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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 March 2014 (20.03.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/043550 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2013/059778

(22) International Filing Date:
13 September 2013 (13.09.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/700,660 13 September 2012 (13.09.2012) US

(71) Applicant: **UNIVERSITY OF UTAH RESEARCH FOUNDATION** [US/US]; 615 Arapeen Drive, Suite 301, Salt Lake City, UT 84108 (US).

(72) Inventors: **HAGEMAN, Gregory**; University of Utah, 65 North Mario Capecchi Drive, Salt Lake City, UT 84132 (US). **PAPPAS, Christian, Matthew**; 932 Wilson Avenue, Salt Lake City, UT 84105 (US).

(74) Agents: **MARTY, Scott, D.** et al.; Ballard Spahr LLP, 999 Peachtree Street NE, Suite 1000, Atlanta, GA 30309 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

(54) Title: METHODS OF PREDICTING SURVIVAL IN A SUBJECT BASED ON THE PRESENCE OF SINGLE NUCLEOTIDE POLYMORPHISMS

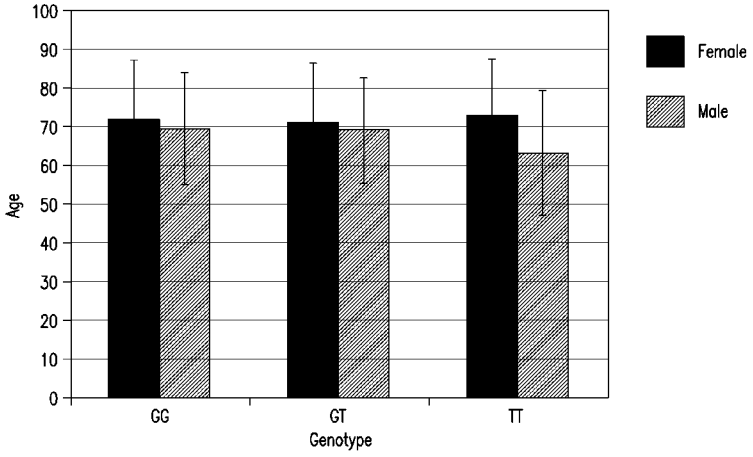


FIG. 1

(57) Abstract: Disclosed herein are methods and compositions for predicting survival in a subject.

WO 2014/043550 A1

METHODS OF PREDICTING SURVIVAL IN A SUBJECT BASED ON THE PRESENCE OF SINGLE NUCLEOTIDE POLYMORPHISMS

BACKGROUND

[0001] Single nucleotide polymorphisms (SNPs) located on chromosome 10 have been associated with a variety of diseases. To date, however, no association has been made between SNPs located on chromosome 10 and a subject's mortality risk. Therefore, what is needed are more effective methods of predicting a subject's mortality risk based on SNPs located on chromosome 10.

SUMMARY

[0002] Described herein are compositions and methods for predicting survival in a subject.

[0003] Also described herein are methods for predicting survival of a subject comprising determining in the subject the identity of the rs10490924 SNP in the ARMS2 gene, wherein the presence of the rs10490924 SNP is predictive of the subject's mortality risk.

BRIEF DESCRIPTION OF THE FIGURES

[0004] The accompanying figures, which are incorporated in and constitute a part of this specification, illustrate several aspects and together with the description serve to explain the principles of the invention.

[0005] Figure 1 shows the average age of death for males vs. females based on genotype at the rs10490924 SNP.

[0006] Figure 2 shows the average age at death for individuals who were homozygous for the disease risk allele at a given SNP compared to those who were homozygous for the non-risk allele.

[0007] Figure 3 shows the average age at death for males and females who were homozygous for the disease risk allele at the rs10490924 SNP compared to those who were homozygous for the non-risk allele at the rs10490924 SNP .

[0008] Figure 4 shows the average age at death for males and females who were homozygous for the disease risk allele at the rs10490924 and rs1410996 SNPs, compared to those who were heterozygous and homozygous at the non-risk allele of the rs10490924 and rs1410996 SNPs.

[0009] Additional advantages of the invention will be set forth in part in the description which follows, and in part will be obvious from the description, or can be learned by practice of the invention. The advantages of the invention will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

DESCRIPTION

[0010] The present invention can be understood more readily by reference to the following detailed description of the invention and the Examples included therein.

[0011] Unless otherwise expressly stated, it is in no way intended that any method or aspect set forth herein be construed as requiring that its steps be performed in a specific order. Accordingly, where a method claim does not specifically state in the claims or descriptions that the steps are to be limited to a specific order, it is in no way intended that an order be inferred, in any respect. This holds for any possible non-express basis for interpretation, including matters of logic with respect to arrangement of steps or operational flow, plain meaning derived from grammatical organization or punctuation, or the number or type of aspects described in the specification.

[0012] Definitions

[0013] As used in the specification and the appended claims, the singular forms “a,” “an” and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “a pharmaceutical carrier” includes mixtures of two or more such carriers, and the like.

[0014] Ranges can be expressed herein as from “about” one particular value, and/or to “about” another particular value. When such a range is expressed, another embodiment

includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint. It is also understood that there are a number of values described herein, and that each value is also herein disclosed as "about" that particular value in addition to the value itself. For example, if the value "10" is disclosed, then "about 10" is also disclosed. It is also understood that when a value is disclosed that "less than or equal to" the value, "greater than or equal to the value" and possible ranges between values are also disclosed, as appropriately understood by the skilled artisan. For example, if the value "10" is disclosed the "less than or equal to 10" as well as "greater than or equal to 10" is also disclosed. It is also understood that throughout the application, data are provided in a number of different formats, and that these data, represent endpoints, starting points, and ranges for any combination of the data points. For example, if a particular data point "10" and a particular data point 15 are disclosed, it is understood that greater than, greater than or equal to, less than, less than or equal to, and equal to 10 and 15 are considered disclosed as well as between 10 and 15. It is also understood that each unit between two particular units is also disclosed. For example, if 10 and 15 are disclosed, then 11, 12, 13, and 14 are also disclosed.

[0015] "Optional" or "optionally" means that the subsequently described event or circumstance may or may not occur and that the description includes instances where said event or circumstance occurs and instances where it does not.

[0016] The word "or" as used herein means any one member of a particular list and also includes any combination of members of that list.

[0017] "Mortality" as used herein means death, such that when describing a subject's mortality risk, what is meant is the subject's risk of death. More specifically, as used herein, when a subject has an "increased mortality risk," that means the subject is likely to die at an earlier age as compared to those subjects that do not have an increased mortality risk.

[0018] "ARMS2" as used herein means the age-related maculopathy susceptibility 2 (ARMS2) gene. ARMS2 can also mean any ARMS2 protein encoded by the ARMS2 gene. ARMS2's NCBI Gene identification number is 387715.

A. METHODS OF PREDICTING SURVIVAL IN A SUBJECT

[0019] The present invention provides methods for predicting the survival or the mortality risk of a subject. Specifically, the methods relate to determining identity of a SNP in the ARMS2 gene and correlating the identity of the SNP with mortality risk associated with the SNP in a population. In some embodiments, the SNP is rs10490924. In some embodiments the subject is a male.

[0020] It is understood and herein contemplated that the disclosed methods utilize tissue samples from the subject to extract nucleic acids and determine the identity of SNPs. Thus, in some aspects the disclosed methods further comprise obtaining a tissue sample from the subject. "Obtaining a tissue sample" or "obtain a tissue sample" means to collect a sample of tissue either from a party having previously harvested the tissue or harvesting directly from a subject. It is understood and herein contemplated that tissue samples obtained directly from the subject can be obtained by any means known in the art including invasive and non-invasive techniques. It is also understood that methods of measurement can be direct or indirect. Examples of methods of obtaining or measuring a tissue sample can include but are not limited to venipuncture, tissue biopsy, tissue lavage, aspiration, tissue swab, spinal tap, magnetic resonance imaging (MRI), Computed Tomography (CT) scan, Positron Emission Tomography (PET) scan, and X-ray (with and without contrast media). In a further aspect, the disclosed methods can comprise purifying nucleic acid from the tissue sample obtained from the subject. Thus, in one aspect the methods disclosed herein comprise extracting nucleic acid from the tissue sample.

[0021] In a further aspect, the present invention provides methods to determine the biological age, as opposed to chronological age, in order to provide a marker for correlating the biological age with disease susceptibility and mortality risk.

[0022] The method of the present invention is useful in a variety of applications. Mortality or survival estimates based on the presence of the SNPs described herein gives a basis for identifying individuals or groups of individuals for therapeutic intervention or medical screening. It is also useful in epidemiological studies for identifying environmental agents, for example dietary factors, pathological agents, chemical toxins associated with the presence of the SNPs described herein and mortality. Similarly, the method may be used to

identify genetic factors, genetic linkage markers, or genes affecting the presence of the SNPs described herein and age related diseases.

[0023] The present disclosure reveals the discovery that certain SNPs are indicative of a subject's mortality risk. Detecting these SNPs represents a previously unrecognized means of predicting a subject's survival, and more specifically, predicting a subject's likelihood of dying at an earlier age than would normally be expected.

[0024] In one aspect, the identity of a SNP in the ARMS2 gene can be determined by sequencing (including but not limited to next generation sequencing methods), amplification methods, hybridization assays (including but not limited to fluorescence in situ hybridization), microarrays, and amplification reactions (including but not limited to reverse transcriptase PCR (RT-PCR), real-time PCR, and real-time RT PCR). It is understood that in some instances it can be advantageous to use cDNA for the amplification process. Thus, in a further aspect the disclosed methods can further comprise steps of obtaining a tissue sample from the subject, extracting nucleic acid from the subject (for RNA extraction this can include the use of a protease such as proteinase K to prevent RNA degradation), and synthesizing complementary DNA (cDNA) from the extracted RNA. It is understood and herein contemplated that the synthesis of cDNA can utilize a reverse transcriptase and a primer pair that will specifically hybridize to the ARMS2 gene or a region of the ARMS2 gene flanking the rs10490924 SNP.

[0025] Numerous scanning technologies been developed over the years and can be used in the disclosed methods to detect a SNP including but not limited to denaturing high-performance liquid chromatography (dHPLC), base excision sequence scanning (BESS), single strand conformation polymorphism (SSCP), heteroduplex analysis (HA),conformation sensitive gel electrophoresis (CSGE), denaturing gradient gel electrophoresis (DGGE), constant denaturing gel electrophoresis (CDGE), temporal temperature gradient gel electrophoresis (TTGE), chemical cleavage of mismatches, cleavase to digest intrastrand structures and produce fragment length polymorphisms (CFLP), enzymatic cleavage of mismatches, template exchange extension reaction (TEER), and mismatch repair enzymes.

[0026] From a technical perspective High-throughput or Next Generation Sequencing (NGS) represents an attractive option for detecting the somatic mutations within a gene. Unlike PCR, microarrays, high-resolution melting and mass spectrometry, which all

indirectly infer sequence content, NGS directly ascertains the identity of each base and the order in which they fall within a gene. The newest platforms on the market have the capacity to cover an exonic region 10,000 times over, meaning the content of each base position in the sequence is measured thousands of different times. This high level of coverage ensures that the consensus sequence is extremely accurate and enables the detection of rare variants within a heterogeneous sample. For example, in a sample extracted from FFPE tissue, relevant mutations are only present at a frequency of 1% with the wild-type allele comprising the remainder. When this sample is sequenced at 10,000X coverage, then even the rare allele, comprising only 1% of the sample, is uniquely measured 100 times over. Thus, NGS can provide reliably accurate results with very high sensitivity, making it ideal for clinical diagnostic testing of FFPEs and other mixed samples.

[0027] Examples of Next Generation Sequencing techniques include, but are not limited to Massively Parallel Signature Sequencing (MPSS), Polony sequencing, pyrosequencing, Reversible dye-terminator sequencing, SOLiD sequencing, Ion semiconductor sequencing, DNA nanoball sequencing, Helioscope single molecule sequencing, Single molecule real time (SMRT) sequencing, Single molecule real time (RNAP) sequencing, and Nanopore DNA sequencing.

[0028] Described herein are a collection of polymorphisms and haplotypes comprised of variations in the ARMS2 gene. One or more of these polymorphisms and haplotypes are associated with mortality risk. Detection of these and other polymorphisms and sets of polymorphisms (e.g., haplotypes) can be useful in designing and performing diagnostic assays for mortality risk. Polymorphisms and sets of polymorphisms can be detected by analysis of nucleic acids, by analysis of polypeptides encoded by ARMS2 coding sequences (including polypeptides encoded by splice variants), by analysis of ARMS2 non-coding sequences, or by other means known in the art. Analysis of such polymorphisms and haplotypes can also be useful in designing prophylactic and therapeutic regimes to increase survival in a subject.

[0029] The term "polymorphism" refers to the occurrence of one or more genetically determined alternative sequences or alleles in a population. A "polymorphic site" is the locus at which sequence divergence occurs. Polymorphic sites have at least one allele. A diallelic polymorphism has two alleles. A triallelic polymorphism has three alleles. Diploid

organisms may be homozygous or heterozygous for allelic forms. A polymorphic site can be as small as one base pair. Examples of polymorphic sites include: restriction fragment length polymorphisms (RFLPs), variable number of tandem repeats (VNTRs), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, and simple sequence repeats. As used herein, reference to a "polymorphism" can encompass a set of polymorphisms (i.e., a haplotype).

[0030] A "single nucleotide polymorphism (SNP)" can occur at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site can be preceded by and followed by highly conserved sequences of the allele. A SNP usually arises due to substitution of one nucleotide for another at the polymorphic site. Replacement of one purine by another purine or one pyrimidine by another pyrimidine is called a transition. Replacement of a purine by a pyrimidine or vice versa is called a transversion. A synonymous SNP refers to a substitution of one nucleotide for another in the coding region that does not change the amino acid sequence of the encoded polypeptide. A non-synonymous SNP refers to a substitution of one nucleotide for another in the coding region that changes the amino acid sequence of the encoded polypeptide. A SNP may also arise from a deletion or an insertion of a nucleotide or nucleotides relative to a reference allele.

[0031] A "set" of polymorphisms means one or more polymorphism, e.g., at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, or more than 6 polymorphisms known, for example, in the ARMS2 gene.

[0032] As used herein, "haplotype" means a DNA sequence comprising one or more polymorphisms of interest contained on a subregion of a single chromosome of an individual. A haplotype can refer to a set of polymorphisms in a single gene, an intergenic sequence, or in larger sequences including both gene and intergenic sequences, e.g., a collection of genes, or of genes and intergenic sequences. For example, a haplotype can refer to a set of polymorphisms in the chromosome 10q26 locus, which includes gene sequence for ARMS2, and intergenic sequences (i.e., intervening intergenic sequences, upstream sequences, and downstream sequences that are in linkage disequilibrium with polymorphisms in the genic region). The term "haplotype" can refer to a set of single nucleotide polymorphisms (SNPs) found to be statistically associated with each other on a single chromosome. A haplotype can

also refer to a combination of polymorphisms (e.g., SNPs) and other genetic markers (e.g., an insertion or a deletion) found to be statistically associated with each other on a single chromosome. A "diplotype" is a haplotype pair. For example, a diplotype can comprise a protective and a risk haplotype, two protective haplotypes, two risk haplotypes, a neutral and a risk haplotype, a neutral and a protective haplotype, or two neutral haplotypes. In some circumstances, one haplotype may be dominant or one haplotype may be recessive.

[0033] ARMS2, also known as age-related maculopathy susceptibility 2 or ARMD8, is a gene encoding the ARMS2 protein. ARMS2's NCBI Gene identification number is 387715. By accessing the NCBI Gene page for ARMS2, one of skill in the art can readily obtain information such as the genomic location of ARMS2, a summary of the properties of the ARMS2 protein, information on cellular localization of ARMS2, information on homologs and variants (for example, splice variants) of ARMS2, as well as numerous ARMS2 reference sequences, such as the genomic sequence of ARMS2, the mRNA sequence of ARMS2 and the protein sequence of ARMS2. All of the information readily obtained from the ARMS2 NCBI Gene entry and the NCBI Gene No. set forth herein are hereby incorporated by reference in their entirety.

[0034] Unless otherwise noted, "ARMS2" as used herein, includes any ARMS2 gene, nucleic acid (DNA or RNA), or protein from any organism that retains at least one activity of ARMS2. In addition, "ARMS2" as used herein, includes any non-coding sequence present in or intergenic sequence present between the genomic DNA or RNA of the ARMS2 gene. When referring to ARMS2, this also includes any ARMS2 gene, nucleic acid (DNA or RNA), or protein from any organism that and can function as an ARMS2 nucleic acid or protein associated with mortality risk. For example, the nucleic acid or protein sequence can be from or in a cell of a subject.

[0035] As used herein, a "nucleic acid", "polynucleotide" or "oligonucleotide" is a polymeric form of nucleotides of any length, may be DNA or RNA, and may be single- or double-stranded. Nucleic acids may include promoters or other regulatory sequences. Oligonucleotides are usually prepared by synthetic means. Nucleic acids include segments of DNA, or their complements spanning or flanking any one of the polymorphic sites known in the ARMS2 gene. The segments are usually between 5 and 100 contiguous bases and often range from a lower limit of 5, 10, 15, 20, or 25 nucleotides to an upper limit of 10, 15, 20, 25,

30, 50, or 100 nucleotides (where the upper limit is greater than the lower limit). Nucleic acids between 5-10, 5-20, 10-20, 12-30, 15-30, 10-50, 20-50, or 20-100 bases are common. The polymorphic site can occur within any position of the segment. A reference to the sequence of one strand of a double-stranded nucleic acid defines the complementary sequence and except where otherwise clear from context, a reference to one strand of a nucleic acid also refers to its complement.

[0036] For certain applications, nucleic acid (e.g., RNA) molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the use of phosphorothioate or 2'-O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. Modified nucleic acids include peptide nucleic acids (PNAs) and nucleic acids with nontraditional bases such as inosine, queosine and wybutosine and acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

[0037] As used herein, "probes" are nucleic acids capable of binding in a base-specific manner to a complementary strand of nucleic acid. Such probes include nucleic acids and peptide nucleic acids (Nielsen et al., 1991). Hybridization may be performed under stringent conditions which are known in the art. For example, see Berger and Kimmel (1987) Methods In Enzymology, Vol. 152: Guide To Molecular Cloning Techniques, San Diego: Academic Press, Inc.; Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Vols. 1-3, Cold Spring Harbor Laboratory; Sambook (2001) 3rd Edition; Rychlik, W. and Rhoads, R. E., 1989, Nucl. Acids Res. 17, 8543; Mueller, P. R. et al. (1993) In: Current Protocols in Molecular Biology 15.5, Greene Publishing Associates, Inc. and John Wiley and Sons, New York; and Anderson and Young, Quantitative Filter Hybridization in Nucleic Acid Hybridization (1985)). As used herein, the term "probe" includes primers. Probes and primers are sometimes referred to as "oligonucleotides."

[0038] The term "primer" refers to a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions, in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. A primer sequence need not be exactly complementary to a template but must be sufficiently complementary to hybridize with a template. The term "primer site" refers to

the area of the target DNA to which a primer hybridizes. The term "primer pair" means a set of primers including a 5' upstream primer, which hybridizes to the 5' end of the DNA sequence to be amplified and a 3' downstream primer, which hybridizes to the complement of the 3' end of the sequence to be amplified. In one aspect, disclosed herein are primers that specifically hybridize to the ARMS2 gene. It is understood tha the disclosed primers can hybridize to an area flanking the SNP of interest (for example the rs10490924SNP) or can overlap the SNP and be able to specifically hybridize to a specific genotype of the rs10490924SNP.

[0039] Exemplary hybridization conditions for short probes and primers is about 5 to 12 degrees C, below the calculated Tm. Formulas for calculating Tm are known and include: Tm = 4˚C x (number of G's and C's in the primer) + 2˚C x (number of A's and T's in the primer) for oligos <14 bases and assumes a reaction is carried out in the presence of 50 mM monovalent cations. For longer oligos, the following formula can be used: Tm = 64.9˚C + 41˚C x (number of G's and C's in the primer-16.4)/N, where N is the length of the primer. Another commonly used formula takes into account the salt concentration of the reaction (Rychlik, supra, Sambrook, supra, Mueller, supra.): Tm = 81.5˚C + 16.6˚C x (log 10[Na+]+[K+]) + 0.41˚C x (% GC) - 675/N, where N is the number of nucleotides in the oligo. The aforementioned formulas provide a starting point for certain applications; however, the design of particular probes and primers may take into account additional or different factors. Methods for design of probes and primers for use in the methods of the invention are well known in the art.

[0040] As used herein, the terms "risk," "protective," and "neutral" are used to describe variations, SNPs, haplotypes, diplotypes, and proteins in a population encoded by genes characterized by such patterns of variations. A risk SNP is an allelic form of a gene, for example an ARMS2 gene, comprising at least one variant polymorphism associated with increased risk for developing a disease, or a symptom thereof, or associated with an increased mortality risk. The term "variant," when used in reference to an ARMS2 gene, refers to a nucleotide sequence in which the sequence differs from the sequence most prevalent in a population. The variant polymorphisms can be in the coding or non-coding portions of the gene. An example of a risk ARMS2 SNP is a T allele at the rs10490924 SNP in the ARMS2 gene. The risk SNP can be naturally occurring or can be synthesized by recombinant techniques. A protective or neutral SNP is an allelic form of a gene, herein an ARMS2 gene,

comprising at least one variant polymorphism associated with decreased risk of developing a disease, or a symptom thereof, or associated with no risk of increased mortality. For example, one protective or neutral ARMS2 SNP is a G allele at the rs10490924 SNP in the ARMS2 gene. The protective or neutral SNP can be naturally occurring or synthesized by recombinant techniques. Thus, the "protective" forms of ARMS2 can provide therapeutic benefit when administered to, for example, a subject with increased mortality risk, and thus can "protect" the subject from early death.

[0041] The term "variant" can also refer to a nucleotide sequence in which the sequence differs from the sequence most prevalent in a population, for example by one nucleotide, in the case of the SNPs described herein. For example, some variations or substitutions in the nucleotide sequence of the ARMS2 gene alter a codon so that a different amino acid is encoded resulting in a variant polypeptide. The term "variant," can also refer to a polypeptide in which the sequence differs from the sequence most prevalent in a population at a position that does not change the amino acid sequence of the encoded polypeptide (i.e., a conserved change). Variant polypeptides can be encoded by a risk haplotype, encoded by a protective haplotype, or can be encoded by a neutral haplotype. Variant ARMS2 polypeptides can be associated with risk, associated with protection, or can be neutral.

[0042] By "isolated nucleic acid" or "purified nucleic acid" is meant DNA that is free of the genes that, in the naturally-occurring genome of the organism from which the DNA of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, such as an autonomously replicating plasmid or virus; or incorporated into the genomic DNA of a prokaryote or eukaryote (e.g., a transgene); or which exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR, restriction endonuclease digestion, or chemical or in vitro synthesis). It also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence. The term "isolated nucleic acid" also refers to RNA, e.g., an mRNA molecule that is encoded by an isolated DNA molecule, or that is chemically synthesized, or that is separated or substantially free from at least some cellular components, for example, other types of RNA molecules or polypeptide molecules.

[0043] By "isolated polypeptide" or "purified polypeptide" is meant a polypeptide (or a fragment thereof) that is substantially free from the materials with which the polypeptide is normally associated in nature. The polypeptides of the invention, or fragments thereof, can be obtained, for example, by extraction from a natural source (for example, a mammalian cell), by expression of a recombinant nucleic acid encoding the polypeptide (for example, in a cell or in a cell-free translation system), or by chemically synthesizing the polypeptide. In addition, polypeptide fragments may be obtained by any of these methods, or by cleaving full length polypeptides.

[0044] Two amino acid sequences are considered to have "substantial identity" when they are at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, or at least about 99% identical. Percentage sequence identity is typically calculated by determining the optimal alignment between two sequences and comparing the two sequences. Optimal alignment of sequences may be conducted by inspection, or using the local homology algorithm of Smith and Waterman, 1981, Adv. Appl. Math. 2: 482, using the homology alignment algorithm of Needleman and Wunsch, 1970, J. Mol. Biol. 48: 443, using the search for similarity method of Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. U.S.A. 85: 2444, by computerized implementations of these algorithms (e.g., in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.) using default parameters for amino acid comparisons (e.g., for gap-scoring, etc.). It is sometimes desirable to describe sequence identity between two sequences in reference to a particular length or region (e.g., two sequences may be described as having at least 95% identity over a length of at least 500 base pairs). Usually the length will be at least about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 amino acids, or the full length of the reference protein. Two amino acid sequences can also be considered to have substantial identity if they differ by 1, 2, or 3 residues, or by from 2-20 residues, 2-10 residues, 3-20 residues, or 3-10 residues.

[0045] Exemplary polymorphic sites in the ARMS2 gene are described herein as examples and are not intended to be limiting. These polymorphic sites, or SNPs, can also be used in carrying out methods of the invention. Moreover, it will be appreciated that these ARMS2 polymorphisms are useful for linkage and association studies, genotyping clinical populations, correlation of genotype information to phenotype information, loss of

heterozygosity analysis, identification of the source of a cell sample and can also be useful to target potential therapeutics to cells.

[0046] It will be appreciated that additional polymorphic sites in the ARMS2 genes, which are not explicitly described herein, may further refine this analysis. A SNP analysis using non-synonymous polymorphisms in the ARMS2 gene can be useful to identify variant ARMS2 polypeptides. Other SNPs associated with increased mortality risk may encode a protein with the same sequence as a protein encoded by a neutral or protective SNP but contain an allele in a promoter or intron, for example, changes the level or site of ARMS2 expression. It will also be appreciated that a polymorphism in the ARMS2 gene may be linked to a variation in a neighboring gene. The variation in the neighboring gene may result in a change in expression or form of an encoded protein and have detrimental or protective effects in the carrier.

[0047] The methods and materials provided herein can be used to determine whether an ARMS2 nucleic acid of a subject (e.g., human) contains a polymorphism, such as a single nucleotide polymorphism (SNP). For example, methods and materials provided herein can be used to determine whether a subject has a variant SNP. Any method can be used to detect a polymorphism in an ARMS2 nucleic acid. For example, polymorphisms can be detected by sequencing exons, introns, or untranslated sequences, denaturing high performance liquid chromatography (DHPLC), allele-specific hybridization, allele-specific restriction digests, mutation specific polymerase chain reactions, single-stranded conformational polymorphism detection, and combinations of such methods.

[0048] Described herein are methods for predicting survival of a subject comprising determining in the subject the identity of a SNP in the ARMS2 gene, wherein the SNP is rs10490924, and wherein the presence of the rs10490924 SNP is predictive of the subject's mortality risk.

[0049] In one aspect, the subject is male. In another aspect, determining a TT genotype at the rs10490924 SNP indicates that a male subject has an increased mortality risk. In yet another aspect, determining a GG genotype at the rs10490924 SNP indicates that a male subject does not have an increased mortality risk. In still another aspect, determining a GT genotype at the rs10490924 SNP indicates that a male subject does not have an increased mortality risk.

[0050] In one aspect, the subject is female. In another aspect, determining a TT genotype at the rs10490924 SNP indicates that a female subject does not have an increased mortality risk. In yet another aspect, determining a GG genotype at the rs10490924 SNP indicates that a female subject does not have an increased mortality risk. In still another aspect, determining a GT genotype at the rs10490924 SNP indicates that a female subject does not have an increased mortality risk.

[0051] As used herein, the term "subject" means an individual. In one aspect, a subject is a mammal such as a human. In another aspect, the subject can be a Caucasian subject. In yet another aspect, the subject can be a male. In a further aspect a subject can be a non-human primate. Non-human primates include marmosets, monkeys, chimpanzees, gorillas, orangutans, and gibbons, to name a few. The term "subject" also includes domesticated animals, such as cats, dogs, etc., livestock (for example, cattle (cows), horses, pigs, sheep, goats, etc.), laboratory animals (for example, ferret, chinchilla, mouse, rabbit, rat, gerbil, guinea pig, etc.) and avian species (for example, chickens, turkeys, ducks, pheasants, pigeons, doves, parrots, cockatoos, geese, etc.). Subjects can also include, but are not limited to fish (for example, zebrafish, goldfish, tilapia, salmon, and trout), amphibians and reptiles. As used herein, a "subject" is the same as a "patient," and the terms can be used interchangeably.

[0052] In one aspect, the SNPs, haplotypes, or diplotypes described herein can be determined from a sample obtained from the subject. In another aspect, the subject's SNPs, haplotypes, or diplotypes can be determined by amplifying or sequencing a nucleic acid sample obtained from the subject. For example, and not to be limiting, described herein are methods of determining the genotype at the rs10490924 SNP in a subject comprising obtaining a nucleic acid sample from the subject. In a further aspect, determining the genotype at the rs10490924 SNP can further comprise amplifying or sequencing the nucleic acid sample obtained from the subject. In one aspect, any of the primers and probes described herein can be used in amplifying or sequencing the nucleic acid sample obtained from the subject.

[0053] In one aspect, predicting survival in a subject can comprise detecting a SNP, or multiple SNPs, that are in linkage disequilibrium with one or more of the ARMS2 polymorphisms described herein. SNPs in LD with one or more of the ARMS2 SNPs

described herein include, but are not limited to rs3750848, rs3750847, rs3793917, rs11200638, rs1049331, rs2284665, and rs932275. In another aspect, predicting survival in a subject can comprise detecting an insertion/deletion that is in LD with one or more of the ARMS2 SNPs described herein. An insertion/deletion that is in LD with one or more of the ARMS2 SNPs described herein includes, but is not limited to, an insertion/deletion beginning in the 3' UTR of ARMS2 (hg19, chr10:124216821).

[0054] As used herein, "linkage" describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome. Linkage can be measured by percent recombination between the two genes, alleles, loci or genetic markers. Typically, loci occurring within a 50 centimorgan (cM) distance of each other are linked. Linked markers may occur within the same gene or gene cluster. As used herein, "linkage disequilibrium" is the non-random association of alleles at two or more loci, not necessarily on the same chromosome. It is not the same as linkage, which describes the association of two or more loci on a chromosome with limited recombination between them. Linkage disequilibrium describes a situation in which some combinations of alleles or genetic markers occur more or less frequently in a population than would be expected from a random formation of haplotypes from alleles based on their frequencies. Non-random associations between polymorphisms at different loci are measured by the degree of linkage disequilibrium (LD). The level of linkage disequilibrium can be influenced by a number of factors including genetic linkage, the rate of recombination, the rate of mutation, random drift, non-random mating, and population structure. "Linkage disequilibrium" or "allelic association" thus means the non-random association of a particular allele or genetic marker with another specific allele or genetic marker more frequently than expected by chance for any particular allele frequency in the population. A marker in linkage disequilibrium with an informative marker, such as one of the ARMS2 SNPs described herein can be useful in predicting a subject's mortality risk. A SNP that is in linkage disequilibrium with a risk, protective, or otherwise informative SNP or genetic marker described herein can be referred to as a "proxy" or "surrogate" SNP. A proxy SNP may be in at least 50%, 60%, or 70% in linkage disequilibrium with risk, protective, or otherwise informative SNP or genetic marker described herein, and in one aspect is at least about 80%, 90%, and in another aspect 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or about 100% in LD with a risk, protective, or otherwise informative SNP or genetic marker described herein.

[0055] Publicly available databases such as the HapMap database and Haploview (Barrett, J. C. et al., Bioinformatics 21, 263 (2005)) may be used to calculate linkage disequilibrium between two SNPs. The frequency of identified alleles in disease versus control populations can be determined using the methods described herein. Statistical analyses can be employed to determine the significance of a non-random association between the two SNPs (e.g., Hardy-Weinberg Equilibrium, Genotype likelihood ratio (genotype p value), Chi Square analysis, Fishers Exact test). A statistically significant non-random association between the two SNPs indicates that they are in linkage disequilibrium and that one SNP can serve as a proxy for the second SNP.

[0056] The discovery that polymorphic sites in the ARMS2 gene are associated with survival has a number of specific applications including, but not limited to, screening individuals to ascertain mortality risk, and identification of new and optimal therapeutic approaches for individuals having an increased mortality risk. Without intending to be limited to a specific mechanism, polymorphisms in the ARMS2 gene can contribute to the survival of an individual in different ways. Polymorphisms that occur within the protein coding region of ARMS2 may contribute to a phenotype by affecting the protein structure and/or function. Polymorphisms that occur in the non-coding regions of ARMS2 may exert phenotypic effects indirectly via their influence on replication, transcription and/or translation. Certain polymorphisms in the ARMS2 gene may predispose an individual to a distinct mutation that is causally related to increased mortality risk. Alternatively, as noted above, a polymorphism in the ARMS2 gene may be linked to a variation in a neighboring gene (including, but not limited to, PLEKHA1). The variation in the neighboring gene may result in a change in expression or form of an encoded protein and have detrimental or protective effects in the carrier.

[0057] Polymorphisms can be detected in a target nucleic acid isolated from a subject. Typically genomic DNA is analyzed. For assay of genomic DNA, virtually any biological sample containing genomic DNA or RNA, e.g., nucleated cells, is suitable. For example, in the experiments described in herein, genomic DNA was obtained from subjects postmortem. Other suitable samples include, but are not limited to, saliva, cheek scrapings, biopsies of retina, kidney or liver or other organs or tissues; skin biopsies; amniotic fluid or CNS samples; and the like. In one aspect RNA or cDNA can be assayed. In one aspect, the assay can detect variant ARMS2 proteins. Methods for purification or partial purification of

nucleic acids or proteins from patient samples for use in diagnostic or other assays are known in the art.

[0058] The identity of bases occupying the polymorphic sites in the ARMS2 gene can be determined in an individual, e.g., in a patient being analyzed, using any of several methods known in the art. For example, and not to be limiting use of allele-specific probes, use of allele-specific primers, direct sequence analysis, denaturing gradient gel electrophoresis (DGGE) analysis, single-strand conformation polymorphism (SSCP) analysis, and denaturing high performance liquid chromatography (DHPLC) analysis. Other well-known methods to detect polymorphisms in DNA include use of: Molecular Beacons technology (see, e.g., Piatek et al., 1998; Nat. Biotechnol. 16:359-63; Tyagi, and Kramer, 1996, Nat. Biotechnology 14:303-308; and Tyagi, et al., 1998, Nat. Biotechnol. 16:49-53), Invader technology (see, e.g., Neri et al., 2000, Advances in Nucleic Acid and Protein Analysis 3826:117-125 and U.S. Pat. No. 6,706,471), nucleic acid sequence based amplification (Nasba) (Compton, 1991), Scorpion technology (Thelwell et al., 2000, Nuc. Acids Res, 28:3752-3761 and Solinas et al., 2001, "Duplex Scorpion primers in SNP analysis and FRET applications" Nuc. Acids Res, 29:20.), restriction fragment length polymorphism (RFLP) analysis, and the like. Additional methods will be apparent to one of skill in the art.

[0059] The design and use of allele-specific probes for analyzing polymorphisms are described by e.g., Saiki et al., 1986; Dattagupta, EP 235,726, Saiki, WO 89/11548. Briefly, allele-specific probes are designed to hybridize to a segment of target DNA from one individual but not to the corresponding segment from another individual, if the two segments represent different polymorphic forms. Hybridization conditions are chosen that are sufficiently stringent so that a given probe essentially hybridizes to only one of two alleles. Typically, allele-specific probes can be designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position of the probe.

[0060] Allele-specific probes can be used in pairs, one member of a pair designed to hybridize to the reference allele of a target sequence and the other member designed to hybridize to the variant allele. Several pairs of probes can be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target gene sequence.

[0061] The design and use of allele-specific primers for analyzing polymorphisms are described by, e.g., WO 93/22456 and Gibbs, 1989. Briefly, allele-specific primers are designed to hybridize to a site on target DNA overlapping a polymorphism and to prime DNA amplification according to standard PCR protocols only when the primer exhibits perfect complementarity to the particular allelic form. A single-base mismatch prevents DNA amplification and no detectable PCR product is formed. The method works best when the polymorphic site is at the extreme 3'-end of the primer, because this position is most destabilizing to elongation from the primer. Any of the primers and probes described herein can be used in the methods described herein.

[0062] In some embodiments, genomic DNA can be used to detect ARMS2 polymorphisms. Genomic DNA is typically extracted from a sample, such as a peripheral blood sample or a tissue sample. Standard methods can be used to extract genomic DNA from a sample, such as phenol extraction. In some cases, genomic DNA can be extracted using a commercially available kit (e.g., from Qiagen, Chatsworth, Calif.; Promega, Madison, Wis.; or Gentra Systems, Minneapolis, Minn.).

[0063] Other methods for detecting polymorphisms can involve amplifying a nucleic acid from a sample obtained from a subject (e.g., amplifying the segments of the ARMS2 gene of an individual using ARMS2-specific primers) and analyzing the amplified gene. This can be accomplished by standard polymerase chain reaction (PCR & RT-PCR) protocols or other methods known in the art. The amplifying can result in the generation of ARMS2 allele-specific oligonucleotides, which span the single nucleotide polymorphic sites in the ARMS2 gene. The ARMS2 specific primer sequences and ARMS2 allele-specific oligonucleotides can be derived from the coding (exons) or non-coding (promoter, 5' untranslated, introns or 3' untranslated) regions of the ARMS2 genes. In one aspect Genomic DNA from a subject can be isolated from peripheral blood leukocytes with QIAamp DNA Blood Maxi kits (Qiagen, Valencia, CA). DNA samples can be screened for SNPs in ARMS2 (A69S, rs10490924). Genotyping can be performed by TaqMan assays (Applied Biosystems, Foster City, California) using 10 ng of template DNA in a 5uL reaction. The thermal cycling conditions in a 384-well thermocycler (PTC-225, MJ Research) can consist of an initial hold at 95°C for 10 minutes, followed by 40 cycles of a 15-second 95°C denaturation step and a 1-minute 60°C annealing and extension step. Plates can be read in the 7900HT Fast Real-Time PCR System (Applied Biosystems).

[0064] Amplification products generated using PCR can be analyzed by the use of denaturing gradient gel electrophoresis (DGGE). Different alleles can be identified based on sequence-dependent melting properties and electrophoretic migration in solution. See Erlich, ed., PCR Technology, Principles and Applications for DNA Amplification, Chapter 7 (W.H. Freeman and Co, New York, 1992).

[0065] Alleles of target sequences can be differentiated using single-strand conformation polymorphism (SSCP) analysis. Different alleles can be identified based on sequence- and structure-dependent electrophoretic migration of single stranded PCR products (Orita et al., 1989). Amplified PCR products can be generated according to standard protocols and heated or otherwise denatured to form single stranded products, which may refold or form secondary structures that are partially dependent on base sequence.

[0066] Alleles of target sequences can be differentiated using denaturing high performance liquid chromatography (DHPLC) analysis. Different alleles can be identified based on base differences by alteration in chromatographic migration of single stranded PCR products (Frueh and Noyer-Weidner, 2003). Amplified PCR products can be generated according to standard protocols and heated or otherwise denatured to form single stranded products, which may refold or form secondary structures that are partially dependent on the base sequence.

[0067] Direct sequence analysis of polymorphisms can be accomplished using DNA sequencing procedures that are well-known in the art. See Sambrook et al., Molecular Cloning, A Laboratory Manual (2nd Ed., CSHP, New York 1989) and Zyskind et al., Recombinant DNA Laboratory Manual (Acad. Press, 1988). In one aspect sequence analysis can be accomplished using sequencing by synthesis.

[0068] A wide variety of other methods are known in the art for detecting polymorphisms in a biological sample. See, e.g., Ullman et al. "Methods for single nucleotide polymorphism detection" U.S. Pat. No. 6,632,606; Shi, 2002, "Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes" Am J Pharmacogenomics 2:197-205; Kwok et al., 2003, "Detection of single nucleotide polymorphisms" Curr Issues Biol. 5:43-60).

[0069] Described herein are nucleic acids, polymorphic sites, adjacent to or spanning the ARMS2 polymorphic sites. The nucleic acids can be used as probes or primers (including Invader, Molecular Beacon and other fluorescence resonance energy transfer (FRET) type probes) for detecting ARMS2 polymorphisms.

[0070] Also described herein are vectors comprising a nucleotide sequence that encodes a full length ARMS2 polypeptide. The vector can also comprise a nucleotide sequence that encodes a sub-domain of the ARMS2 polypeptide. Therefore, the ARMS2 polypeptides can comprise the amino acid sequence found at NCBI Protein accession number NP_001093137, or a variant thereof (e.g., a risk variant or a protective variant) and may be a full-length form or a truncated form. The nucleic acid may be DNA or RNA and may be single-stranded or double-stranded.

[0071] Some nucleic acids can encode full-length, variant forms of an ARMS2 polypeptide. The variant ARMS2 polypeptide can differ from NP_001093137.1 at an amino acid encoded by a codon including one of any non-synonymous polymorphic position known in the ARMS2 gene. In one aspect, the variant ARMS2 polypeptide differs from NP_001093137.1 at an amino acid encoded by a codon including one of the non-synonymous polymorphic positions described herein. It is understood that variant ARMS2 genes can be generated that encode variant ARMS2 polypeptides that have alternate amino acids at multiple polymorphic sites in the ARMS2 genes.

[0072] Expression vectors for production of recombinant proteins and peptides are well known in the art (see Ausubel et al., 2004, Current Protocols In Molecular Biology, Greene Publishing and Wiley-Interscience, New York). Such expression vectors can include the nucleic acid sequence encoding the ARMS2 polypeptide linked to regulatory elements, such as a promoter, which drives transcription of the DNA and is adapted for expression in prokaryotic (e.g., E. coli) and eukaryotic (e.g., yeast, insect or mammalian cells) hosts. A variant ARMS2 polypeptide can be expressed in an expression vector in which a variant ARMS2 gene is operably linked to a promoter. The promoter can be a eukaryotic promoter for expression in a mammalian cell. The transcription regulatory sequences can comprise a heterologous promoter and optionally an enhancer, which is recognized by the host cell. Commercially available expression vectors can be used. Expression vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences

useful for insertion into the host genome, and the like. For example, and not to be limiting, an ARMS2-A69S plasmid can be synthesized by PCR using oligonucleotides as template to produce human ARMS2 amino acid sequence using codons optimized for expression in *Escherichia coli* bacterial cells. The sequence can contain an A69S mutation. The PCR product can contain NdeI and XhoI restrictions sites at the ends which can be used to ligate the PCR product into the pET-21a(+) plasmid.

[0073] Also described herein are isolated host cells comprising a vector that encodes a full length ARMS2 polypeptide. The host cells can also comprise a vector that encodes a sub-domain of the ARMS2 polypeptide. Suitable host cells can include bacteria such as E. coli, yeast, filamentous fungi, insect cells, and mammalian cells, which are typically immortalized, including mouse, hamster, human, and monkey cell lines, and derivatives thereof. Host cells may be able to process the ARMS2 gene product to produce an appropriately processed, mature polypeptide. Such processing may include glycosylation, ubiquitination, disulfide bond formation, and the like.

[0074] Expression constructs containing an ARMS2 gene can be introduced into a host cell, depending upon the particular construction and the target host. Appropriate methods and host cells, both prokaryotic and eukaryotic, are well-known in the art. For example, full-length human ARMS2 can be expressed in NEB T7 Express lysY cells. Constructs expressing ARMS2 protein can be transformed and expressed using BL21-AI™ One Shot® Chemically Competent E. coli (Invitrogen #C607003). Constructs that produce a protein whose native fold contains disulfide bonds can be expressed using either NEB SHuffle Express competent E. coli (New England BioLab #C3028H) or NEB SHuffle T7 Express competent E. coli (New England BioLabs #C3026H) for constructs driving protein expression through the T7 promoter).

[0075] Also described herein are adenoviral vectors comprising a nucleotide sequence that encodes a full length ARMS2 polypeptide, including variants. In one aspect, described herein are infectious adenoviral particles that encode a full length or variant ARMS2 polypeptide. Further described herein are isolated host cells containing therein an adenoviral particle that encodes a full length or variant ARMS2 polypeptide. Suitable host cells can include mammalian cells, such as mouse, hamster, human, and monkey cell lines, and derivatives thereof. Host cells may be able to process the ARMS2 gene product to

produce an appropriately processed, mature polypeptide. Such processing may include glycosylation, ubiquitination, disulfide bond formation, and the like. The adenoviral construct can allow for the expression of ARMS2 polypeptide in mammalian systems. In one aspect, a mouse can be infected with the adenoviral particle and ARMS2 polypeptide or a variant thereof can be expressed in order to drive phenotypes associated with increased mortality risk, including death. In another aspect, described herein are transgenic animals comprising the TT genotype at the rs10490924 ARMS2 SNP.

[0076] Also described herein are purified or isolated proteins comprising the amino acid sequence of the full length ARMS2 polypeptide. In one aspect, the purified or isolated protein can comprise the amino acid sequence of an ARMS2 polypeptide sub-domain. A protein assay can be carried out to identify the proteins and to characterize polymorphisms in a subject's ARMS2 genes, such as the rs10490924 SNP. Methods that can be adapted for detection of the ARMS2 protein, and in particular the ARMS2-A69S protein, are well known. These methods include analytical biochemical methods such as electrophoresis (including capillary electrophoresis and two-dimensional electrophoresis), chromatographic methods such as high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, mass spectrometry, and various immunological methods such as fluid or gel precipitin reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmnunoassay (RIA), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, western blotting and others. For example, and not to be limiting, a number of well-established immunological binding assay formats suitable for the practice of the invention are known (see, e.g., Harlow, E.; Lane, D. Antibodies: a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory; 1988; and Ausubel et al., (2004) Current Protocols in Molecular Biology, John Wiley & Sons, New York N.Y. The assay can be competitive or non-competitive. Typically, immunological binding assays (or immunoassays) utilize a “capture agent” to specifically bind to and, often, immobilize the analyte. In one aspect, the capture agent can be a moiety that specifically binds to a variant ARMS2 polypeptide or subsequence. The bound protein may be detected using, for example, a detectably labeled anti-ARMS2 antibody. In one aspect, at least one of the antibodies is specific for a variant form of an ARMS2 polypeptide. In one aspect, the variant polypeptide can be detected using an immunoblot (Western blot) format.

[0077] Also described herein are purified polyclonal antibodies or fragments thereof that bind an ARMS2 polypeptide. Also described herein are isolated antibodies or fragments thereof that bind an ARMS2 polypeptide.

[0078] Described herein are antibodies that hybridize ARMS2 including, but not limited to antibodies that bind to amino acids 42-58 of ARMS2 (See NCBI Gene Accession No. 387715).

[0079] The antibodies described herein can recognize and hybridize to a reference ARMS2 polypeptide or a variant ARMS2 polypeptide, in which one or more non-synonymous single nucleotide polymorphisms (SNPs) are present in the ARMS2 coding region. In one aspect, the antibodies can specifically hybridize to a variant ARMS2 polypeptide or fragments thereof, but not an ARMS2 polypeptide without a variation at the polymorphic site. The antibodies can be polyclonal, and can be made according to standard protocols. Antibodies can be made by injecting a suitable animal with a wild type or variant ARMS2 polypeptide, or fragment thereof, or synthetic peptide fragments thereof.

[0080] Also described herein are methods of detecting ARMS2 in a subject comprising detecting ARMS2 levels using an antibody that specifically hybridizes to ARMS2. In one aspect, the antibodies described herein can specifically hybridize to the A69S variant ARMS2 protein. Methods to identify antibodies that specifically hybridizes to a polypeptide are well-known in the art. For methods, including antibody screening and subtraction methods; see Harlow & Lane, Antibodies, A Laboratory Manual, Cold Spring Harbor Press, New York (1988); Current Protocols in Immunology (J. E. Coligan et al., eds., 1999, including supplements through 2005); Goding, Monoclonal Antibodies, Principles and Practice (2d ed.) Academic Press, New York (1986); Burioni et al., 1998, "A new subtraction technique for molecular cloning of rare antiviral antibody specificities from phage display libraries" Res Virol. 149(5):327-30; Ames et al., 1994, Isolation of neutralizing anti-C5a monoclonal antibodies from a filamentous phage monovalent Fab display library. J Immunol. 152(9):4572-81; Shinohara et al., 2002, Isolation of monoclonal antibodies recognizing rare and dominant epitopes in plant vascular cell walls by phage display subtraction. J Immunol Methods 264(1-2):187-94. Immunization or screening can be directed against a full-length protein or, alternatively (and often more conveniently), against a peptide or polypeptide fragment comprising an epitope known to differ between the variant and wild-type forms.

Polyclonal antibodies specific for an ARMS2 polypeptide can be useful in diagnostic assays for detection of the variant forms of ARMS2, or as an active ingredient in a pharmaceutical composition.

[0081] Also described herein are methods of screening for polymorphic sites in other genes that are in linkage disequilibrium with a risk, protective, or otherwise informative SNP or genetic marker described herein, including but not limited to the polymorphic sites in the ARMS2 genes. These methods can involve identifying a polymorphic site in a gene that is in linkage disequilibrium with a polymorphic site in the ARMS2 gene, wherein the polymorphic form of the polymorphic site in the ARMS2 gene is associated with survival, (e.g., increased mortality risk), and determining haplotypes in a population of individuals to indicate whether the linked polymorphic site has a polymorphic form in linkage disequilibrium with the polymorphic form of the ARMS2 gene that correlates with survival, or increased mortality risk.

[0082] Polymorphisms in the ARMS2 gene, such as those described herein, can be used to establish physical linkage between a genetic locus associated with a trait of interest and polymorphic markers that are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Mapping a genetic locus associated with a trait of interest facilitates cloning the gene(s) responsible for the trait following procedures that are well-known in the art.

[0083] As used herein, "physical linkage" describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome. Linkage can be measured by percent recombination between the two genes, alleles, loci or genetic markers. Typically, loci occurring within a 50 centimorgan (cM) distance of each other are linked. Linked markers may occur within the same gene or gene cluster.

[0084] In one aspect, described herein are biological compounds, in particular, proteins, peptides, or nucleic acids, that are differentially present in samples from subjects with an increased mortality risk, as compared to age-matched control subjects (individuals without the increased mortality risk). These proteins therefore be associated with survival (e.g., increased mortality risk), and termed survival-associated biomarkers or biomarkers. These biomarkers can be present at different levels in individuals with an increased mortality

risk as compared to individuals without the increased mortality risk. These biomarkers can be present in individuals with an increased mortality risk at either elevated or reduced levels compared to individuals without an increased mortality risk. An exemplary biomarker shown to be present in individuals with an increased mortality risk at different levels compared to age-matched control individuals is ARMS2. Therefore, described herein are methods of determining ARMS2 expression levels in a subject. In one aspect, the methods comprise determining the ARMS2 – A69S expression levels in a subject. In practicing the methods described herein, biomarkers can be obtained in a sample, preferably a fluid sample, of the individual. The biomarkers are preferentially obtained in a sample of the individual's saliva, cheek scrapings, biopsies of retina, kidney or liver or other organs or tissues; skin biopsies; amniotic fluid or CNS samples; and the like.

[0085] As used herein, the term "biomarker" can refer to a protein found at different levels in a sample from a subject with an increased mortality risk compared to an age-matched control subject. The term "biomarker" can also refer to nucleic acid sequences, for example DNA or RNA sequences, such as the ARMS2 nucleic acid sequences described herein.

[0086] The biomarkers described herein can be in any form that provides information regarding presence or absence of a variant or SNP of the invention. For example, a disclosed biomarker can be, but is not limited to, a nucleic acid molecule, for example a DNA or RNA molecule, a polypeptide, or an antibody.

[0087] The term "level" refers to the amount of a biomarker in a sample obtained from an individual. The amount of the biomarker can be determined by any method known in the art and will depend in part on the nature of the biomarker (e.g., electrophoresis, including capillary electrophoresis, 1- and 2-dimensional electrophoresis, 2-dimensional difference gel electrophoresis DIGE followed by MALDI-ToF mass spectroscopy, chromatographic methods such as high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, mass spectrometry (MS), various immunological methods such as fluid or gel precipitin reactions, single or double immunodiffusion, immunoelectrophoresis, radioimmunoassay (RIA), enzyme-linked immunosorbant assays (ELISA), immunofluorescent assays, Western blotting and others, and enzyme- or function-based activity assays. It is understood that the amount of the biomarker

need not be determined in absolute terms, but can be determined in relative terms. For example, the amount of the biomarker may be expressed by its concentration in a sample, by the concentration of an antibody that binds to the biomarker, or by the functional activity (i.e., binding or enzymatic activity) of the biomarker.

[0088] The level(s) of a biomarker(s) can be determined as described above for a single biomarker or for a "set" of biomarkers. A set of biomarkers refers to a group of more than one biomarkers that have been grouped together, for example and not for limitation, by a shared property such as their presence at elevated levels in patients having an increased mortality risk compared to controls (e.g., difference between 1.25- and 2-fold, difference between 2- and 3-fold, difference between 3- and 5-fold, and difference of at least 5-fold), or by function.

[0089] The term "difference" as it relates to the level of a biomarker of the invention refers to a difference that is statistically different. A difference is statistically different, for example and not to be limiting, if the expectation is <0.05, the p value determined using the Student's t-test is <0.05, or if the p value determined using the Student's t-test is <0.1. The difference in level of a biomarker between an individual having an increased mortality risk and a control individual or population can be, for example and not to be limiting, at least 10% different (1.10 fold), at least 25% different (1.25-fold), at least 50% different (1.5-fold), at least 100% different (2-fold), at least 200% different (3-fold), at least 400% different (5-fold), at least 10-fold different, at least 20-fold different, at least 50-fold different, at least 100-fold different, at least 150-fold, or at least 200-fold different.

[0090] Survival-associated biomarkers can be detected in any of a number of methods including immunological assays (e.g., ELISA), separation-based methods (e.g., gel electrophoresis), protein-based methods (e.g., mass spectroscopy), function-based methods (e.g., enzymatic or binding activity), or the like. In one aspect, determining ARMS2 expression levels comprises using an antibody that specifically binds to ARMS2. In one aspect, the antibody specifically binds to ARMS2 – A69S. Other methods are known to those of skill in the art guided by this specification. The particular method for determining the levels will depend, in part, on the identity and nature of the biomarker protein. In one aspect, normal or baseline values (or ranges) can be established for biomarker expression levels. Normal levels can be determined for any particular population, subpopulation, or

group of organisms according to standard methods well known to those of skill in the art. Generally, baseline (normal) levels of biomarkers can be determined by quantifying the amount of biomarker in biological samples (e.g., fluids, cells or tissues) obtained from normal (healthy) subjects. Application of standard statistical methods used in medicine permits determination of baseline levels of expression, as well as significant deviations from such baseline levels. It will be appreciated that the assay methods do not necessarily require measurement of absolute values of biomarker, unless it is so desired, because relative values can be sufficient for many applications of the methods described herein. Where quantification is desirable, described herein are reagents such that virtually any known method for quantifying gene products can be used.

[0091] In one aspect, the method for separating and determining the levels of the one or more biomarkers described herein, including, but not limited to, ARMS2, can involve obtaining a biological sample from an individual, separating and determining the levels of the biomarkers by 2-dimensional difference gel electrophoresis (DIGE), and identifying the biomarkers by MALDI-ToF mass spectroscopy. In another aspect, the biomarkers separated by DIGE can be identified by comparison to a known separation pattern of biomarkers using DIGE.

[0092] In a further aspect, the method for separating, detecting, and determining the levels of the biomarkers described herein, including, but not limited to, ARMS2, involves obtaining a biological sample from an individual, separating the proteins by chromatography, if appropriate, capturing the proteins on a biochip (i.e., an adsorbent of a SELDI probe), and detecting and determining the levels of the captured biomarkers by mass spectrometry (i.e., ToF-MS).

[0093] A biochip can comprise a solid substrate and can have a generally planar surface to which a capture reagent (also called an adsorbent or affinity reagent) can be attached. Frequently, the surface of a biochip comprises a plurality of addressable locations, each of which can have the capture reagent bound thereto.

[0094] A "protein biochip" as used herein refers to a biochip adapted for the capture of proteins. Protein biochips are known to those of skill in the art, including, but not limited to, those produced by Ciphergen Biosystems, Inc. (Fremont, Calif.), Packard BioScience Company (Meriden, Conn.), Zyomyx (Hayward, Calif.), Phylos (Lexington, Mass.) and

Biacore (Uppsala, Sweden). Examples of such protein biochips are described in, e.g., U.S. Pat. Nos. 6,225,047, 6,329,209 and 5,242,828, and PCT Publication Nos. WO 99/51773 and WO 00/56934.

[0095] In one aspect, the biomarkers of the invention can be detected by mass spectrometry (MS) methods. Examples of mass spectrometers include, but are not limited to, time-of-flight (ToF), magnetic sector, quadrupole filter, ion trap, ion cyclotron resonance, electrostatic sector analyzer, and hybrids of these.

[0096] In one aspect, the mass spectrometer can be a laser desorption/ionization mass spectrometer. In laser desorption/ionization mass spectrometry, the analytes (i.e., proteins) are placed on the surface of a MS probe, which engages a probe interface of the mass spectrometer and presents an analyte to ionizing energy for ionization and introduction into the mass spectrometer. A laser desorption mass spectrometer employs laser energy, typically from an ultraviolet laser, but also from an infrared laser, which desorbs the analytes from the surface, and volatilizes and ionizes the analytes, thereby making them available to the ion optics of the mass spectrometer.

[0097] A mass spectrometry method for use in the methods described herein can be "Surface Enhanced Laser Desorption and Ionization" or "SELDI," as described, for example, in U.S. Pat. Nos. 5,719,060 and 6,225,047. SELDI refers to a method of desorption/ionization gas phase ion spectrometry in which the analyte (i.e., at least two of the biomarkers) is captured on the surface of a SELDI MS probe. There are several versions of SELDI, including "affinity capture mass spectrometry," "Surface-Enhanced Affinity Capture" or "SEAC," "Surface-Enhanced Neat Desorption" or "SEND," and "Surface-Enhanced Photolabile Attachment and Release" or "SEPAR".

[0098] SEAC involves the use of probes having a material on the probe surface that captures analytes (i.e., proteins) through non-covalent affinity interactions (i.e., adsorption) between the material and the analyte. The material is variously called an "adsorbent," a "capture reagent," an "affinity reagent, or a "binding moiety." Such probes are called "affinity capture probes" having "adsorbent surfaces." The capture reagent can be any material capable of binding an analyte. The capture reagent can be attached directly to the substrate of the selective surface, or the substrate can have a reactive surface that carries a reactive moiety capable of binding the capture reagent, e.g., through a reaction forming a

covalent or coordinate covalent bond. Epoxide and carbodiimidizole can be reactive moieties used to covalently bind protein capture reagents, such as antibodies or cellular receptors. Nitriloacetic acid and iminodiacetic acid can be reactive moieties that function as chelating agents to bind metal ions that interact non-covalently with histidine containing peptides. Adsorbents can be generally classified as either chromatographic adsorbents or biospecific adsorbents.

[0099] A "chromatographic adsorbent" refers to an adsorbent material typically used in chromatography. Chromatographic adsorbents include, for example, anion and cation exchange materials, metal chelators (e.g., nitriloacetic acid or iminodiacetic acid), immobilized metal chelates, hydrophobic interaction adsorbents, hydrophilic interaction adsorbents, dyes, simple biomolecules (e.g., nucleotides, amino acids, simple sugars and fatty acids) and mixed mode adsorbents (e.g., hydrophobic attraction/electrostatic repulsion adsorbents).

[00100] A "biospecific adsorbent" refers to an adsorbent comprising a biomolecule, e.g., a nucleic acid molecule (e.g., an aptamer), a polypeptide, a polysaccharide, a lipid, a steroid or a conjugate of these (e.g., a glycoprotein, a lipoprotein, a glycolipid, or a nucleic acid (e.g., DNA)-protein conjugate). In some aspects, the biospecific adsorbent can be a macromolecular structure, such as a multi-protein complex, a biological membrane or a virus. Examples of biospecific adsorbents include, but are not limited to, antibodies, receptor proteins and nucleic acids. Biospecific adsorbents can have higher specificity for a target analyte than chromatographic adsorbents. Further examples of adsorbents for use in SELDI can be found in U.S. Pat. No. 6,225,047. A "bioselective adsorbent" refers to an adsorbent that binds to an analyte with an affinity typically of at least 10^{-8} M.

[00101] Protein biochips produced by Ciphergen Biosystems, Inc. comprise surfaces having chromatographic or biospecific adsorbents attached thereto at addressable locations. Ciphergen PROTEINCHIP arrays include NP20 (hydrophilic); H4 and HSO (hydrophobic); SAX2, Q10 and LSAX30 (anion exchange); WCX2, CM10 and LWCX30 (cation exchange); IMAC3, IMAC30 and IMAC40 (metal chelate); and PS10, PS20 (reactive surface with carboimidizole, expoxide) and PG20 (protein G coupled through carboimidizole). Hydrophobic PROTEINCHIP arrays have isopropyl or nonylphenoxy-poly(ethylene glycol)methacrylate functionalities. Anion exchange PROTEINCHIP arrays have quaternary

ammonium functionalities. Cation exchange PROTEINCHIP arrays have carboxylate functionalities. Immobilized metal chelate PROTEINCHIP arrays have nitriloacetic acid functionalities that adsorb transition metal ions, such as copper, nickel, zinc, and gallium, by chelation. Preactivated PROTEINCHIP arrays have carboimidizole or epoxide functional groups that can react with groups on proteins for covalent binding.

[00102] Protein biochips are further described in U.S. Pat. Nos. 6,579,719 and 6,555,813, PCT Publication Nos. WO 00/66265 and WO 03/040700, U.S. Patent Application Nos. US 20030032043 A1, US 20030218130 A1 and US 20050059086 A1.

[00103] In one aspect, a probe with an adsorbent surface can be contacted with the sample for a period of time sufficient to allow proteins present in the sample to bind to the adsorbent. After the incubation period, the substrate can be washed to remove unbound material. Any suitable washing solutions can be used; for example, aqueous solutions can be employed. The extent to which proteins remain bound to the adsorbent can be manipulated by adjusting the stringency of the wash. The elution characteristics of a wash solution can depend, for example, on pH, ionic strength, hydrophobicity, degree of chaotropism, detergent strength, temperature, and the like. Unless the probe has both SEAC and SEND properties (as described herein), an energy absorbing molecule can then applied to the substrate with the bound proteins.

[00104] The biomarkers bound to the substrates can be detected in a gas phase ion spectrometer such as a ToF mass spectrometer. The biomarkers can be ionized by an ionization source such as a laser, the generated ions can be collected by an ion optic assembly, and then a mass analyzer can disperse and analyze the passing ions. The detector can then translate information of the detected ions into mass-to-charge ratios. Detection of a biomarker can involve detection of signal intensity. Thus, both the quantity and mass of the biomarker can be determined.

[00105] SEND involves the use of probes comprising energy absorbing molecules that are chemically bound to the probe surface ("SEND probe"). The phrase "energy absorbing molecules" (EAM) denotes molecules that are capable of absorbing energy from a laser desorption/ionization source and, thereafter, contribute to desorption and ionization of analyte molecules in contact therewith. The EAM category includes molecules used in MALDI, frequently referred to as "matrix," and is exemplified by cinnamic acid derivatives,

sinapinic acid (SPA), cyano-hydroxy-cinnamic acid (CHCA) and dihydroxybenzoic acid, ferulic acid, and hydroxyaceto-phenone derivatives. In one aspect, the EAM can be incorporated into a linear or cross-linked polymer, e.g., a polymethacrylate. For example, the composition can be a co-polymer of α-cyano-4-methacryloyloxycinnamic acid and acrylate. In another aspect, the composition is a co-polymer of α-cyano-4-methacryloyloxycinnamic acid, acrylate and 3-(tri-ethoxy)silyl propyl methacrylate. In another aspect, the composition can be a co-polymer of α-cyano-4-methacryloyloxycinnamic acid and octadecylmethacrylate ("C18 SEND"). SEND is further described in U.S. Pat. No. 6,124,137 and PCT Publication No. WO 03/64594.

[00106] SEAC/SEND is a version of SELDI in which both a capture reagent and an EAM can be attached to the sample presenting surface. SEAC/SEND probes can therefore allow the capture of analytes through affinity capture and ionization/desorption without the need to apply an external matrix. The C18 SEND biochip is a version of SEAC/SEND, comprising a C18 moiety which functions as a capture reagent, and a CHCA moiety which functions as an EAM.

[00107] SEPAR involves the use of probes having moieties attached to the surface that can covalently bind an analyte, and then release the analyte through breaking a photolabile bond in the moiety after exposure to light, e.g., to laser light (see U.S. Pat. No. 5,719,060). SEPAR and other forms of SELDI can be readily adapted to detecting a biomarker or biomarker profile, pursuant to the methods described herein.

[00108] In another MS method, the biomarkers can first be captured on a resin having chromatographic properties that bind biomarkers. This can include a variety of methods. For example, the biomarkers can be captured on a cation exchange resin, such as CM CERAMIC HYPERD F resin, the resin can be washed, biomarkers can be eluted and the eluted biomarkers can be detected by MALDI. Alternatively, this method can be preceded by fractionating the sample on an anion exchange resin, such as Q CERAMIC HYPERD F resin, before application to the cation exchange resin. In another aspect, the sample on an anion exchange resin can be fractionated and detected by MALDI directly. In yet another aspect, the biomarkers can be captured on an immuno-chromatographic resin comprising antibodies that bind particular biomarkers, resin can be washed to remove unbound material, the

biomarkers can be eluted from the resin and the eluted biomarkers can be detected by MALDI or by SELDI.

[00109] Analysis of analytes by ToF-MS generates a time-of-flight spectrum. The time-of-flight spectrum ultimately analyzed typically does not represent the signal from a single pulse of ionizing energy against a sample, but rather the sum of signals from a number of pulses. This reduces noise and increases dynamic range. This time-of-flight data can then be subject to data processing using Ciphergen's PROTEINCHIP software, or any equivalent data processing software. Data processing can include TOF-to-M/Z transformation to generate a mass spectrum, baseline subtraction to eliminate instrument offsets and high frequency noise filtering to reduce high frequency noise.

[00110] Data generated by desorption and detection of biomarkers can be analyzed with the use of a programmable digital computer. The computer program can analyze the data to indicate the number of biomarkers detected, the strength of the signal, or the determined molecular mass for each biomarker detected. Data analysis can include steps of determining signal strength of a biomarker and removing data deviating from a predetermined statistical distribution. For example, the observed peaks can be normalized, by calculating the height of each peak relative to some reference. The reference can be background noise generated by the instrument and chemicals such as the energy absorbing molecule which can be set at zero in the scale.

[00111] The computer can transform the resulting data into various formats for display. The standard spectrum can be displayed, but in one aspect only the peak height and mass-to-charge information can be retained from the spectrum view, thereby yielding a cleaner image and enabling biomarkers with nearly identical molecular weights to be more easily seen. In another aspect, two or more spectra can be compared, conveniently highlighting unique biomarkers and biomarkers that are up- or down-regulated between samples. Using any of these formats, it can readily be determined whether a particular biomarker is present in a sample.

[00112] Analysis can involve the identification of peaks in the spectrum that represent signal from an analyte. Peak selection can be done visually, but software is available, for example, as part of Ciphergen's PROTEINCHIP software package, which can automate the detection of peaks. In general, this software functions by identifying signals

having a signal-to-noise ratio above a selected threshold and labeling the mass of the peak at the centroid of the peak signal. In one aspect, many spectra can be compared to identify identical peaks present in some selected percentage of the mass spectra. One version of this software clusters all peaks appearing in the various spectra within a defined mass range, and assigns a mass (M/Z) to all the peaks that are near the mid-point of the mass (M/Z) cluster.

[00113] Software used to analyze the data can include code that applies an algorithm to the analysis of the signal to determine whether the signal represents a peak in a signal that corresponds to a biomarker described herein. The software also can subject the data regarding observed biomarker peaks to classification tree or ANN analysis, to determine whether a biomarker peak or combination of biomarker peaks is present that indicates the status of the particular clinical parameter under examination. Analysis of the data can be "keyed" to a variety of parameters that are obtained, either directly or indirectly, from the mass spectrometric analysis of the sample. These parameters include, but are not limited to, the presence or absence of at least two peaks, the shape of a peak or group of peaks, the height of at least two peaks, the log of the height of at least two peaks, and other arithmetic manipulations of peak height data.

[00114] An example protocol for the detection of biomarkers described herein is as follows. The biological sample to be tested can be obtained a subject, depleted of albumin and IgG or pre-fractionated on an anion exchange chromatographic resin or other chromatographic resin, as appropriate, and then contacted with an affinity capture SELDI probe comprising a cation exchange adsorbant (e.g., CM10 or WCX2 PROTEINCHIP array from Ciphergen Systems, Inc.), an anion exchange adsorbant (e.g., Q10 PROTEINCHIP array from Ciphergen Systems, Inc.), a hydrophobic exchange adsorbant (e.g., HSO PROTEINCHIP array from Ciphergen Systems, Inc.), or an IMAC adsorbant (e.g., IMAC3 or IMAC30 PROTEINCHIP array from Ciphergen Systems, Inc.). The SELDI probe can be washed with a suitable buffer that retains the biomarkers of the invention, while washing away unbound biomolecules. The biomarkers specifically retained on the SELDI probe can then be detected by laser desorption/ionization mass spectrometry.

[00115] The biological sample, e.g., serum, plasma or urine, can be depleted of albumin and IgG or subjected to pre-fractionation before binding to a SELDI probe. In one aspect, pre-fractionation can involve contacting the biological sample with an anion exchange

chromatographic resin. The bound biomolecules can then be subjected to stepwise pH elution using buffers at various pH. Various fractions containing biomolecules can be collected and subjected to binding to a SELDI probe.

[00116] In a further aspect, if analysis of particular proteins and various forms thereof are desired, antibodies which recognize specific proteins can be attached to the surface of a SELDI probe (e.g., pre-activated PS10 or PS20 PROTEINCHIP array from Ciphergen Systems, Inc.). The antibodies capture the target proteins from a biological sample onto the SELDI probe. The captured proteins can then be detected by, for example, laser desorption/ionization mass spectrometry. The antibodies can also capture the target proteins on immobilized support, and the target proteins can be eluted and captured on a SELDI probe and detected as described herein.

[00117] Antibodies to target proteins are either commercially available or can be produced by methods known in the art, e.g., by immunizing animals with the target proteins isolated by standard purification techniques or with synthetic peptides of the target proteins.

[00118] In some aspects it will be desirable to establish normal or baseline values (or ranges) for biomarker expression levels. Normal levels can be determined for any particular population, subpopulation, or group of organisms according to standard methods well known to those of skill in the art. Generally, baseline (normal) levels of biomarkers are determined by quantifying the amount of biomarker in biological samples (e.g., fluids, cells or tissues) obtained from normal (healthy) subjects. Application of standard statistical methods used in medicine permits determination of baseline levels of expression, as well as significant deviations from such baseline levels.

[00119] It will be appreciated that the assay methods described herein do not necessarily require measurement of absolute values of a biomarker, unless it is so desired, because relative values are sufficient for many applications of the methods described herein. Where quantification is desirable, the presently described methods provide reagents such that virtually any known method for quantifying gene products can be used.

[00120] In one aspect, described herein are methods for determining the a subject's increased mortality risk by determining levels of at least one survival-associated biomarker in a sample from the individual, and comparing the levels of the biomarker in the sample to

reference levels of the biomarker characteristic of a control population of individuals without an increased mortality risk, where a difference in the levels of the biomarker between the sample from the individual and the control population indicates that the individual has an increased mortality risk. A biomarker can be, but is not limited to, ARMS2 protein. For example, the biomarker can be ARMS2 protein expressed at elevated levels in individuals with an increased mortality risk. In another aspect, the biomarker can be an ARMS2 nucleic acid, such as DNA or RNA.

[00121] Also disclosed herein are imaging agents, wherein the agent specifically binds an ARMS2, or a variant ARMS2, encoding nucleic acid. For example, disclosed are arrays comprising polynucleotides capable of specifically hybridizing to one or more ARMS2 SNPs described herein. Also disclosed are imaging agents, wherein the agent is capable of specifically hybridizing to one or more of the one or more ARMS2 SNPs including, but not limited to, rs10490924.

[00122] It is understood that the methods described herein, of predicting a subject's survival, comprising determining in the subject the identity of one or more SNPs described herein, can be performed in combination with any of the other methods described herein.

B. ARRAYS

[00123] Also described herein are arrays comprising polynucleotides capable of specifically hybridizing to ARMS2 or a variant ARMS2 encoding nucleic acid. For example, described are arrays comprising polynucleotides capable of specifically hybridizing to the rs10490924 SNP.

[00124] Also described herein are solid supports comprising one or more polypeptides capable of specifically hybridizing to an ARMS2 or a variant ARMS2 peptide.

[00125] Solid supports are solid-state substrates or supports with which molecules, such as analytes and analyte binding molecules, can be associated. Analytes, such as calcifying nano-particles and proteins, can be associated with solid supports directly or indirectly. For example, analytes can be directly immobilized on solid supports. Analyte capture agents, such as capture compounds, can also be immobilized on solid supports. For example, described herein are antigen binding agents capable of specifically binding to an ARMS2 or a variant ARMS2 peptide.

[00126] A preferred form of solid support is an array. Another form of solid support is an array detector. An array detector is a solid support to which multiple different capture compounds or detection compounds have been coupled in an array, grid, or other organized pattern.

[00127] Solid-state substrates for use in solid supports can include any solid material to which molecules can be coupled. This includes materials such as acrylamide, agarose, cellulose, nitrocellulose, glass, polystyrene, polyethylene vinyl acetate, polypropylene, polymethacrylate, polyethylene, polyethylene oxide, polysilicates, polycarbonates, teflon, fluorocarbons, nylon, silicon rubber, polyanhydrides, polyglycolic acid, polylactic acid, polyorthoesters, polypropylfumerate, collagen, glycosaminoglycans, and polyamino acids. Solid-state substrates can have any useful form including thin film, membrane, bottles, dishes, fibers, woven fibers, shaped polymers, particles, beads, nanoparticles, microparticles, or a combination. Solid-state substrates and solid supports can be porous or non-porous. A preferred form for a solid-state substrate is a microtiter dish, such as a standard 96-well type. In preferred embodiments, a multiwell glass slide can be employed that normally contain one array per well. This feature allows for greater control of assay reproducibility, increased throughput and sample handling, and ease of automation.

[00128] Different compounds can be used together as a set. The set can be used as a mixture of all or subsets of the compounds used separately in separate reactions, or immobilized in an array. Compounds used separately or as mixtures can be physically separable through, for example, association with or immobilization on a solid support. An array can include a plurality of compounds immobilized at identified or predefined locations on the array. Each predefined location on the array generally can have one type of component (that is, all the components at that location are the same). Each location will have multiple copies of the component. The spatial separation of different components in the array allows separate detection and identification of the polynucleotides or polypeptides described herein.

[00129] Although preferred, it is not required that a given array be a single unit or structure. The set of compounds may be distributed over any number of solid supports. For example, at one extreme, each compound may be immobilized in a separate reaction tube or container, or on separate beads or microparticles or nanoparticles. Different modes of the

disclosed method can be performed with different components (for example, different compounds specific for different proteins) immobilized on a solid support.

[00130] Some solid supports can have capture compounds, such as antibodies, attached to a solid-state substrate. Such capture compounds can be specific for calcifying nano-particles or a protein on calcifying nano-particles. Captured calcifying nano-particles or proteins can then be detected by binding of a second, detection compound, such as an antibody. The detection compound can be specific for the same or a different protein on the calcifying nano-particle.

[00131] Methods for immobilizing antibodies (and other proteins) to solid-state substrates are well established. Immobilization can be accomplished by attachment, for example, to aminated surfaces, carboxylated surfaces or hydroxylated surfaces using standard immobilization chemistries. Examples of attachment agents are cyanogen bromide, succinimide, aldehydes, tosyl chloride, avidin-biotin, photocrosslinkable agents, epoxides and maleimides. A preferred attachment agent is the heterobifunctional cross-linker N-[γ-Maleimidobutyryloxy] succinimide ester (GMBS). These and other attachment agents, as well as methods for their use in attachment, are described in Protein immobilization: fundamentals and applications, Richard F. Taylor, ed. (M. Dekker, New York, 1991);, Johnstone and Thorpe, Immunochemistry In Practice (Blackwell Scientific Publications, Oxford, England, 1987) pages 209-216 and 241-242, and Immobilized Affinity Ligands; Craig T. Hermanson *et al.*, eds. (Academic Press, New York, 1992) which are incorporated by reference in their entirety for methods of attaching antibodies to a solid-state substrate. Antibodies can be attached to a substrate by chemically cross-linking a free amino group on the antibody to reactive side groups present within the solid-state substrate. For example, antibodies may be chemically cross-linked to a substrate that contains free amino, carboxyl, or sulfur groups using glutaraldehyde, carbodiimides, or GMBS, respectively, as cross-linker agents. In this method, aqueous solutions containing free antibodies are incubated with the solid-state substrate in the presence of glutaraldehyde or carbodiimide.

[00132] A preferred method for attaching antibodies or other proteins to a solid-state substrate is to functionalize the substrate with an amino- or thiol-silane, and then to activate the functionalized substrate with a homobifunctional cross-linker agent such as (Bis-sulfo-succinimidyl suberate (BS^3) or a heterobifunctional cross-linker agent such as GMBS. For

cross-linking with GMBS, glass substrates are chemically functionalized by immersing in a solution of mercaptopropyltrimethoxysilane (1% vol/vol in 95% ethanol pH 5.5) for 1 hour, rinsing in 95% ethanol and heating at 120 °C for 4 hrs. Thiol-derivatized slides are activated by immersing in a 0.5 mg/ml solution of GMBS in 1% dimethylformamide, 99% ethanol for 1 hour at room temperature. Antibodies or proteins are added directly to the activated substrate, which are then blocked with solutions containing agents such as 2% bovine serum albumin, and air-dried. Other standard immobilization chemistries are known by those of skill in the art.

[00133] Each of the components (compounds, for example) immobilized on the solid support preferably is located in a different predefined region of the solid support. Each of the different predefined regions can be physically separated from each of the other different regions. The distance between the different predefined regions of the solid support can be either fixed or variable. For example, in an array, each of the components can be arranged at fixed distances from each other, while components associated with beads will not be in a fixed spatial relationship. In particular, the use of multiple solid support units (for example, multiple beads) will result in variable distances.

[00134] Components can be associated or immobilized on a solid support at any density. Components preferably are immobilized to the solid support at a density exceeding 400 different components per cubic centimeter. Arrays of components can have any number of components. For example, an array can have at least 1,000 different components immobilized on the solid support, at least 10,000 different components immobilized on the solid support, at least 100,000 different components immobilized on the solid support, or at least 1,000,000 different components immobilized on the solid support.

[00135] Optionally, at least one address on the solid support is the sequences or part of the sequences set forth in any of the nucleic acid sequences described herein. Also disclosed are solid supports where at least one address is the sequences or portion of sequences set forth in any of the peptide sequences described herein. Solid supports can also contain at least one address is a variant of the sequences or part of the sequences set forth in any of the nucleic acid sequences described herein. Solid supports can also contain at least one address as a variant of the sequences or portion of sequences set forth in any of the peptide sequences described herein.

[00136] Also disclosed are antigen microarrays for multiplex characterization of antibody responses. For example, disclosed are antigen arrays and miniaturized antigen arrays to perform large-scale multiplex characterization of antibody responses directed against the polypeptides, polynucleotides and antibodies described herein, using submicroliter quantities of biological samples as described in Robinson *et al.*, Autoantigen microarrays for multiplex characterization of autoantibody responses, Nat Med., 8(3):295-301 (2002), which is herein incorporated by reference in its entirety for its teaching of constructing and using antigen arrays to perform large-scale multiplex characterization of antibody responses directed against structurally diverse antigens, using submicroliter quantities of biological samples.

[00137] Protein variants and derivatives are well understood to those of skill in the art and can involve amino acid sequence modifications. For example, amino acid sequence modifications typically fall into one or more of three classes: substitutional, insertional or deletional variants. Polypeptide variants described herein will typically exhibit at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity (determined as described below), along their length, to the polypeptide sequences set forth herein.

C. KITS

[00138] Also described herein are kits for performing the methods described herein. The kits described herein can comprise an assay or assays for detecting one or more SNPs in a nucleic acid sample of a subject, or obtained from a subject, wherein the one or more SNPs are the SNPs described herein. For example, and not to be limiting, the one or more SNPs can be the rs10490924 SNP in the ARMS2 gene. The kits described herein can further comprise amplification reagents for amplifying the ARMS2 locus.

[00139] For example, and not to be limiting, the kits described herein can comprise an assay for detecting a SNP in a nucleic acid sample of a subject, wherein the SNP is rs10490924 in the ARMS2 gene. In another aspect, the kits described herein can comprise an assay for detecting the genotype at the rs10490924 SNP. For example, and not to be limiting, the kits described herein can be used to detect the TT genotype at rs10490924 SNP. In one aspect, the kits can further comprise instructions for correlating the assay results with the subject's mortality risk.

[00140] Also described herein are kits comprising one or more primers or probes for detecting a SNP in a nucleic acid sample of a subject, wherein the SNP is rs10490924 in the ARMS2 gene. The kits described herein can comprise any of the primers or probes described herein. In one aspect, primers and probes disclosed herein can specifically hybridize to rs10490924 or possess sufficient specificity to distinguish between particular genotype of re10490924 (i.e., have specificity sufficient to distinguish TT genotype, GG genotype, and GT genotype). In one aspect, the primers and probes can be purchased from a commercial source, such as Applied Biosystems. In a futher aspect, the kits described herein can further comprise instructions for correlating the presence of rs10490924 to the subject's mortality risk. It is further disclosed herein that the kits described herein can include a forward and reverse primer pair and a reverse transcriptase and optionally a reverse/transcriptase for the synthesis of cDNA.

D. EXAMPLES

[00141] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how the compounds, compositions, articles, devices and/or methods claimed herein are made and evaluated, and are intended to be purely exemplary of the invention and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers (e.g., amounts, temperature, etc.), but some errors and deviations should be accounted for.

1. Rs10490924 Genotype Associates with Earlier Mortality

[00142] In the methods and examples described herein, all SNPs were genotyped with predesigned or custom TaqMan assays (Applied Biosystems, Foster City, California) using 10 ng of template DNA in a 5uL reaction. The thermal cycling conditions in the 384-well thermocycler (PTC-225, MJ Research) consisted of an initial hold at 95°C for 10 minutes, followed by 40 cycles of a 15-second 95°C denaturation step and a 1-minute 60°C annealing and extension step. Plates were read in a 7900HT Fast Real-Time PCR System (Applied Biosystems). Subsequently, statistical calculations were done using Microsoft Excel.

[00143] Genotypes from 2,725 donor individuals (eye donors and people who donated other organs) were ascertained postmortem at disease-related SNP locations on chromosome 1 (rs800292, rs1061170, rs1410996, rs1409153, rs10922153, rs698859), chromosome 6 (rs547154), chromosome 10 (rs10490924) and chromosome 19 (rs2230199). The average age at death for individuals who were homozygous for the disease risk allele of a given SNP was calculated and compared to those who were homozygous for the non-risk allele (See FIG. 2). rs1410996 and rs10490924 showed an age at death difference of greater than 0.5 years. Donors who died as a result of accidents or suicides were subsequently removed from the analysis and the average age at death was recalculated for males and females of each rs1410996 and rs10490924 genotype group. A Student's t-test was performed to compare the average age at death values of homozygous groups. Male donors with TT genotypes at the rs10490924 SNP showed an average age at death 6.2 years younger than males with the GG genotype (p=0.0015) (See FIG. 4).

CLAIMS

What is claimed is:

1. A method for predicting survival of a subject comprising determining in the subject the identity of a SNP in the ARMS2 gene, wherein the SNP is rs10490924, and wherein the presence of the rs10490924 SNP is predictive of the subject's mortality risk.

2. The method of claim 1, wherein the subject is male.

3. The method of claim 2, wherein a TT genotype at the rs10490924 SNP indicates that subject has an increased mortality risk.

4. The method of claim 32, wherein a GG genotype at the rs10490924 SNP indicates that the subject does not have an increased mortality risk.

5. The method of claim 2, wherein a GT genotype at the rs10490924 SNP indicates that the subject does not have an increased mortality risk.

6. The method of claim 1, wherein the subject is a female subject.

7. The method of claim 1, further comprising obtaining a nucleic acid sample from a subject.

8. The method of claim 7, further comprising sequencing the nucleic acid sample obtained from the subject.

9. A kit comprising an assay for detecting a SNP in a nucleic acid sample of a subject comprising one or more probes that specifically hybridize to the SNP, wherein the SNP is rs10490924 in the ARMS2 gene.

10. The kit of claim 9, further comprising instructions for correlating the assay results with the subject's mortality risk.

11. A kit comprising:

a. one or more primers for detecting a SNP in a nucleic acid sample of a subject, wherein the SNP is rs10490924 in the ARMS2 gene; and

b. instructions for correlating the presence of rs10490924 to the subject's mortality risk.

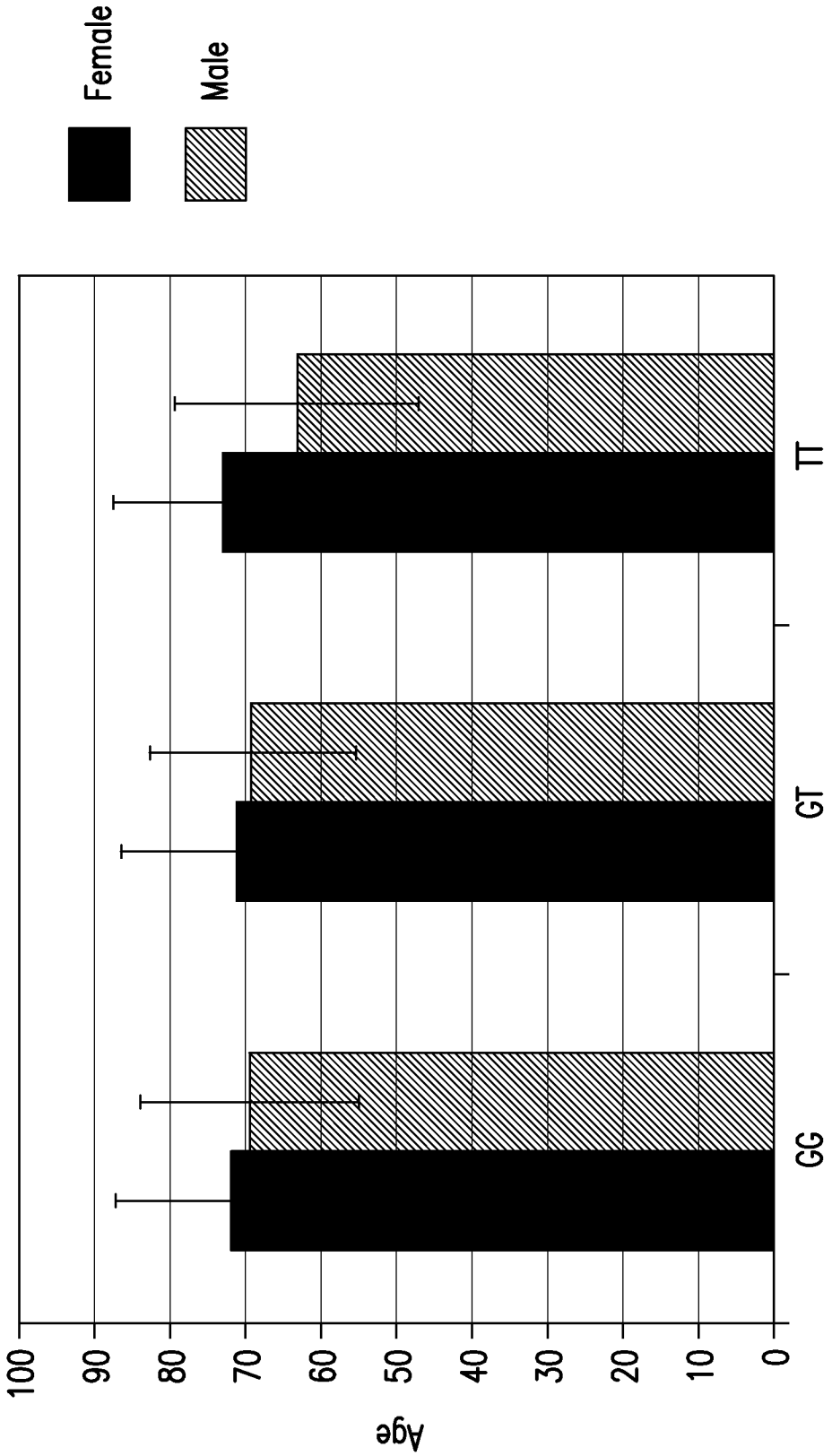


FIG. 1

SNP	Risk Gtype	Ave.Age of Death	Prot Gtype	Ave.Age of Death	Difference
rs10490924	TT	65.2	GG	67.9	2.7
rs1061170	CC	67.5	TT	67.4	−0.1
rs2230199	GG	68.1	CC	67.9	−0.2
rs800292	GG	68	AA	67.5	−0.5
rs1410996	GG	67.2	AA	68.4	1.2
rs1409153	CC	67.7	TT	67.4	−0.3
rs698859	TT	67.9	CC	67.5	−0.4
rs10922153	GG	67.4	TT	67.5	0.1
rs547154	GG	67.2	TT	only 20 samples	
*data from approx 2800 donor samples					

FIG. 2

rs10490924			
	GG	*TT*	Difference
Female	70.25	68.8	1.45
Male	66.4	62.3	4.1

FIG. 3

With accidents/suicides removed (n=2562):					
Total:					
Female (n=1101)	72.0				
Male (n=1436)	69.5				
rs10490924	*GG*	*GT*	*TT*	*GG vs TT p-value*	
Female	71.9 (n=585)	71.3 (n=334)	73 (n=54)	0.685	
Male	69.7 (n=788)	69.3 (n=456)	63.5 (n=68)	0.0015	
rs1410996	*AA*	*AG*	*GG*	*AA vs GG p-value*	
Female	70.8 (n=175)	71.6 (n=457)	72.2 (n=325)	0.373	
Male	70.5 (N=233)	69.0 (n=596)	69.1 (n=447)	0.255	

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US13/59778

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68 (2013.01)
USPC - 435/6.11, 6.1, 4, 6.12, 6.18, 91.2; 536/23.5, 24.31
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8): C12Q 1/68 (2013.01)
USPC: 435/6.11, 6.1, 4, 6.12, 6.18, 91.2; 536/23.5, 24.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MicroPatent (US-G, US-A, EP-A, EP-B, WO, JP-bib, DE-C,B, DE-A, DE-T, DE-U, GB-A, FR-A); Google Scholar/ Google Patent; Proquest; PubMed; Search terms ---- Mortality, death, Single nucleotide polymorphisms, SNP, Rs10490924, Age related maculopathy susceptibility 2, ARMS2, LOC387715

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	XU, Y et al. Association Of CFH, LOC387715, And HTRA1 Polymorphisms With Exudative Age-Related Macular Degeneration In A Northern Chinese Population. Molecular Vision, 2008, Vol. 14, pages 1373-1381; abstract; page 1374, column 2, paragraph 2; page 1376, column 1, paragraph 5; page 1337, column 2, paragraph 2; Tables 1-3.	9 --- 1-8, 10-11
Y	CUGATI, S et al. Visual Impairment, Age-Related MacularDegeneration, Cataract, And Long-Term Mortality. Arch Ophthalmol. 2007, Vol. 125, No. 7, pages 917-924; abstract; page 917, column 1, paragraph 1; page 920, column 1, paragraph 1.	1-8, 10-11

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 December 2013 (10.12.2013)	16 DEC 2013
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774