccc	cgcctcgatc	tctgcagttt	13920
aagggcacag	caggtggggg	agagaagtca	gactcaagcc	tcagtgaat	tccaggctct	13980
gcccctgttt	tactgtctga	tctgaactag	ttacctaac	tttctgagct	caagaaagac	14040
ctgcactga	ctgaccgtgg	ggctctccag	attactcccc	gtccttccgt	tccccacat	14100
tcctcagatg	cctgccttac	tcccacacat	gccatcagca	tcagcacagt	ccccacaca	14160
tgctgggtgg	atgcttattg	catatttatt	gtaatgtctc	tgaggttatg	ttttctgtta	14220
tgtttttgat	atagagtctt	gctctgtagc	ccaggctgga	gagcagtggc	acaatcatag	14280
ctacttcaag	tggcacaatc	aagccatcct	cccatctcag	cttccagata	gctgggatta	14340
caggtgcctg	ccactacacc	cagctaattt	tttgtaattt	tagtagagat	ggggtttcac	14400
catcttggcc	aggctggtct	tgaactctg	acctcaagt	atccacctgc	cttggcctcc	14460
caaagtgcct	ggattacagg	tgtgagccac	ggcgcccagc	ccattgtatc	cttctgtga	14520

-continued

ctttgtgtct	tcatagctta	gtcctacta	ataaatgaga	acatacaata	tttggttttc	14580
cattcctgag	ttacttcact	cagaataatg	gtgtccaact	tcattccagat	tgttacggat	14640
gccattattc	ccttcctttt	tacagctgag	taatattcca	tggcatatac	atatataggg	14700
gtgtgtgtgt	gtgtgtgtgt	gtgtgtatat	acacacacat	tttcttcac	cacttggttg	14760
ctgatgagca	tttaggttgg	ttccatattt	tttcgggttg	aaattatgca	gctataaaca	14820
tgcatgtgag	agtatctttt	ttatataata	acttttctcc	tctgggtaga	taccctggtag	14880
tgggattact	ggatcaaatg	gtattctact	tttagttttt	taaggagtgg	ccatactgtt	14940
ttccatagt	gttgtactag	tttacattcc	caccaccagt	gtaaaagt	ttccctgtca	15000
ccatattcac	accatttttag	ctaggctggg	cttgaactcc	tgacctcaag	tgatccacct	15060
gctttggcct	cccaaagtgc	tgggattaca	gggtgtgagc	actgcacct	tctattcttc	15120
tttgattttt	aaattatggc	cattcttgca	ggagtaagg	ggcaacttat	tgtagtttta	15180
atttgcat	ctctgataat	tagtgacgtt	gagcacattt	tcataatttt	ttggccattt	15240
cattttgtat	atcttctttt	gagaattgtc	tattcatgtc	ctttgccac	tttttgatgg	15300
gattattttt	ttttcttgc	tgattgttt	gagttccttg	tagaaactgg	atattagtcc	15360
tttgtcaaat	gcacagtttg	caaatatttt	ctcccacttt	gaggggtgtc	tgttttctct	15420
gctgattatt	tattttgctg	tgcagaagct	ttttacttta	attaggtccc	atctctttat	15480
ttttgttttc	gttgcattta	tttttgggtt	cttggctctg	aattctttgc	ctaagccaat	15540
gtctggagga	gtttttccga	tggtatcttc	tagaattttt	atgggttcag	gttttagatt	15600
taagtctttg	atccatcttg	agttgat	tgtatgagga	tactgtttca	ttcttctaca	15660
tggtggttgc	caattatcct	ggcaccattt	gttgaatagg	gtgtcctttc	cgcactatac	15720
gtttttcttt	gctttgtcaa	agactagtta	gctgaaagta	tttggcttta	tttctgggtt	15780
ctctattctg	ttcccttgat	ctatgtgcct	atttttatac	cgggtaccatt	ctgccattat	15840
gtaccttttt	tttttttttt	tttttttttt	tttgagatgg	agtctctctc	tgccacccaa	15900
gctggagtgc	agtgggtgcga	tcttggtctg	ctgcaagctc	cacctcctgg	gttcatgcta	15960
ttctctgccc	tcagcctccc	gagtagctgg	gactacaggc	gcccaccacc	acacctggct	16020
aatttttttt	tttttttttt	tttgtatttt	agtagagatg	gggtttcact	gtgttagcca	16080
ggatggtctc	gatctcctga	cctcgtgatc	tgctgcctt	agcctcccaa	gggtgctggga	16140
ttacagatgt	gagccactgc	accagccct	tgtacctttt	aagtgaacca	tttagcccat	16200
tttcattcaa	tgtagtatt	gagatataaa	gtgctgttct	attcatcagg	caagttgttg	16260
cctaaatacc	ttattttttt	cattgtgtta	ttgttttata	ggccctaagg	cttataaaac	16320
aatatgctgt	aaggagggtc	tattttgggt	tatttcgagg	tttgtttca	agacttagaa	16380
ctcgttttag	catttcttgt	agtactggct	tggtagtggc	gaattctttc	agcatttggt	16440
tgctgaaaa	agactttatc	tctccttc	ttatgaagct	tggttttgct	ggatacaaaa	16500
ttcttgacta	acaattattt	tgtttaagga	ggctaacaac	aggactccag	tcccttctgg	16560
cttgatgggt	ttctgctgat	aaattagctg	ttaatctcat	agattttcct	ttacaggtta	16620
ccttatgctt	ttgtctcaca	gctcttaaga	ttatttcctt	cttcttgact	ttagataccc	16680
tgatgacatg	tgctagggtg	aggatcattt	tgtgatgaat	ttcccagatg	ttctttgaac	16740
ttctgtatc	tggatatcta	gatctctagc	aaggccaagg	aagttttcct	caattattcc	16800

-continued

ctcaaataag	ttttccaaac	tttttagattt	ctcttcatcc	tcagcaacac	caattattct	16860
taaatttggc	catttaacat	aatctcaaat	ttcttggagg	ctttgctcat	tttttaaaat	16920
tcttttttct	ttgtctttgt	ctgattgggt	taatttgaaa	gctgtgtctt	caagctctga	16980
agttctttct	tctacttgtt	ctagtctatt	gttgaaactt	tccagtgcct	tttgtatttc	17040
tttcagtggt	tctttcattt	ccagaagctg	tgattgtttt	ttctttgtga	tatctatttc	17100
tctggagaat	tttcatccat	attctgtatt	ttttaaaatt	tctttaagtt	ggttttcacc	17160
ttctctgtct	atctccttga	atagcttaat	aatcaacctt	ctaaattctt	tatatggcaa	17220
ttcagagatt	ttttgttggc	ttgtatctat	tgctggggag	ctagcaatag	ctttgggggg	17280
tgttatagaa	atatgttttg	tcatattacc	agaattactt	ttccggttcc	ttctcatttg	17340
ggtagactat	ttcagtgga	agatctgaaa	ctcaactcct	tctgttcagg	ttcttttgtc	17400
ccacgggtta	ttccttgatg	tggtactctc	cttctcctag	gaatagggtt	tcctgagagc	17460
cagactgcag	tgattgttat	tgctcttctg	ggtagagcca	cccactgggg	ctataccagc	17520
ctttgggttg	gtgctgggga	atgtctgcaa	agagtcctgt	gatgtgatca	tctctgtgtc	17580
tcctagccat	ggataccagc	acctactctg	gtgtaggtga	caggcgagtg	aagtagactc	17640
cgtgagagtc	attggttgta	gttctcttta	gtgtgctggt	ttctccaatg	ctgggttatgc	17700
tagcagtgaa	gctgtcatgt	ggagagactg	aggacctctg	gtcagtcagg	atgttgccggg	17760
cagtggtatt	agctcttgte	ttctccttcc	tgggaatagg	gttattctgt	catgagttgc	17820
tgtaatggcc	tgagttgggt	ggtctccagc	caggaaagtg	cactttcaag	agagcaccag	17880
ctgcggtagt	agaagggaga	tacaagcttg	ccctagggtg	accagggtaa	gtattctagt	17940
ttctcaggtg	attggcaggg	ccataaagct	cccacgagtt	tatgtctttt	gtgttcagct	18000
accaggataa	gtagagaaat	accatcaggt	ggagggcagg	gttagctggg	tctgagctca	18060
gactttcctt	aagcagggtt	tgctgcggcc	actgtgaggg	atgcagattg	tcctaaggcc	18120
aatgggggtt	gttacagagg	ggattatggc	tgcctctgct	tgccatata	gttcgccagg	18180
gaagtggggg	atagctggca	gcgagaggcc	tcaccagat	cacatgcagt	tggcaaggct	18240
ggccttgctc	ccgcagtgcc	ccactaaca	tgccaagttt	agattcaggc	agcctgtgcg	18300
cagaactcag	accttgctcc	aggccataag	cttctccact	gagaaagcaa	gcattggcttt	18360
caggccttgc	ctctcccat	ctgccacag	gttcagtggt	ggctcctggg	ctcatatttg	18420
cagaagttcc	tgttctcccc	ctggattctg	ctcaagaaaa	ttcaagccca	gtagaaatta	18480
ttagaaaatc	catctggaag	cttctttcac	cctgtgaccc	ctcccaaatt	ctgctggctg	18540
ctttcccaa	aggcctctgt	gagataccgc	cagggatggt	gcctggggt	tgagctggac	18600
actgggaatg	cctacaggat	tcttccact	tctgcttcta	cttttacatt	ttgcatggtt	18660
ctctgaatcc	atttcagcta	tgggtaaggt	taaatccctc	tcctgtaatc	tggattttca	18720
gattccccag	gggcagatgt	tgtttgagg	cagattttcc	ctccctcaca	ctttgggaac	18780
tcccagtttt	ttgcctgtct	catagaattg	gcaaggagca	tgccacttct	ttcaaaggat	18840
ctgtgaatcc	tttcagtttt	cctggtacat	tcctgtgggt	gttcttgga	cacattatta	18900
catgatgtgg	gtctccacat	gctgtccggt	ccttccaagt	gggagctgca	tgtagccct	18960
gtctcctatc	tgccatcttc	ctcagacttc	taacattaac	ttttaaaaag	cttgggtggga	19020
agagaccagg	gccaagtctc	tccttcttgg	agatgaaagc	caagaagcac	ttcctttggg	19080

-continued

acaagtcaga	aaaatcctga	ccaaatcaat	gtgtgttctt	ggcgattctg	gcccaggcct	19140
atttcctccc	catctgagga	attgcacaaa	cacagcagaa	cccttctgct	ggcccaggac	19200
atggagaggc	tggagtgaag	ttgatcagag	ctgggatgga	tcctggcttc	tgtccctgca	19260
cagctgtggg	cctctaggat	tatgccttgg	agactggcat	ctggcgtctg	gcatctgtcc	19320
aatgggaata	gtaattcaat	catataggag	gatggatgaag	gccttagagt	gaactccagg	19380
tcaaagggtg	gggaccaagt	accatgatcc	tgccgaaaag	tccagataga	cctggaccag	19440
gggcagaacc	aggggtcagg	atgggcataa	gtgggcctaa	ttggaaggta	agtaggaagg	19500
ctgggacacc	agggccagga	ctagggtgag	aggagggagg	caccaaggac	accaaacgta	19560
aggaggtcat	tgacttgcag	aggacacatg	agtgggagag	ggatatttct	tcctccctg	19620
atttgatccc	atcagtgtat	ggcatagtgg	gccgtgagtg	acggtttccc	ccacaccaag	19680
cctgagaact	gtgagaactg	tggagttttt	gcgtgccctt	tgggaccctc	ttgtgttcta	19740
catagagtag	cccctgaccc	tgggaaagca	gtgtggctta	gtgaccactt	cgggagtcag	19800
aatgggtggc	tgccaatttt	tgtaatttca	ggggttctga	ggaaacccat	aagttgccaa	19860
caaagagga	aaacacttcg	tctaccatct	tgtccaaatt	caggtctttt	cagcattagc	19920
aatttatata	gatggttctg	aaagtgtctt	gagtcttcta	gaggctccat	aaaattgcac	19980
atggccttgg	caagcacaga	ggtgaagcca	tcactcctta	ggcatcgatt	aggetggtga	20040
catctgaaa	tgagacctgg	aaggttggag	gtgcaggtgg	taggtcagga	ggcagtaagg	20100
tggctggaga	atccatgtgg	actctccaaa	actaatagcc	aggtctcttc	ctccagtcaa	20160
ctggtgcggt	gcttggaaac	agtgaccatc	agtcagaagg	tcaaagggca	tgaggcattg	20220
gaagagagca	gatgccaggc	aaagtgtatc	tgtgtaggtc	tgtttttggt	tttgtttgaa	20280
ttctgagtac	tcagagttac	caaaactgca	atgtctgagg	gcacccttat	ctttgacact	20340
aactgcaaat	ttaaggggat	tcctaaatcc	gtccttagat	tttataattc	actagaggaa	20400
ttcacgtaac	tcactgaaaa	ctgttattgc	atgactgtgg	tttatcacag	ggaaaaaaca	20460
cagggtaaaa	tcagcctaag	gaagacacac	agggcagagt	ctaggagggt	ttcagggtgtg	20520
aagcttcggt	gtcctcagga	ggctgcgttc	tcctgtgtca	atgtgggaca	aaacacatgg	20580
ggtgttgcca	accagggaag	ctcaccagtg	taaggaaaaa	ttcctgtgaa	tcacattaaa	20640
tcagggaaga	agcctttatt	caggattgtt	gttacagggg	agggagactg	agcccagctc	20700
ccaatacagt	aaagacagag	ggggagtcac	agccaacacc	agagtgaggg	gcagcagatg	20760
gaacaggact	gacaggaggc	accagagtcg	ggggttcatg	ctaaactggc	ctaacaggat	20820
tttccctaaa	ggcaggtcag	acactcacc	atcaaggggc	aggtgaggag	agtgatcaga	20880
tagcaagtgt	gtcagacag	taagaaggga	ggaccctccc	taaactgact	cggggggatt	20940
cttgctgagc	cagggttgtg	caggccacgc	aagcacggga	tgggcatgga	aggccaaggt	21000
tgggtgcctag	aggagaagag	gtcagaaga	gcccactggg	agtctggtca	aacagaaagt	21060
ctttctcaac	ccatgcctta	gtgtccagag	tgtttgctgg	ggcttgacca	caggctgcac	21120
tcatggctga	cctggagtgt	ccagcccctc	cctgaggtct	cacaccttta	gtccccagat	21180
cctctggtgg	tgagaagaga	tacccaatag	ccccaaaccc	ccatcatcca	tcacatttgc	21240
tatactgtcc	tgtagccaaa	ccaccaggga	aacaaagaca	ttcctaccag	gcagggcatt	21300
ccaggggcct	ggagatcact	tcgcagggtg	aggggacaa	gggaagacct	ctcctgggat	21360

-continued

aaagttaatt	cttcaactgcg	caattgtagt	gcagtgaaga	ttctattccc	agaggccggg	21420
gaaggtttgc	aattgctttg	gtgcagtggg	gcattgggtaa	acttgcacct	tagctataac	21480
agcagacaga	acaacatag	tacaaatata	ggtagtcagt	tgtgcaactg	tgtattggta	21540
gacacaatga	cagcacatca	ccggctaacc	ccacaatatg	gggtgactac	ctgtacaaca	21600
atgtcaccaa	gcccttgggt	taagcaatga	tacatggctg	ggttggtaga	ataagtctct	21660
gcataatggg	gcctaccaat	ccccatggag	atggagatgg	ctgtgggcct	gtgtcaccca	21720
cccttggcct	tatttgccat	tcaactggac	tggcagatgc	atctcagggc	atccttaatt	21780
gccacagtat	atagtcccc	atttagacac	gacccccagt	aactggaaaa	cggtaaaagc	21840
tagagattac	cagatatgtg	catactgggt	gttctacccc	ttggccatct	ttgcctccaa	21900
agtggaact	gtaaatgcag	agctgcaatt	cgtgtgctgg	ctaaacacac	agctgcagcc	21960
ttcaatgata	ctcaccatcg	ttttacactt	cccaatcaag	aaaccaccca	gacagggcat	22020
gttcctctgc	aaaatcgtat	gacccaaaat	catacaactg	ctgcccattg	tggagcttgt	22080
gcattaatta	aaacaaaatg	ttgtgtatat	agccctcatt	actctcataa	cataacacag	22140
gctctgctga	cgtagatac	ccagagtcca	gccacagagt	ctttgtctta	tgatcttata	22200
acctcatggt	ttaatcaact	cccagacacg	tggagaaaact	ttgtatctag	cgtaattgca	22260
attgcgttca	taatcctttg	ttgaggttag	tgttgcctgc	ggggtgtctg	gctgcaatgc	22320
tcttcacaaa	aactggcccc	aggaaaaatc	cagaaaaaca	gtgtgatgag	attgtgaggg	22380
ccattcaagg	ctatgcaggg	gtgtggagag	catggtgggc	atgtcccgtc	agacagcaat	22440
aactgaagca	tcccctaaga	ataaccctac	attccttggg	gaatttttgc	tcagtgttcc	22500
aaggtagtga	tcccgggaat	ggccaatcca	gatgtttaca	tcatacttat	gaagaactct	22560
gttccttgga	tcagaggttg	tgcaagggaat	caagaccttt	tgttttgggc	taggtggagg	22620
ttccttgga	gaggtgctaa	gtggaggttg	ctctgggaag	aatatcatat	aacctgcatt	22680
catttgacaa	acaggagggg	atttcttctg	ttgcctgctg	ctcctgggca	acctgtacat	22740
aagctccctg	aataaagctt	atgtctcacc	tgctgtctcc	aggtcttttc	ttctgtctct	22800
ccaagtctgt	gtcctgtcag	ctcattagca	taggggtcca	gcattgacaaa	tataaaacgg	22860
tgtttaacat	tctttggatt	atgtttatat	aaatgtgtta	ttaatatgtg	ttccaaaatt	22920
gtacgagatt	ctataggtct	aatatgtctt	ggcatatggt	atcagaaaact	attattatta	22980
ttatgttaag	tagttgtttg	ccacagaaat	aaagtaattt	ccttgtcgaa	tgtgtcttta	23040
ccaatgtgtg	tctcatactt	ttgttatctg	caaaaagtgt	ttactttaca	tctcaaaaaa	23100
cagtttatga	tcagctacag	attgtgccat	cggactaaga	aaaaaataaa	tcttccagga	23160
ctgtaattca	aaagttagtg	tggtcatgaa	gattgctaac	ccaatatcaa	gtggaagaag	23220
gttattacat	gggaatgaat	taattggagaa	caaaaataaa	tttttaataa	ctttttgttt	23280
aatatacgac	tgattctttt	gttttgtttt	tcttagtcaa	gaaaacttct	ttttcccttt	23340
gggtccatg	tagcttttaa	caattgaata	aagtatactc	ttgtgagcaa	aatttgaagc	23400
atatttcttt	ctcactatct	gatttctcca	gaatttggaa	actatttgtg	agtattctta	23460
gttcattggc	atagagttat	ttgcataagt	tcagttggaa	tctgttttct	ttcataacag	23520
gacacaacta	gagacactgg	ttattttacc	aaggctttga	ctggcatggc	atatttccac	23580
atgtacgtag	gctgctttga	gtaattgggtg	ctggcttata	gagccgatga	aagcctcttt	23640

-continued

gaaaaat	ttt	cctcaaac	tt	tgccata	taca	gttccttt	tac	agggttc	ctg	tctgtg	gta	23700
agtaaga	aat	gtcact	ttt	ct	gacagg	ccca	ggaagccc	aa	gttac	cttg	gac	23760
agaaagga	ac	tcaccta	ata	tatatag	gca	tctgcag	gca	aataata	aatc	actgg	ctgag	23820
cttgagatt	tt	taaaagg	cct	actctga	gat	atcttat	gaa	aaagaaa	ata	ttgtag	caaa	23880
gtcaatt	tc	aaaagg	agcc	catatg	ccaa	ataagt	tattc	ttgctg	cact	ttata	caaat	23940
gatcagg	cca	agaac	catca	gactaaa	act	tctgtt	gcaa	ataaatt	agc	cctact	atga	24000
tttgtct	ttg	gtagaaa	tac	aggacta	gag	agagaaa	agg	tatgttt	caa	aagaa	actag	24060
agaatgc	cctg	ttattg	gatt	ctagct	ttt	ccattg	ttt	ctagcc	gtat	tgtttg	tcca	24120
cagtttat	ct	ggacta	catata	ctgaatt	cctt	tcctgg	ccac	aagtct	ccaa	actaat	at	24180
tgctttt	ctt	tttct	ctcat	ttttct	gact	tggaat	cacc	agaaat	ccaa	actgtg	cttt	24240
tcttaa	agtc	ctgcaa	actg	aagctt	gaaa	actgtg	tgtat	agtga	aggaa	gagaa	agacc	24300
ctctcat	att	tttttat	atat	tgtttt	gtac	tcagtac	ctg	tttta	agaaa	aaaca	acaag	24360
gaagta	aaaac	caaagac	agg	cagccc	ggcg	ccaggccc	ga	aaccagg	cct	gggccc	gcct	24420
ggccta	aaac	cattagt	ttaa	aaatca	actt	atgatt	caga	agccg	atg	ttcat	agat	24480
tccttac	att	atgtaga	aga	acattg	tga	actcc	ctg	ctgtt	ctg	ttt	ctgac	24540
gacctg	tgt	tgagccc	ct	gtcac	gtacc	gctgt	ctg	tcaaat	caat	catgac	ccct	24600
tcgtgt	gaaa	tcttt	agtg	tgtgag	ccct	taaa	aggga	agaaat	gtg	cactcag	gga	24660
gctcgga	ttt	taaggc	agta	gcttg	ccgat	gctccc	agct	gaataa	agcc	cttc	cttcta	24720
ctactcg	gtg	tctgag	agg	tttgt	ctgca	gctcg	tctg	ctaca	ataga	atcag	acca	24780
aaactg	tga	cagagac	tca	tttg	cttcta	cagtgt	tttg	tccaaa	agat	gttga	agaac	24840
ggccagg	cg	agtg	ctc	gcctg	taac	ccagc	gctt	cggagg	ccga	ggcagg	caga	24900
tcacaac	gcc	aagagat	cga	gaccat	gctg	gccaa	gatg	tgaaac	gctg	tctgt	actaa	24960
aaataca	aaaa	attag	ctggg	cttg	gtggtg	cg	tg	cctg	ta	cttggg	aggg	25020
tgaggc	agga	gaatc	acttg	aaccc	gggag	gtggag	gtg	cagtga	gccc	agatc	acaac	25080
actgc	actcc	agcct	ggcaa	cagag	caaga	atctgt	caaa	aaaaaaaa	aaaaaa	agt		25140
tgttga	agaa	caagg	agaaa	acgtg	aaagg	aaaat	caaat	cttggg	accc	caaaa	actcac	25200
tggaac	ag	taag	cttaga	aacgg	agtca	ctccaaa	acc	tg	cctt	gtt	tttg	25260
agcagat	att	tg	caagaca	aaagg	caaca	tg	ttcccc	ag	gtag	cctc	ct	25320
ctcaca	agga	aactc	ctt	gt	gggtccc	ag	ag	ttttt	acc	ctgag	gcaga	25380
tttcatt	gtg	acaat	gcaaa	ttag	caact	atctt	tactg	gtacat	gata	aataa	caaga	25440
ctctg	ccca	ccctc	agaaa	aatgc	acata	tg	cctg	cttc	ctctact	cta	tg	25500
tatctt	atat	aaaat	gtaga	cgaact	gtg	atgag	acgaa	tgaata	aattg	actgt	tcctc	25560
tacccc	ctcc	tttcat	agac	aacct	gtgga	ttca	aggagg	gcta	aatcaaa	gcttc	acaag	25620
aatgtg	acca	ctcac	ctcat	tacct	atcct	gtctt	tttt	ctt	ctc	ctc	ctg	25680
cctgat	ttta	cacatt	ttaa	tactg	gaggc	ctcaaa	accc	tctgt	gaaa	acgag	tggac	25740
cccaga	acct	cctgt	gactt	gtgtg	actt	ttc	ctgg	gtg	cate	ctcaac	cttgg	25800
taaa	catcta	aactg	actga	gacgt	gtccc	agacc	ctct	tggtt	tacaa	gtctt	gctc	25860
ctcc	ctgag	tgtg	gatga	aagca	acgtc	ctct	cttct	gcctc	ag	ttt	cctcat	25920

-continued

aaaatcagat cataatcctg gccattgag gggatgatta taaaactcaa tcaaatggag	25980
aagcatttgc aaatggaatg tattgcaaa ggtattgtgc agtccctgtg taatcaccat	26040
tctcttcctg cctcataagc tgggtgtgtg gtgtgtatgg gtgggggtgag gtggacacga	26100
tgcccccaga gatgtttgct caggatgaaa taggccccca tggaaggggtc aggttggcct	26160
ctgtggttta gctgcctttg aggcccttct gcctggggca aagcacaaaa gtccaatagc	26220
aatctgggtg catgacttcc tagtcatgga agtgtggctg agcagtggtc tctagtccct	26280
ccacctgtg tgctcggga gtggaacacg gggcagtggtc ctctgctggg gaggatgctg	26340
gggatctct gaaaggaggc aatgcctgat tttgtcactg aacgatgagc atgatttttc	26400
caggcgggtg aagcaggtgc tcccggcaga agaccccata taagcaaata tctagaagcc	26460
acaaggcatg gccttcaaca agagcaggcc tgggtcgggtg gctcatgcca agtggggagg	26520
atcacttgag gcctggagtt caagagtagc ccagtcaata tagtgaggca ccactctctaa	26580
tataattttt ttaaaaaatta gctgggtggt tggctgatac ctgtagaacc agctactcag	26640
gaggctgagg tggggggatg gcttaagccc aggggtttga agctgctgtg cgcgatgact	26700
gtggcactgc aattcagtct tgggtgacaga ctgagaccct gtctcaaaaa cataaccaaa	26760
ataaccccaa cggcattaca ataaatatac aataagcatc cacaaggcat gtgtgggaga	26820
ctgtgcttat gccgatggca cgggcaggag taaggcaggc atcaggggaa tgtggatgca	26880
cgggaggatg gggagtaatc tgcagggtcc catagtatgt cagtggcagg tctttctcct	26940
tgagaccaca gcagaccccc agccctgagg atgcgaggca ggtgggttgg atgagaggga	27000
tctggatgtc tgggtctcagg ctgctcctct aagggcagca gcaaggaggg tggggctgga	27060
ctgaagggtg aggggagcgt gtttctgtgt cttgggtcatg agctgctggg aagttgtgac	27120
tttactttc ctttctgaat tcctcggtat atcttgggga ctggaggacc tgtctgttta	27180
ttatacagac gcataactgg aggtgggac caccacagctc agaacagctg gatcttgctc	27240
agtctctgcc aggggaagat tccttggtta gttggggaca gttgacctgc ctccatctta	27300
ttctcacagc tttttagatg aagctgggag aggttaggtg actggttcag atagacagga	27360
aagtagcggc aaagcctgga tttgcaccca ggcttgatc tgattgcttc cctctcctg	27420
ctgggctcct aaactataga ggggtgggga aacagctcag atgggttcac aaagctccct	27480
gggctcagcc ctggttttga ctgctgagac acgactgagt gaaccagaa ggaaggtttg	27540
tgttgggggt gccctctct ttgggaaatt tgaaggaaacc aaagaagggtg gaggagagga	27600
ggtgaccag atccagcct gttcttatgg ccaggccaca ccaaggctgg cccaggggtg	27660
ctgctgctga ggcccaatgg gggcaacaga attttttttc aggtacaagg agttcagagg	27720
tcactctacc caccctgagt ctgagacca acttggtgcca ggcctgtgg ggaactgag	27780
gccctgggag cggaatgact tgcccagggt cccgtgtcag gaggtgagc aggtagatca	27840
taggactcca cgtctctggc cctcaacct gtccctctgt tcagacttct ggggtgatgg	27900
agaagaaca ggctgtgctg tgtccctaact gggaaacgtg gctgagacag gggagtgaga	27960
aggggtgcgt gcagaatggt gcctgtggca tgatgccagc tttgcaatca tgagattcaa	28020
aagccacact gtggaattgt gagtataact acaggagtga gagcttagat ctctgtgttc	28080
atagaaggag aatcaaacct ggaaactttt ggattattaa aatcaaacat tttgtcttca	28140
atattagagt cacaagggca gatgttgcc tccctctgc cctccaagag ctgtgtgact	28200

-continued

ctgggaaatt	cacctccct	cccagcctga	gtattttccc	ctgtaaaaa	cctgatatag	28260
ctgccaaaag	tccatgaaat	caggggggta	atttctgtga	aatctgagtc	atggctcttat	28320
ttgttcttca	ccaaaaagag	atgatgaagc	ctgtaacccc	agtgtgtgtg	gaggcagagg	28380
cgggaggatg	acttgaggcc	aggagtcca	gacagcctgg	gaaacacagg	agatgctgtc	28440
tctatgaaat	caaaaattaa	caaattagcc	cagcatggta	gtgtgtacct	gtagtcccag	28500
ctactcagaa	ggttgaagga	ggcggatcac	ttgagcccaa	cagatctagg	ctgcaatgag	28560
ctaggatggt	gtcaaaagccc	tgcagcctgg	gtgacagagc	aaaactgttt	attaaaaact	28620
aaacgaaagg	cggggcgccg	tggctcacgc	ctgtaatccc	agcacttttg	gaggccaaga	28680
cagggtggtca	cgaggtcagg	atctcgagac	catcctggct	aacatggtga	aacctgtct	28740
ctactaaaa	tacaaaaaaa	ttagccgggt	gtcgtggcgg	gcacctgtag	tcccagctac	28800
tcgggaggct	gaggcaggag	aatggcgtga	acccgggagg	cagagcttgt	agcagacct	28860
gatcgcccca	ctgcactcca	gcctgggtga	tagagcgaga	ctccatttca	aaaaaaaaat	28920
gaaatgaaaa	tgggtgatttc	tcaggaggag	cacactgtct	caaccctct	tttctgtct	28980
aaggaggagg	ccctgcagcg	acatggaggg	agctgctttg	ctgagagtct	ctgtcctctg	29040
catctggtga	gcttggtcca	gagccctgag	gcgggaggga	gggctccctg	tctgggctgc	29100
acaattggac	tgggcttggg	ctgggccagg	gccatctggg	cttcttctag	gaaccaaagt	29160
cacttcccgg	aattgacca	ccggcagact	cgcccaggga	ggaccctctg	tatcctagtt	29220
gaggctaacg	gtcaatatcc	ggcctcaata	ttcagcagag	gctccaggtc	agggctctgag	29280
ccaggccaac	aatgaccaag	gaggatggga	tccagggtgc	agctcctcac	aagtgtcggg	29340
ggaatccgag	gcaccagctc	tctctaccct	cctgtctctc	tgtctctctc	tggctcctct	29400
gtgtcctcat	ggctcatgag	atggtgttga	gatggctccc	ctggcctagg	ggatcacaca	29460
gtgaacaaa	caggaagaag	gcgaagggat	gcccctctcc	tgtcaccctt	gtcccttaca	29520
gcaagaaggg	cccagacgcc	cctctgcata	ctcccctggg	gaactgctgc	ccaggactgg	29580
gtcccccttt	tacccttgct	gcatggagtc	cccagaagac	aaacatctgt	gtgtctgaac	29640
cctgagacaa	aggcaggaaa	gggaaagagg	gaggcgagtg	gcttttgagg	agggggcttt	29700
agtatgagag	ctggaggatg	gaaccccatc	agggggcccg	ggaaccactg	agctgttaaa	29760
ataaagtctg	caaacaaaga	ccagctgctg	gaagtgggtg	tgcaggagg	tgcgcagaga	29820
cacacggtga	gaaaagaaca	atggtaatgc	ttggagccgc	ccctaactgg	gatgggcctg	29880
aaatggtatt	gttattattt	atagtatcat	tattagtcac	tttcatctta	tttgtaccct	29940
ccctctatct	ctcctctcca	ccttttccta	acattctatc	accagtttta	tgtcttccat	30000
tagcaacttt	gtagctgtaa	ataatttact	tacaacttcc	atataccctc	agttgtaccc	30060
agtatttctt	aacttccctc	tttaaaaaaa	tgataaatatt	aatcctcctt	ctccttcggt	30120
cagtgtctca	catcccaact	gtacactttg	tgttttaaat	tttgtactaa	ttggtgttga	30180
taacaagtat	atttagttct	catatttata	tgttgtgtat	atgttttggt	ttttttttgt	30240
cttaaaacag	agactactga	gagccagtaa	gactgaaaga	aagggcatct	tcatatactg	30300
ctagtttgag	tgtgaattgg	tacaatttgc	ctgcagtgtg	gctctaggta	taaagccttt	30360
tgagttgttg	atagcctctg	acctgtcat	tctactccta	caaatatctt	aaggaaataa	30420
tctcaaatgt	gggcaaagtt	ttatgcgaag	aaggatgttc	ttttagcat	aatttatatg	30480

-continued

aggaattaga	aacacctga	ggtccatcag	ctagggaaaag	acagcagagg	ggacagggca	30540
gcaccaggag	ggcttggcag	ccacagcagg	ggctgggctg	gcatctggca	agatccagg	30600
ctttgtggga	caaagaatct	tagagaagca	acggggacac	aggctgtcca	gtgcagcacc	30660
agtgaattca	ccaagatgaa	gagttgagca	cagagtgtga	gacaaggatt	tgcaatgagg	30720
agtcatttac	catgagcgtg	tatacatgac	tcaatattgt	atatttaagg	gagtagagca	30780
cagtcattga	aaacatggg	ttgaagcaga	agtcctggct	ctagcactat	tgtaaaaaca	30840
ctggcaagtt	ccctaacttt	tgtggattcc	tctgtgtttg	aatgggaaaag	tggtacccat	30900
ctcctggcct	gtgtgtgatc	atcgaatgag	acacatgtaa	gaggctctgc	acatcgcaga	30960
catgcaatac	acactggcta	tcaccagcta	gttatcacac	atgctcccag	gcctcgggtca	31020
ttgtcagaac	cttcctcccc	atctctatct	ctgtgtatag	actctgtagt	ttctgtaatg	31080
atcagatggc	tgcccgtcct	ctgattatct	tctcctcata	ccccaacagg	atgagtgcac	31140
ttttccttgg	tgtgggagtg	agggcagagg	aagctggagc	gaggtgagtg	tctgcaaata	31200
gcagatgatg	ggggctcagt	gatagacaaa	caatctggtt	taggttggtc	ttggtgtttg	31260
cttcaccccc	agagagattt	gcagagcccc	ggctgctcaa	gcaagcctcc	ctctgtccac	31320
tggtggctca	catttgatcc	cacaatttgt	tttctcctcc	tcaaggggtgc	aacaaaacgt	31380
tccaagtggg	acagatactg	gagatcctca	aagtaagccc	ctcggtgact	gggctgctgg	31440
caccatggac	ccaggtaggc	tcacctctcc	ttccctgctg	gttcctactc	gcacttagaa	31500
tggaggggtg	cacctccaca	gctgagtgg	cccagaaga	accggatgg	actaggaggc	31560
caatgagtct	gccccaaagca	gcagaagagg	tcacgtggag	tccccgacc	caggggtctg	31620
ggggtcatct	gcatcctgac	ctctccttag	ggtgacacgg	cccagggtgc	cacttcctcc	31680
ctaccctggc	acctagcaaa	tgactcaagt	ggggcaggac	atccttgggg	aagtggaaac	31740
agcggggggg	ggggggagtg	tggggagagc	aggggtccagg	aggagaggag	ggctgggaga	31800
ggaggttgac	cagcagcccc	gacctgtcac	ctcccgctct	gctccaagct	gacctggggg	31860
tgtttccacc	tgaggctggt	cccctccctc	cagctcttgc	cctgccctag	cccctttggt	31920
ctgagtggcc	ctgtgggaat	gtatcgggcc	tgattcagct	tccagcttaa	ctggaccccg	31980
gtctgtcctt	tgccacatcc	cctctgtgtg	acctggcgct	ccttacctcc	cttctcctgt	32040
accatgagaa	taatatttcc	ccatgttgtc	ttgaggatca	aatataacac	ggctacctgg	32100
cctggccagt	gcctgccaca	caatagggcc	tcctaagg	ctggttccca	tcccctctac	32160
tggtcccttg	ttgattccag	gtggaagcca	cccgtcccc	cccagctgt	gttagggcac	32220
ccgccctgcc	ctgcccatca	caagcacctg	cttaccagg	atgactgcct	caggcagggg	32280
ctgagttcag	tcattcaaaa	tacccccggt	caaagaaaaa	tctgaggccc	aatagaattt	32340
taaaagtgta	attgagtaag	aaaggattgc	tgaattggga	aacagcaaac	tggaggaggc	32400
ttcgtgctat	ggtgaaaaag	tgctggagac	aaggtttata	gggtgaatat	gaaactgcag	32460
taaggacatt	atttgattgg	ttgcacttac	aaaactgcct	gtttgggttt	acctgtgttg	32520
aaaaatccctg	gtcacgttgg	tgccagtgtg	agtttaggga	ggctccgaa	tacaggtgag	32580
accggaactt	agccttgccc	aggtttttga	tgctgaagag	agaaaaaagt	caaagcaaaa	32640
gcaaatcagc	ttttagtgc	aaacagaagt	gcaggctcag	gggtgggagt	gcaggcggac	32700
tcagaagagc	cagtcggggc	cagtggaatt	tggggttctt	gtctttgctg	gtttattaaa	32760

-continued

ttaaggtgtg	gagtaagcat	aaggatttct	gggaagaggt	ggggatttct	tggaattggg	32820
agggcaatgc	agttttgtac	caaacatggg	catacatgga	acggctgtgg	tgccgggtggg	32880
tgtgggggtt	agtgaagcaa	ggagggtatg	attgagtgtg	aggcaaaacg	tggggtcagac	32940
ccagcgccga	gttgatcttc	actggtctta	gctgaattgg	ccacttcttg	ttgttcagga	33000
tcctgtgggc	ccatcccctc	cccctgtccc	tacaggtggg	ttcagcaact	ccttctgtgt	33060
ctgtcatgtg	aacctgctgc	ctggagttct	tgtttcttct	gtgacctccc	agtatccaca	33120
tctcagcatg	gcccctctgtc	agctggcttt	aagttttgtt	tgtatctaac	agaagtaccc	33180
atctgatatg	accccagttg	agtttcccat	agcttgacgt	ttcagggggg	tacagctgtt	33240
ctaaaggcct	cgtaggtatt	gcctgtccag	gaatgtttca	gggctgggtt	ttgttttaat	33300
tttgctttta	tatccaccag	ccccaccaca	tcaagaaata	cacagattta	gtgccagagg	33360
ttccttccat	ccaaacctgg	gtttccttcc	tccttttggg	gaagaaggaa	aaagaaaaaa	33420
caagacaatc	gctcgcccag	ccaagtacgc	attgcgttca	tttccctaag	ccccatttcc	33480
tcaggagggc	ttcctcccct	tacagagttg	ccacggcaac	caagtcagct	ctgctcagga	33540
gccgcccctg	acagcaaaact	gggctgtgct	gagtcctgtc	caagacgagg	ggagggcagg	33600
gaagattaga	agccgagagc	actcagccga	ccccggaatc	cttacgaagg	gctgggctgg	33660
gggactgagg	aaaaactctc	cccccaaaac	aagatgatct	gtggtttaat	aacaggccca	33720
gctgggtcca	gaggtgacag	tggagagccg	tgtaccctga	gaccagcctg	cagaggacag	33780
aggcaacatg	gaggtgcctc	aaggatcagt	gctgagggtc	ccgcccccat	gccccgtcga	33840
agaacccccct	ccactgcccc	tctgagagtg	cccaagacca	gcaggaggaa	tctcctttgc	33900
atggtagttg	tactcaatgc	tgtggagttt	gagacattat	ttttaaaact	taaatagccc	33960
cacaggtaag	ctataaggga	gaatgcagag	accacaggca	ttttcaaaac	ataaggagtt	34020
ttaatgttaa	ggagtttaag	gagaatgtca	aagaacatt	ggccttgagg	tcaaatcctg	34080
agcttctctc	ttgatgggtt	tgggcgggtg	ctctcgctgt	ctggatttta	ctttcttcat	34140
ctgtgaatta	agaatgaggt	ggccgggcac	tgtggctcac	gcctgtaatc	ccaacatttg	34200
ggaggcccgag	gcgggcagat	cacaaggcca	ggagatcgag	acgatcctgt	ctaaccagat	34260
gaaacccctgt	ctctactaaa	aatacaaaaa	aaattaccag	gcctgggtggc	aggcgccctgt	34320
agtcccagct	actcaggagg	ctgaggcagg	agaatggcat	gaaccagga	ggcagagcct	34380
gcagtgagcc	aagatcgcac	cactgcactc	cagcctgggt	gagagagcaa	gactccatcc	34440
ccccaaaaaa	aagaatgagg	tctcttgtgt	gtgcacggca	agaacaatt	ttacatcatg	34500
cctcagtgca	cacacaaata	catactggtc	cccagttcct	tatctgcaat	tccatnatte	34560
ccagatccct	gcaaaccaca	cgttttggtg	gcagctgact	tcaggagtaa	tctgtctctg	34620
tactgacttg	agttctttgt	agacctacca	ttatttttta	catttaattgt	gaatatttgc	34680
ttgtttcact	gttgaaacat	gtgtgggtgt	gattacaggg	agcagtcacg	cctctgctgg	34740
gaatggtcag	gaatatacca	tatgcacccc	gtgtcctttc	taaaatagtg	gaaattctga	34800
atgcctaaac	acatctggat	tccaggaatt	tacattaagg	gatatatatt	tgaataaacc	34860
cacaaaaaca	atactcacc	ctgctacgct	tggtacactc	tgatattttc	aaattgattt	34920
tcttttttaa	atgctgcttt	tgaaccaata	atttcataac	ccaatgatgt	gttgatattc	34980
acattttgaa	gacactggta	taaagtgttt	gacccaacag	gcatgtggca	gatggaagtg	35040

-continued

atcactggga	tgtttccact	tccttggcct	ggaaggctca	cttcatgttg	gtgctgagct	35100
gaaaaggatt	cattcctctc	caaagcacca	tggcactccc	ctggtgtttc	aatgttctca	35160
ggctgccata	gtaaaatata	acaggctggg	tgacttcaac	atctggcttt	tattttctca	35220
cagatccaga	ggcaagaagt	ccaaggtgaa	ggtgccaaaa	gggttgaggt	ctggggagga	35280
ctctcttgct	ggcttatgga	agctgccttt	gcactgtgtg	ctcacacagt	cttcttggtg	35340
ccctttctgt	tcttacaata	ggacactagc	cctctaatag	aattagagcc	ctacccttgt	35400
gatctcattc	agtcttaatt	gttctcctaa	aggccagatc	tccaaatgta	gtcacctgag	35460
tttaaagctt	cagcatatga	attttggggg	cacacaattc	cctctatagc	actaatgact	35520
ccataaaaca	agtcccatat	cacagctgtc	caggaaaatt	aacatgtagt	catctcttat	35580
gcaatggctg	ttatgcactc	ccacactttc	ctccaacctt	atcctttctt	ctttccaact	35640
agagagcagt	atctttattg	aggatgccat	taagtatttc	aaggaaaaag	tgagcacaca	35700
gaatctgcta	ctcctgctga	ctgataatga	ggcctggaac	ggattcgtgg	ctgctgctga	35760
actgcccagg	taagctccat	ggggttacct	ccattgggca	ctccggcgat	gccaccacgc	35820
tctccctggg	gctagttttc	cctgacaggc	acacctcttc	caggcagccc	cctctgctgg	35880
gcttgaggat	gaccttcttg	gcactccagg	aagaatattt	cctccatgtc	cttcgctggc	35940
agtgagttag	ctcaggctga	gggatagga	agcccacagg	aacaggggtc	ttctgtccat	36000
ctgctgggtg	tgggaacttt	gccagttact	ttctcacttt	cagcctttct	tcataaatgg	36060
agttttattg	tgagaaaaat	gagatcgtgc	atgtaaagtg	cttagcataa	tgactgacac	36120
acagtaggtg	cacagttaac	gttaacatct	tacaagcttg	gtggaaacag	aaccaggggt	36180
aaggatcttc	ttcttgaaag	tgggagccca	gcaggacttg	acctttggga	caagtcagaa	36240
aaatcctgac	ctaaccaatg	tgtgttcctg	gggatgctgg	ctggagccct	ttttctcatc	36300
tgaggaattc	cataaacacg	gcaggaccct	tctgctggtc	caggatgccg	agaggctgga	36360
gtggagtgtc	acagcgctgg	gtctctttcc	agctctgtct	gtgtccagtc	cctgggctct	36420
gggcttatgt	cagagactgg	aaactgtcat	cagtctaata	ggaatagcaa	ctcaatgtga	36480
taggagtgtg	gtgaaggcct	tagaggcaac	tccagggtcaa	agggtgggaa	ccaacaccaa	36540
gatcctgccc	aaaggccacg	agagaccggg	atcagcgta	gagcctggag	tcaggatgga	36600
gcaagagtca	gccttggttg	taggacaagc	aggaaggctg	gggcaccagg	gctgggacta	36660
gggtgtggaa	agagaggcac	caagggtctg	aaacttaagg	aggatcatgaa	cttgacagag	36720
acatgtgagt	gggagaggga	tatttcttcc	ctgtctggat	gatcccaaca	ctcatgggca	36780
tatagggtca	tggttgcctg	tgtctctcac	accacgcctt	ggaactggca	agttctgggc	36840
taccttttga	gaccttcttc	tcttctacac	agagaaaccc	ctgacattga	caagcagtat	36900
gtcttagtga	ccattttcat	agtagaatta	gttgctgcca	atttgtgtaa	tttcagggat	36960
actgagcaga	ctcacaaact	cccaaccaag	gaggaaaaca	cttctcttac	catcttgctc	37020
aaattcaggt	cttttcagca	ttagcaattt	atatagggtg	ttctgaaagt	gttttgaatc	37080
ttctagaggt	tccataaaat	tgcaaatggc	cttggtgtgc	acacaggtga	agctgtcact	37140
actaaggcat	cagttaggct	ggtgacacct	gcgagtgaga	cctggaaggt	tgggggtgca	37200
ggtggtaggt	caggaggcag	taaggtggct	ggagcgccca	tgtggagcca	gaatgaatag	37260
ccagatctcc	cactccagtc	agctggggca	gtggccagaa	ccagtagcct	tcagccaggt	37320

-continued

gaaagggcat	gaggcattgg	aagggagcag	atgccaggca	aagtgaactc	ctgtctctcc	37380
ggggttttgt	ttttgtttga	attcaaagta	ctcagagata	gccagactgc	aatgtctgag	37440
ggcacccctta	cctttgacac	cacctgcaaa	tttaagggga	tttctaaatc	catcctttga	37500
ttttgtaatt	cactggaagg	actcacataa	ctcactgaaa	actgctatth	cacaattgtg	37560
gttgatcaca	cggaaagaac	acagattaaa	atcagcctaa	gaaagagaca	cacggggcag	37620
agtttgggag	ggttttaagt	gtgaagcttc	attgtcctca	ggacgctgtg	ttctcctgtg	37680
tgaatgtggg	acaaaacaca	tgcggtgttg	ccaaccaggg	aagctcacc	agttaaggaa	37740
aaattcctgt	gaatcacatt	aaaccagaaa	tgaagcctth	tttcaggatt	gttggttacag	37800
gggaggggaga	ctgagcccag	ctccaaatat	agtaaagaca	gagggggagt	cacagccaac	37860
accaggggtg	gggggggaga	tggaacagga	ctgataggag	acaccagagt	gggggggttc	37920
tgctaaattg	gcctaacagg	attctcacta	aaggcaggtc	agacactcac	ccatcaatgg	37980
tgggggtgaa	gagcataatc	agatagcaaa	gcgtaatcag	acagcaaggg	gggaagaccc	38040
tccttaaaact	gacttgggga	gattcttgct	gagctgggtg	gtgcaggctc	agcaagcagt	38100
gggtggctgt	ggaaggccaa	ggtcggcgcc	tagtgagaga	gaggctcaga	agagcccagc	38160
tggagttggt	caaggagaaa	gtccttctth	ctcaaccat	gccttagtgt	ccagagtatc	38220
tgctggggct	ggatcacaa	ctgcctcat	ggctgacctg	gagtctccag	cccctccctg	38280
aggtctaacc	cctttagtcc	ccgggtcctc	tgggatcagg	aaagacacca	aatagcccca	38340
aaccccatca	tccatcagat	tcctatacgg	tcccgtggcc	aatcaccag	gcaaagacag	38400
atgttcttac	caggcaggac	attccagggg	cctggagatc	acttcccagg	agctgaggac	38460
aaagggaaag	cctctctttg	ggtaaagttc	atgcttctact	ttcttgccat	cctgggaata	38520
aggtctctth	gtctcactct	cactgctctt	tgttcactctg	cagagtthtc	cctctccaaa	38580
cactttggca	gtcacggtgt	tgacctgccc	tgaaatgagg	aggcatctth	atcaaacacc	38640
taccatgaaa	gagcatctga	gatgccttha	agatatttat	ttccatatgg	aataaatatt	38700
gaggagtatc	tatgatgaac	aaaatgtgth	tgagtttgaa	agataattaa	ttccatttgt	38760
aaaatttgtt	tctatttgat	cgatagtga	agctctgcaa	tttacacagg	atagagatca	38820
catggagctc	atttcacaga	cagggaact	gcatttctth	ttttttttth	tttttttgag	38880
ttggagtctt	gctctgtctc	ccaagctgga	gtgcagtggc	gcgatctcgg	ctcactgcaa	38940
gctctgcctc	ccgggtcaag	caattctcct	gcctcagcct	cctgagtagc	tgggactaca	39000
ggcgcccacc	accacgccc	gctaattth	tgaattthtg	gtagagacag	ggtttcaccg	39060
tattagtctg	gatggtctca	atctcctgac	cttggtgatcc	accgccttg	gcctcccaaa	39120
gtgctgggat	tacaggtgtg	agccaccaca	ccgagccaaa	actgcatttc	ttaatcctth	39180
aacctttcct	tgtgcaggaa	tgaggcagat	gagctccgta	aagctctgga	caaccttgca	39240
agacaaatga	tcatgaaaga	caaaaactgg	cacgataaag	gccagcagta	cagaaaactgg	39300
ttcttgaaag	agtttctctg	gttgaaaagt	gagcttgagg	ataacataag	aaggctccgt	39360
gcccttgctg	atgggggtta	gaagggtcac	aaaggcacca	ccatcgccaa	tgtgggtgtct	39420
ggctctctca	gcatttctct	tggcatcctg	accctcgtctg	gcattgggtct	ggcacccttc	39480
acagagggag	gcagccttgt	actcttgga	cctgggatgg	agttgggaat	cacagccgct	39540
ttgaccggga	ttaccagcag	taccatggac	tacggaaaaga	agtgggtggac	acaagcccaa	39600

-continued

gcccacgacc	tggtcatcaa	aagccttgac	aaattgaagg	aggtagggga	gtttttgggt	39660
gagaacatat	ccaactttct	ttccttagct	ggcaataactt	accaactcac	acgaggcatt	39720
gggaaggaca	tccgtgccct	cagacgagcc	agagccaatc	ttcagtcagt	accgcatgcc	39780
tcagcctcac	gcccccggtt	cactgagcca	atctcagctg	aaagcgggtga	acaggtggag	39840
agggttaatg	aaccagcat	cctggaaatg	agcagaggag	tcaagctcac	ggatgtggcc	39900
cctgtangct	tctttcttgt	gctggatgta	gtctacctcg	tgtacgaatc	aaagcactta	39960
catgaggggg	caaagtcaga	gacagctgag	gagctgaaga	aggtaggctca	ggagctggag	40020
gagaagctaa	acatnctcaa	caataattat	aagattctgc	aggcggacca	agaactgtga	40080
ccacagggca	gggcagccac	caggagagat	atgcctggca	ggggccagga	caaaatgcaa	40140
actttttttt	ttttctgaga	cagagtcttg	ctctgtcgcc	aagttggagt	gcaatgggtgc	40200
gatctcagct	cactgaagc	tctgcctccc	gtgttcaagc	gattctcctg	ccttgccctc	40260
ccaagtagct	gggactacag	gcgcctacca	ccatgcccag	ctaatttttg	tatttttaat	40320
agagatgggg	tttcaccatg	ttggccagga	tggtctcgat	ctcctgacct	cttgatctgc	40380
ccaccttgcc	ctcccaaagt	gctgggatta	caggcgtgag	ccatcgcttt	tgacccaaat	40440
gcaaacattt	tattaggggg	ataaagaggg	tgaggtaag	tttatggaac	tgagtgttag	40500
ggactttggc	atttccatag	ctgagcacag	caggggaggg	gttaatgcag	atggcagtgc	40560
agcaaggaga	aggcaggaac	attggagcct	gcaataaggg	aaaaatggga	actggagagt	40620
gtggggaatg	ggaagaagca	gtttacttta	gactaaagaa	tatatggggg	ggccgggtgt	40680
agtggtctcat	gcctgtaate	cgagcacttt	gggaggccaa	ggcgggcgga	tcacgaggtc	40740
aggagatcga	gaccatcctg	gctaacacag	tgaaccccg	tctctactaa	aaatacaaaa	40800
aattagccgg	gcatggtggc	gggcgcctgt	agttccagct	aactgggcgg	ctgaggcagg	40860
agaatggcgt	gaacctggga	ggtggagctt	gcagtgaacc	gagatatcgc	cactgcactc	40920
cagcctgggt	gacagagcga	gactccatct	caaaaaaaaa	aaaaaaaaaga	atatattgac	40980
ggaagaatag	agaggaggct	tgaaggaacc	agcaatgaga	aggccaggaa	aagaaagagc	41040
tgaaaatgga	gaaagcccaa	gagttagaac	agttggatac	aggagaagaa	acagcggctc	41100
cactacagac	ccagccccag	gttcaatgtc	ctccgaagaa	tgaagtcttt	ccctggtgat	41160
ggccccctgc	cctgtcttct	cagcatccac	tctcccttgt	cctcctgggg	gcanatctca	41220
gtcaggcagc	ggcttctctga	tgatgggtcat	tggggtggtt	gtcatgtgat	gggtccccctc	41280
caggttacta	aagggtgcat	gtccccctgt	tgaacactga	agggcagggtg	gtgggccatg	41340
gccatgggtcc	ccagctgagg	agcagggtgc	cctgagaacc	caaacttccc	agagagtatg	41400
tgagaaccaa	ccaatgaaaa	cagtcaccatc	gctcttacc	ggtaagtaaa	cagtcagaaa	41460
attagcatga	aagcagttta	gcattgggag	gaagctcaga	tctctagagc	tgtcttgtcg	41520
ccgccagga	ttgacctgtg	tgtaagtccc	aataaactca	cctactcacc	aagctggact	41580
tggtcgagtc	attcttttgt	ctctcagctc	tttccaagct	ttgggggcca	tcagtctcaa	41640
gtttttctcg	taatgggtgt	gatgggtggt	gtgatggcag	cagagtgggtg	ggggccttgg	41700
caagtggctg	gtgctcttcc	agaccagggg	acaccgatgg	accgatgaag	ttccagggga	41760
acttgaagtg	ctgtgttgga	aagaggttaa	tttgggaaga	aaggaaaaat	gaaaggagat	41820
gccaggaaca	ggtcaaggaa	ccttgagggtg	tgtccaggag	ttcaaccccg	cctgtggtca	41880

-continued

cgagacctcg cctctctgag cctcagtttc ctcacacagct gtagaagggt acctggacaa	41940
ggtgatgtct caggtccacc cagctctccc catcttgtgt cctggagact tgggtccaat	42000
cagcactgac tgatggctgt gcccttgttg tgccgatgga ggctcccctg ggctctgggt	42060
gcctgacttc ccttcctcat gatccttctt ccagggtctc gggaaacgag gcttccatgg	42120
ccccctctgc ttgctcactg gactctcact tccccagctg gtcatttccc tgagacctca	42180
aaacccacc aacccactg gggcaccac agaaagcttg tacatgtctg gctctggctg	42240
agcacagggt tggtttccat ctgcgagtca gagcccgga ttgacatggg acgctcgagg	42300
ctgactagt tccagagggc ctgtgtgcag attgggagaa gattatcgag taatagagg	42360
gcaaagagg acatccagac agcaaacacc cccatccagt ctgtcctcac gttttgcca	42420
gactcctgca agaacttcca gaatgggtctt ccatgtgcaa ctgccctcc ctgccatgt	42480
cctggatcat tctccacact gcgcagggtt gggctttaac aagatcatgt tttgggtttt	42540
tgggggtttt ttgtttgtt gtttggtctc actctgttgt gaggtggcg tcccgtggca	42600
tgagcatggc tcaactgcac ttggacctc cgccctcaa ctaccctcc acctgaccc	42660
cctgagtagc tgggaccatc cgtgtgtgcc acattgcca ggctggctc aaactcctga	42720
gttcagggtga tctgttcgcc ttggcattcc aatgggtgc aattacagac acgagccacc	42780
tgccccacct aagaacatag atttgatcga gcaccactt ctgaccattc aatggcttc	42840
cattactctt agcacaaga tgtaggccat gatgtgcct aaagtagcct gttgcggcct	42900
ggggctctgt gaaggaagag ccagaggcaa cgttctggtg cagagggttt ctaggaagt	42960
attctaggag cagaactgag gatccaggag ggaaggagga aacgttgata caagaagttt	43020
ttgaatgggc catccatag ggcaatgaga ccttctgagg tgccttcaag gagcgctcc	43080
agaatctttt tttccaggga ggagacatgg aggcctttat ccaactgctt agctgtgtcc	43140
caggggttct ggtacattgt gtctttgttc tcattggtt caaagaactt cttgatttct	43200
gccttaattt tgtattttac ccagtaatca ttcaggaaca cgttgttcaa tttccatgta	43260
gttgtgcggt ttttagtgag tttcttaac ctgagttcta atttgattgc actgtggtct	43320
gagacactgt ttgttatgat ttcattctt ttgcatttgc tgaggaatgt tttacttcca	43380
attatgttgt cgactttaga ataagtgtta tacggtgctg agaagaaagt atattctgtt	43440
gatttgggac ggagagtctt gtagatgtct gttagggtcca ctcagtccag ggctgagttc	43500
aagtctgaa catccttgtt aattttctgt ctggttgacc tgtctaatat tgacagtggg	43560
gtgttaaagt ctccactgc tattgtgttg gagtctaagt ctctttgtag gtctctaaga	43620
actgttttta tgaatctgcg tgctccgta ttgggtgcat atatatttag gatagttagc	43680
tcttcttgtt gcactgatgt ttgttggtt aaaagatctt ttttgatctt tgttggttta	43740
aagtctgttt tatcagagac taggattgca acccctgctt tttgtttgt ttgtttgttt	43800
gtttgtttt ctttccattt gcttggtaaa tgttctcca tccctttatt ttcagcctat	43860
gtgtgttttt gcacatgaga tgggtctcct gaatacagta catcaacggg tcttgacttt	43920
atccaatttg ccagtctgtc tgttttaatt ggggcattta gccatttac atttaagggt	43980
catattgtta tgtgtgaatt tgatcctgtc atcatgatgc tagctgatta ttttgacat	44040
tagttgatgc agtttcttca tggatcatt ggtctttata ttttggtgtg gttttgcagt	44100
gtctggtact ggtttttctt ttccatattt agtgcctcct tcaggagctc atgtaaggcc	44160

-continued

ggcctgggtgg	tgacaaaatc	cctcagcatt	tgcttgtctg	gaaaggattt	tatttctcct	44220
ttgcttatga	agcttagttt	ggctggatat	gaaattcttg	gttgaaaatt	ctttcttta	44280
agaatgttga	atattggtcc	ccactctctt	ccagcttata	tagggtttct	gcagagagat	44340
ccactgttag	tctgatgggc	ttcccttcgt	aggtaacgtg	atctttctct	ctggctgccc	44400
ttaacatttt	ttccttcgtt	tcaaccttgg	agaatctgat	gattatgtgt	cttgggggtg	44460
ctcttcttga	agagtatctt	agtgggtgtc	tctgtatttc	ctgaatttga	atgttggcct	44520
ggcttgcctg	ttggggaagt	tctcctggat	aatatcctaa	agtgtgtttt	ccagcttgtt	44580
tccgttctcc	ccatcacttt	caggtaacac	aatcaattgt	aggtttaatc	ttttcacata	44640
gtcccatatt	tcttggaggc	tttatttgtt	ccattgtcat	tcttttttct	ctaattttgt	44700
cttcacacat	tgtttcagta	agttgatctt	caatctttga	tattccgctt	gatcaatttg	44760
gctattgata	gttgtgtatc	cttcacgaag	ttctcatgct	gtgtttttca	gtcccatcag	44820
gtcatttaag	ttcttctcta	aactgggtat	tctagtttag	agctcctgta	acctttcatc	44880
aagggttcta	gttcccttgc	attgggttac	aacatgctcc	tttagctcag	aggagtctct	44940
cattacccat	cttctgaagc	ctacttctgt	caattcgtca	aattcattct	ccatccagtt	45000
ttgtgcctt	gctggagagg	agttgcaatc	atttgagaaa	gagacattct	ggtctttgga	45060
attttcagcg	tttttgtgct	ggtttttcct	tatcttcattg	aatttatctt	tgatctttga	45120
ggatgatgac	cttcggatgg	gcttttgggg	gaggngagg	ccttttagtt	gatcttgatg	45180
ttattgcttt	ctgtttgtta	gtttttcatc	taataatcag	gccccctctg	tgcaggctctg	45240
ctgcagtttg	ctggagggtc	actccagacc	ttttttgcct	gggtatcacc	agcagaggct	45300
acagaacagc	aaagatagct	gcctgctcct	tcctctggaa	gcttcgttcc	agaggggcac	45360
tgacttgata	tcagccagag	ctctcctgta	tgagggtgtc	ccctcaggag	gcacgggagt	45420
cagggaccca	cttgaggagg	cagtctgtcc	cttagtagtg	cttgagcgtc	gtgctgggaa	45480
atgcgctgct	ctcttcagag	ccagcaggca	ggaacgttta	agtcactga	agttgcgccc	45540
acagctgccc	cttcccttag	gtgcctctc	tgtccgggga	gatgggagtt	ttatctataa	45600
gccccctgact	ggggctgctg	cctttcttct	agagatgccc	tgccagttag	gagggaatcta	45660
gagaggcagt	ctggccacag	ccgctttgct	gtgctgtggt	gagttccgcc	cagtccgaac	45720
tccccacct	ccttagcact	gtcagggaaa	agctacctac	tcaagcctca	gtaatggcgg	45780
atgccctct	ctctaccaag	ctctagtggc	ccagggtcaac	ttcaggctgc	tgtgctggag	45840
cgagaatttc	aagccagtgt	ttcttagctt	gctgggctcc	gtgggagtgg	gacctgctga	45900
gcaagaccac	ttggtccct	ggcttcagcc	ctctttccag	gggagtgaac	ggttctgtct	45960
tgtgtgggtt	ccaggcgcca	ctgggggtaca	aaaaaaaaat	tcctgcagct	agctcagtgt	46020
ctttcccaaa	cagccaccca	gttttgtgct	tgaaccacag	ggccctggtg	gtgtaggcac	46080
atgagggaat	ctcctggtct	gtggattgca	aaaactggga	aatgcatagt	atctggactg	46140
gatagcacag	tcctcatagg	tttcccttgg	ctgggggagg	gaggcccccg	gctccttgca	46200
cttccctgggt	gaggcaaac	cccacctga	ttctgctcgc	cctccgtggg	ctgtaccac	46260
tgcctaacca	gtcccaatga	gatgaactgg	gtacctcctg	tggaaatgca	gaaatcactc	46320
gccttctgtg	ttgatcttgc	tgggagctgc	agtctggagc	tgttcctatt	cagccatctt	46380
gccagatctc	ccctctatct	tctttttttt	tttttttttt	ttgagacaga	gtcttgcctc	46440

-continued

gtctcccagg	ctggagtgca	gtagcacaat	ctcggtcac	tgcaagctcc	gcctcccag	46500
ttcacgccat	tctcctgcct	cagcctcctg	agtagcggac	tacaggctcc	tgcttttttg	46560
ttttgttttt	tgtattttta	gtagagacag	ggtttcacca	tgtagccag	gatggctctg	46620
atctcatgac	ctcatgatcc	gcccgcctca	gcctcccata	gtgctgggat	tacaggcgtg	46680
agccaccacg	tctggccgat	ctcacccat	cttaagctat	ttccagaaac	agaatgacta	46740
ggtcagaggg	tattaacatc	gaaaaatgtt	tccataccta	gaaccacatc	gttatgcaaa	46800
taaattgtat	caatttacac	tcacaccagc	tttacaggag	gggtgttcatt	ccttcagtct	46860
taagaattct	cagtattctt	tctctttaaa	aagaaaaaca	gatctttaca	cggagaaaaag	46920
tttttcaagt	tgggcttg	attatcctgc	ctgcaaagat	gagggaaaag	agcgaaagtt	46980
gatcatgttt	aggcattact	ttaagaagac	aaaagagcat	ttttcaacca	ccccttctga	47040
gaccagacta	aaatcacagt	caaggcaatt	ataggtataa	tagcagctga	tatggcttgg	47100
ggctgggtcc	ctgccccaat	ctcacatcaa	actgtaatcc	ccagcattga	tggtgatgcc	47160
tgctggaagg	caactgggtt	atgggagtg	atttcccctt	tggtgctatt	ctcttgatag	47220
tgagttccct	tcaccttctg	ccatgattgg	aagttccctg	aggcctcccc	agaagctgat	47280
accaccatgc	ttcctgtaca	gcccgcagaa	ctgtgagcca	attaagcccc	ttttctttat	47340
aaattaccca	gtctcaggta	tttcttttca	gtagtgtgag	aatggactaa	tacagcaact	47400
aatgctaata	gtggcgaaca	atgtgtggac	tttactttgc	accactaact	gctctaagat	47460
gttagaaaaa	actcaccgta	ctctttcaag	aacctctgaa	gtgtgtatca	tttgtaccac	47520
cattaacaga	taaggaaact	gaaacagatg	ttcaatgact	tggccaaggt	cagaaatcta	47580
ctgatggcta	aagactcgg	tccgacccag	tcatatggat	tccagaactc	acacattgcc	47640
ttggtaccct	ctctgctgac	aaaaaacttg	aatgagacca	acaggagatc	agagatttca	47700
acagctat	agaagacaga	actgactgaa	gaaatgacaa	ccgaataaag	atgggtgaag	47760
gcagctgcga	tccacagcaa	tatggagagg	acatccagcc	aagaagaatc	acctgctagg	47820
cagagcctca	gagagaggag	aagttcacca	gccccatgac	ccatatagga	ttcaaggcag	47880
cctcgagtct	acaccaatga	aaaaatcaga	gagttcatga	ctcaatcaaa	aagcccaga	47940
aataatatcc	ttgtattctg	acacttttag	gggttcccat	ccaatcacc	cacaggtaag	48000
cctgccagt	gacttgtctt	cctctacat	aggaatggat	ctaggatcag	atatatgaga	48060
aagacaaaat	aaggaaaaga	gaccaagact	aacaaatgca	aatgagctcc	agaagaaatg	48120
aataattttg	aaaatggaca	ataattttga	aaaaatcaaa	ctcatgtcct	tggagagatt	48180
taaactcacg	ctggccgggt	gcagtgactc	acacctgtaa	tcccagcact	ttgggaggcc	48240
gaggcaggca	gatcacctga	agtcgggagt	ttgagaccag	ccttaccac	atggagaaac	48300
ccctgtctct	actaaaaata	caaaattagc	tgggcatggt	ggcgcatgcc	tgtaatccca	48360
gctactcggg	aggctgagac	aggagaatcg	cttaaacctg	ggaggcggag	gttacggtga	48420
cctgagttcg	tgccattgca	ctccagcctg	ggcaacaaga	gtgaaactct	gtttcaaaat	48480
aaaaataaat	tgtatgcttg	agaacaagat	gtactgaaaa	gaagggggaa	acaatagaac	48540
aaaaaaaaat	tcttggaat	taaatgattg	acaaaataaa	caactcaaga	gagggttgaa	48600
aaagtctaga	aaagcttcca	gaaaataaac	tagaaagatg	acaggagaaa	tggagctaaa	48660
gcggaattga	aatgttacca	tacatcatga	cttcctggaa	agctacttga	gaaagagcta	48720

-continued

cagaggggagt	gggaaggacg	tgcgtgtgtg	tgcacgtgtg	tgcacccatg	gtgagtgggt	48780	
ggggataacc	aagagagaaa	aagacatggg	atccaggaaa	caagcaactc	aaccagtgga	48840	
gaaggggaatc	caaggatcac	ggctgagcat	gcagcaggcc	taaagaacaa	ccagtccagg	48900	
atggaataag	aaggggaagg	ttctgtgggg	acttcacaga	gtacctcata	tgattcacaa	48960	
tttgagggaa	gctgaggaaa	tcatgaaggc	aaacagtgca	agggggggga	aaacaaaggc	49020	
aattagaaac	atcaggaaaa	actaaagctg	tgccacaaag	catgtgcagc	catattgtac	49080	
aactcaactc	tatagtaaac	agtattttaa	acatcacaaa	aatgtaaaaa	cgcttgtttg	49140	
atTTTTaact	ttgagcacia	aacagaacat	ttaaataatga	gaaccaaata	aacttgttgc	49200	
caatcagtag	caaaattaaa	gtacaatggt	agtgtctggg	aaggagggca	ctgaaaagga	49260	
agattaatat	cctcgtttta	taaaatgggg	agtcaagaaa	cactggccat	agtgaggggc	49320	
taagaaaaac	attgcaagta	aattgtttac	aaagtttcta	atggggatta	taaaactagc	49380	
gatataatat	tagaagaagg	tggggagggg	agttagaagt	gtaaagataa	cttaataata	49440	
ataaataaat	aataagtaat	gaataaaaa	attaataata	aaaatagtaa	taaaaaaatt	49500	
atatccta	at	aggaagcaaa	taaatagtga	tgtctaaatt	tcaaaatgga	gaaaaataaa	49560
aatgcaatag	cttatattta	ttgaccatct	gctttgcacc	aggaaaccttc	taaatgcttt		49620
gcaggcatta	acttattttt	agtcttcaca	acaactgtta	gaaagatact	attatcacct		49680
gcatgcttca	gatgaggaaa	ctgaagctca	gagaagttaa	gctagtgaca	gcacctaaact		49740
agtgcacccc	ggaattcaaa	ctcaggcttc	ctatctccag	agcatgcact	ctcactatct		49800
catgtcccag	cataagaaaa	ttattttaa	atatttattt	actccctaaa	gaaaaagcta		49860
gggcccggcg	tgggtggctca	cacctgta	at	cccagcactt	tgggaggctg	agggtgggtg	49920
atcacctgag	gtcaggaggt	caagaccagc	ctgggcaaca	tggtgaaacc	ccgtctctac		49980
taaaagtaca	aaaattagct	gggcacagt	gcattgtgcct	gtaatcccag	ttactcggga		50040
ggctgaggca	ggagaatcgc	ttaagccagg	gaggtggagg	ttgcagttag	ccaagatcgc		50100
gccactgcac	tccagcctgg	gtgacagagg	gagactccgt	ctcaaaaaaa	aaaaaaaaaa		50160
gaaaaaaagaa	aaagctaaaa	ttgggactaa	ttgctcttaa	tttgggtaag	tggagccgtc		50220
tgtactaatt	gattgattga	ttgggaagg	gtaagtagag	ccctctgtac	taattgattg		50280
gtcttaacca	tctgtgcac	tcacttatta	ctttattttt	ttaatttaag	aggacagaat		50340
tcagtttata	gtagttagta	actgctaacc	tcaatagcaa	gttaggttct	ggtgtcacct		50400
tccttgaagg	ttatggagca	atgtgaacat	gtgccagtga	gaaataagcc	tgctctgat		50460
atggaagtgc	tgggaaggga	agcatgtagt	ccctttaa	at	gatatggatc	ccacctctgg	50520
gaggcctgta	atcccagcac	ctgtaatccc	agcactttgg	gaggccaagg	cggatagagc		50580
acaaggctcag	gagttcgaga	ccggcctggc	caatatgggtg	aaacccgctc	tctactaaaa		50640
atacaaaaat	tagctgggca	tcggtggtgg	tgctgtagt	cccagctact	cgggaggctg		50700
aggcaggaga	attgcttgaa	cccgggaggc	agaggttgcc	gtgagccatg	attacaccac		50760
tgacccccag	cctggagaca	gagtgagact	ttgtctcaa	gaaaaaaaaa	gaaaagaagt		50820
gggggcctca	cacacagttt	cctggataac	aggaactatt	gtaagagatc	ccaccaaac		50880
acaaccatgc	agaatgaatc	acagccctgc	acaaaggcca	cctctatgag	gacatctgcc		50940
caacaactgt	ctgtttaacc	ttgagctaag	gtgacccttg	ttattaattc	ttgtacctag		51000

-continued

ggatcattat ttaaaataac ttacataatt ctccccattt tggetttaaa aacctctgct 51060

<210> SEQ ID NO 4
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Single strand DNA oligonucleotide

<400> SEQUENCE: 4

acaagcccaa gccacgacc 20

<210> SEQ ID NO 5
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Single strand DNA oligonucleotide

<400> SEQUENCE: 5

cctggcccct gccaggcata 20

<210> SEQ ID NO 6
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: N can be A or G

<400> SEQUENCE: 6

tcaagctcac ggatgtggcc cctgtangct tctttcttgt gctggatgta gt 52

<210> SEQ ID NO 7
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: N can be G or T

<400> SEQUENCE: 7

caggagctgg aggagaagct aaacatnctc aacaataatt ataagattct gc 52

<210> SEQ ID NO 8
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: N can be A or G

<400> SEQUENCE: 8

ccccagttcc ttatctgcaa ttccatnatt ccagatccc tgcaaaccac ac 52

<210> SEQ ID NO 9
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27)..(27)

-continued

<223> OTHER INFORMATION: N can be A or G

<400> SEQUENCE: 9

accttcggat gggcttttgg gggagngag gtccttttag ttgatcttga tg

52

<210> SEQ ID NO 10

<211> LENGTH: 52

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (27) .. (27)

<223> OTHER INFORMATION: N can be C or T

<400> SEQUENCE: 10

cactctccct tgctctctg ggggcanatc tcagtcaggc agcggcttcc tg

52

What is claimed is:

1. A method of determining predisposition of a subject to develop a kidney disease, comprising:

identifying in a sample of the subject at least one APOL1 polypeptide variant which is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on said *Trypanosoma brucei rhodesiense* under identical assay conditions,

wherein presence of said APOL1 polypeptide variant indicates increased predisposition of the subject to develop the kidney disease,

thereby determining the predisposition of the subject to develop the kidney disease.

2. A method of determining predisposition of a subject to develop a kidney disease, comprising:

identifying in a sample of the subject at least one APOL1 nucleotide mutation in the APOL1 genomic sequence set forth in SEQ ID NO:3 or at least one APOL1 polypeptide variant, wherein said at least one nucleotide mutation or polypeptide variant being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1,

wherein presence of said APOL1 nucleotide mutation or said APOL1 polypeptide variant indicates increased predisposition of the subject to develop the kidney disease, thereby determining the predisposition of the subject to develop the kidney disease.

3. The method of claim 1, wherein said APOL1 polypeptide variant comprises a mutation in the C-terminal helix of said APOL1 polypeptide.

4. The method of claim 1, wherein said APOL1 polypeptide variant has a reduced binding ability to serum resistance-associated protein (SRA) expressed by *Trypanosoma brucei rhodesiense* as compared to the binding ability of said APOL1 wild type polypeptide under identical assay conditions.

5. The method of claim 1, wherein said APOL1 polypeptide variant comprises the amino acid sequence set forth in SEQ ID NO:1 with a mutation selected from the group consisting of G342, M384, N388-del and Y389-del.

6. The method of claim 2, wherein said LD between said APOL1 nucleotide mutation and said S342G mutation is characterized by a Lewontin correlation coefficient (D') higher than 0.5.

7. The method of claim 6, wherein a significance of said LD is characterized by $LOD \geq 2$.

8. The method of claim 2, wherein said APOL1 nucleotide mutation is selected from the group consisting of the guanine-containing allele of single nucleotide polymorphism (SNP) rs73885319 (SEQ ID NO:6), the guanine-containing allele of SNP rs60910145 (SEQ ID NO:7), the guanine-containing allele of SNP rs9622363 (SEQ ID NO:8), the adenine-containing allele of SNP rs60295735 (SEQ ID NO:9) and the cytosine-containing allele of SNP rs58384577 (SEQ ID NO:10).

9. The method of claim 1, wherein said kidney disease comprises end stage kidney disease.

10. The method of claim 1, wherein said kidney disease comprises HIV-associated nephropathy (HIVAN).

11. A method of designing a life style change to a subject with the risk of kidney disease, comprising:

(a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation,

(i) wherein said variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on said *Trypanosoma brucei rhodesiense* under identical assay conditions;

(ii) wherein said APOL1 nucleotide mutation is included in the APOL1 genomic sequence set forth in SEQ ID NO:3, and wherein said at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, wherein presence of said APOL1 polypeptide variant or said APOL1 nucleotide mutation indicates increased predisposition of the subject to develop the kidney disease, and;

(b) designing the life style change based on presence or absence of said APOL1 polypeptide variant or said APOL1 nucleotide mutation,

thereby designing life style change to a subject with the risk of kidney disease.

12. The method of claim 11, wherein said subject is infected with HIV.

13. A method of designing a nephrotoxic anti-retroviral treatment regimen to a subject infected with HIV, comprising:

(a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation,

(i) wherein said variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on said *Trypanosoma brucei rhodesiense* under identical assay conditions;

(ii) wherein said APOL1 nucleotide mutation is included in the APLO1 genomic sequence set forth in SEQ ID NO:3, and wherein said at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1,

wherein presence of said APOL1 polypeptide variant or said APOL1 nucleotide mutation indicates increased predisposition of the subject to develop the kidney disease, and;

(b) designing the nephrotoxic anti-retroviral treatment regimen based on presence or absence of said APOL1 polypeptide variant or said APOL1 nucleotide mutation, thereby designing a nephrotoxic anti-retroviral treatment regimen to the subject infected with HIV.

14. A method of determining if a subject is suitable for donating a kidney for transplantation, comprising:

(a) identifying in a sample of the subject at least one APOL1 polypeptide variant or at least one APOL1 nucleotide mutation,

(i) wherein said variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on said *Trypanosoma brucei rhodesiense* under identical assay conditions;

(ii) wherein said APOL1 nucleotide mutation is included in the APLO1 genomic sequence set forth in SEQ ID NO:3, and wherein said at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1,

wherein presence of said APOL1 polypeptide variant or said APOL1 nucleotide mutation indicates increased predisposition of the subject to develop a kidney disease, and;

wherein increased predisposition of the subject to develop said kidney disease indicates that the subject is not suitable for donating the kidney for transplantation,

thereby determining if the subject is suitable for donating a kidney for transplantation.

15. A kit for determining predisposition to a kidney disease, comprising a reagent capable of specifically detecting at least one APOL1 nucleotide mutation in the APLO1 genomic sequence set forth in SEQ ID NO:3, wherein said at least one nucleotide mutation being in linkage disequilibrium (LD) with the S342G mutation in the APOL1 polypeptide set forth in SEQ ID NO:1, and/or a reagent capable of specifically detecting at least one APOL1 polypeptide variant, wherein said APOL1 polypeptide variant is characterized by a higher trypanolytic activity on *Trypanosoma brucei rhodesiense* as compared to the trypanolytic activity of wild type APOL1 polypeptide as set forth in SEQ ID NO:1 on said *Trypanosoma brucei rhodesiense* under identical assay conditions.

16. The kit of claim 15, wherein said APOL1 nucleotide mutation is selected from the group consisting of the guanine-containing allele of single nucleotide polymorphism (SNP) rs73885319 (SEQ ID NO:6), the guanine-containing allele of SNP rs60910145 (SEQ ID NO:7), the guanine-containing allele of SNP rs9622363 (SEQ ID NO:8), the adenine-containing allele of SNP rs60295735 (SEQ ID NO:9) and the cytosine-containing allele of SNP rs58384577 (SEQ ID NO:10).

17. The kit of claim 15, wherein said reagent comprises a polynucleotide capable of specifically detecting said APOL1 nucleotide mutation.

18. The kit of claim 15, wherein said reagent comprises an antibody capable of specifically binding an APOL1 variant comprises said nucleotide mutation and not to said APLO1 wild type polypeptide.

19. The kit of claim 15, wherein said APOL1 polypeptide variant comprises the amino acid sequence set forth in SEQ ID NO:1 with a mutation selected from the group consisting of G342, M384, N388-del and Y389-del.

20. The method of claim 1, further comprising informing the subject on the state of the predisposition to develop the kidney disease.

21. The method of claim 2, further comprising informing the subject on the state of the predisposition to develop the kidney disease.

22. The method of claim 2, wherein said nucleotide mutation creates an APOL1 protein variant.

23. The method of claim 2, wherein said nucleotide mutation is detected by a DNA detection method.

24. The method of claim 22, wherein said APOL1 protein variant is detected by a protein detection method.

* * * * *