

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/11* (2006.01)(21) International Application Number:
PCT/US2009/057511(22) International Filing Date:
18 September 2009 (18.09.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/098,112 18 September 2008 (18.09.2008) US(71) Applicant (for all designated States except US):
GEISINGER CLINIC [US/US]; 100 North Academy Avenue, Danville, Pennsylvania 17822 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ELMORE, James R.** [US/US]; 100 North Academy Avenue, Danville, Pennsylvania 17822 (US). **CAREY, David J.** [US/US]; 100 North Academy, Danville, Pennsylvania 17822 (US).(74) Agents: **ZAHER, Alfred, W.** et al.; 600 New Hampshire Avenue, NW, Washington, District of Columbia 20037 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

— with sequence listing part of description (Rule 5.2(a))

(54) Title: GENETIC VARIANTS ASSOCIATED WITH ABDOMINAL AORTIC ANEURYSMS

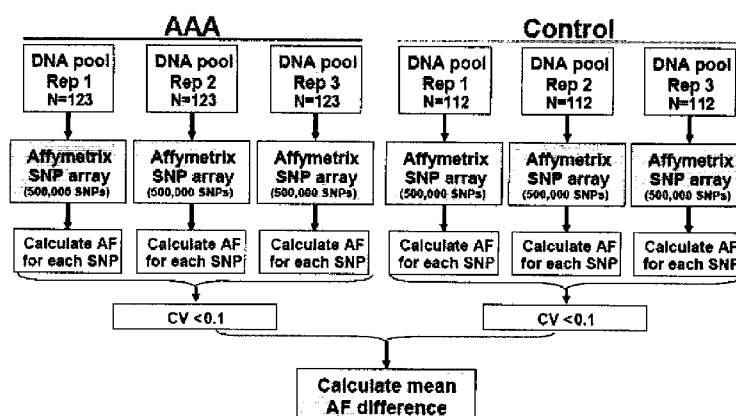


Figure 1

(57) Abstract: Methods and kits for determining the risk of developing an abdominal aortic aneurysm are described. The methods of the present invention involve analyzing nucleic acid samples from an individual to determine the presence of one or more polymorphisms associated with abdominal aortic aneurysms.

Genetic Variants Associated With Abdominal Aortic Aneurysms

Statement of Priority

[0001] This application claims priority to U.S. Provisional Patent Application No. 61/098,112, filed September 18, 2008, the disclosure of which is hereby incorporated by reference herein.

Field of the Invention

[0002] The present invention relates to methods and kits for identifying individuals who are at risk for developing abdominal aortic aneurysms (AAAs) by detection of genetic variants associated with AAA formation. The methods of the present invention allow for individuals showing a genetic susceptibility for developing AAAs to undergo monitoring and treatment as necessary.

Background of the Invention

[0003] Abdominal aortic aneurysms (AAAs) are a leading cause of death in the elderly, usually from aneurysm rupture. (1,2) Many patients with AAAs are undiagnosed. The lack of diagnosis increases their risk of death from AAA, since surgical treatments to repair aneurysms before they rupture are safe and effective. The most important risk factors for AAAs are old age, male gender, smoking, and family history of AAA.

[0004] The underlying causes of aneurysm formation are not known. Investigations into the pathophysiology of AAA have focused on remodeling of the extracellular matrix and inflammation as important mechanisms in AAA formation. (3-8) Family and epidemiologic studies demonstrate a substantial genetic risk for AAA. (4,9)

[0005] AAA is a complex disease with risk influenced by an unknown number of genes that act in combination with environmental factors such as smoking. Identifying specific genetic factors that are associated with AAA would be extremely useful in developing improved methods to diagnose and treat the disease.

[0006] While a number of candidate gene association studies that examined genes selected on the basis of their predicted role in AAA have been reported, (4,9) none of these studies have conclusively identified any AAA-associated variants. Previously, whole genome linkage analyses using family-based or affected-relative-pair methods identified two AAA-associated genetic loci designated as the AAA1 (on 19q13) and AAA2 (4q31) susceptibility loci. (4,10,11) The specific AAA-associated genetic variants in these regions have not yet been identified.

[0007] Another approach to identifying genetic variants associated with complex disease is a genome wide association study (GWAS) in unrelated groups of cases and matched controls, preferably using large numbers of genetic markers. The genetic markers of choice for GWAS are single nucleotide polymorphisms (SNPs). Most SNPs are bi-allelic, stable in the genome, and easily assayed. Several million common SNPs have been identified. In a typical GWAS, the majority of SNP alleles that are assayed have no known functional effects (e.g., they do not alter protein coding information). However, because genetic variants that are physically close in the genome will generally be inherited together (are in linkage disequilibrium, linkage disequilibrium [LD]), SNPs that are nearby a functional genetic variant can serve as markers for the functional variant.

[0008] A previous GWAS identified a genetic variant on chromosome 9 that is associated with a moderate increase in AAA risk. (12) A further description of this

variant can be found in International Publication Number WO2008/102380, which is hereby incorporated herein by reference. However, there remains a need in the art for genetic variants associated with an increase in AAA risk, including panels of variants, which would allow for more accurate characterization of the risk.

[0009] Summary of the Invention

[0010] It is an object of the present invention to provide methods for determining an individual's genetic predisposition, or risk, of developing an abdominal aortic aneurysm (AAA). The methods of the present invention involve isolating a nucleic acid sample from an individual and determining the presence or absence of an allele associated with AAA risk.

[0011] In one aspect of the present invention, a single polymorph of the present invention may be analyzed to determine the individual's risk of developing an AAA. If an individual has at least one allele associated with AAA risk at the polymorphic position analyzed, then the individual is indicated as having a risk of developing an AAA.

[0012] In another aspect of the present invention, more than one polymorph may be analyzed to determine the individual's risk. If more than one polymorph is to be analyzed, the number of alleles associated with AAA risk may be determined. If the total number of alleles associated with AAA risk is at least half of the total number of alleles analyzed, then the individual is indicated as having a risk of developing an AAA.

[0013] Alleles associated with AAA risk in the present invention include a C allele at rs12039875 and a C allele at rs7635818. Other polymorphs in linkage disequilibrium with these alleles may also be used in determining genetic predisposition to an AAA according to the present invention.

[0014] Individuals who are indicated as having a risk of developing an AAA by the methods of the present invention may then be further tested using methods well known in the art for determining the presence of an AAA, such as ultrasound scanning. If no AAA is found to be currently present, those individuals indicated as having a risk of developing an AAA may be subsequently monitored to determine if an AAA develops. Further, individuals indicated as having a risk of developing an AAA may be considered candidates for pharmacotherapeutic agents which may prevent or slow the growth of AAAs.

[0015] It is a still further object of the present invention to provide kits for determining an individual's risk of developing an AAA. The kits of the present invention may include reagents, nucleic acids and probes as necessary for determining the presence or absence of alleles associated with AAA risk.

[0016] Description of the Drawings

[0017] Figure 1: Flow chart for the pooled DNA genome wide association study (GWAS) of Example 1. AF, allele frequency; CV, coefficient of variation.

[0018] Figure 2: Genome wide association identifies a candidate AAA-associated region on chromosome 3p12.3. *Panel A:* The vertical arrow indicates a region on chromosome 3 that contains SNPs with significant allele frequency differences between AAA cases and controls. The known genes in this region are indicated below the line with their gene names and accession numbers. Vertical lines and numbers above the line indicate chromosomal position (in millions of base pairs). *Panel B:* Transformed P values of case-control allele frequency differences corresponding to the region indicated by the horizontal bar in panel A are shown. Each point represents the sum of 5 adjacent -

$\log_{10}$ P values. The physical map is shown for reference. Note that the peak of $-\log_{10}$ P value sums located ~200,000 base pairs upstream of the *CNTN3* gene.

[0019] Figure 3: Genotype analysis and haplotype structure of the AAA-associated region on chromosome 3p12.3. Vertical arrows indicate the physical positions of the 10 SNPs from this region that were individually genotyped in 451 cases and 279 controls from the Geisinger Vascular Clinic (GVC) sample set. The position of the *CNTN3* gene is shown. The inverted triangle is the HapMap R^2 LD plot for the CEU population (gray indicates highest R^2 values). The table shows the reference numbers, map positions and X^2 P values for AAA association of the genotyped SNPs. The significantly associated SNPs are outlined by boxes and lie within a CEU haplotype block characterized by strong LD.

[0020] Figure 4: LD analysis of AAA-associated SNPs. *Top:* $-\log_{10}$ transformed P values for AAA association in the Geisinger Clinic population of the 10 SNPs individually genotyped in the chromosome 3p12.3 region. *Bottom:* LD correlation R values of these SNPs; the most significantly associated SNP and regions of strong LD with that SNPs ($R^2 \geq 0.95$) are indicated in red.

[0021] Figure 5: Odds ratios and P values for AAA association of SNP s7635818. The graph shows the calculated odds ratios (with 95% confidence intervals) for all samples combined, the set of all GVC samples, and the GVC samples from subjects with 20 or more years of smoking history. The total sample size and X^2 P values are also shown. These values were calculated using Helix Tree Software, Version 6.4 assuming a dominant genetic model. The dashed vertical lines indicate an odds ratio of 1.

[0022] Figure 6: Expression of contactin-3 transcripts in aortic tissue. Total RNA isolated from AAA tissue, control (normal) aortic tissue, aortic occlusive disease tissue (AOD) or peripheral blood mononuclear cells from patients with AAA (PBMC) were used as templates for reverse transcriptase and PCR amplification using contactin-3 specific primers. The products were resolved on an agarose gel and stained with ethidium bromide. The expected position of the contactin-3-specific product is indicated by the arrow. Two independent samples of each type were analyzed. Identical results were obtained with a different pair of contactin-3 specific PCR primers.

[0023] Figure 7: Genome wide association identifies a candidate AAA-associated region on chromosome 1q41. *Top panel:* Transformed P values of case-control SNP allele frequency differences on chromosome 1. Each point represents the rolling sum of 5 consecutive $-\log_{10}$ P values. The horizontal bar indicates the candidate region in 1q41. *Lower panel:* Transformed P values of case-control SNP allele frequency differences in a region of ~1.5 million bp corresponding to the horizontal bar in the top panel. The figure above the graph shows known genes in this region.

[0024] Figure 8: AAA-associated haplotype on chromosome 1q41. Vertical arrows indicate the locations of the 6 SNPs from this region that were individually genotyped in a subset of 451 cases and 279 controls from the Geisinger Vascular Clinic (GVC) sample set. The physical map of the *KCNK2* gene is shown. The inverted triangle is the HapMap R^2 LD plot for the CEU population. The table insert shows the reference numbers, map positions and X^2 P values for AAA association of the genotyped SNPs. The significantly associated SNPs are outlined by boxes and overlaps a CEU haplotype block characterized by strong LD.

[0025] Figure 9: AAA-associated haplotype. *Top:* $-\log_{10}$ transformed P values for AAA association in the GVC samples (451 cases, 279 controls) of the 6 SNPs individually genotyped in the chromosome 1q41 region. *Bottom:* LD correlation values of these SNPs in the GVC population. The LD correlation values closely mirror the AAA association P values, with LD and genetic association peaking at SNP rs12039875.

[0026] Figure 10: Expression of *KCNK2* transcripts in aortic tissue. Total RNA was isolated from AAA tissue, control (normal) aortic tissue, and aortic occlusive disease (AOD) tissue. The RNA was copied to cDNA by reverse transcriptase and then used as templates for PCR amplification using primers specific for each of the 3 *KCNK2* isoforms and for *GAPDH*. The PCR products were resolved on an agarose gel and stained with ethidium bromide. The expected positions of the products from the 3 *KCNK2* transcripts and *GAPDH* are indicated. The N-terminal sequences of the 3 *KCNK2* isoforms are shown at the top (isoform a (SEQ ID NO: 8), isoform b (SEQ ID NO: 9) and isoform c (SEQ ID NO: 10). Each lane represents an independent sample. NTC, no template control. The graph shows *KCNK2* expression levels in the tissue samples standardized to *GAPDH* levels. Horizontal bars indicate mean values for each group. The mean value in normal aortic tissue was set to 1.0.

[0027] Figure 11: Additive effect of AAA-associated SNP alleles. AAA cases (938) and controls (1,126) were genotyped for 3 AAA-associated SNPs, rs10757278 (9p21), rs7635818 (3p12.3), and rs12039875 (1q41). *Top panel:* distribution of total AAA-associated alleles for these 3 SNPs in the case (hatched bars) and control (solid bars) samples. The case and control distributions were significantly different by Chi-squared test for trend ($P < 0.0001$). *Middle panel:* Relative odds ratios for AAA disease were

calculated for individuals with 0-2, ≥ 3 , or 6 AAA risk alleles. Error bars indicate 95% confidence intervals. The odds ratio for the 0-2 risk allele group was set = 1. *, $P < 0.001$; **, $P < 0.01$ by Chi-square analysis. *Lower panel:* Linear regression analysis of AAA SNP allele burden vs. AAA disease odds. Filled circles indicate disease odds calculated for each risk allele value (from 0 to 6). The solid line is the regression line (non-zero slope, $P < 0.01$, $R^2 = 0.80$). The dashed lines are the 95% confidence intervals.

[0028] Detailed Description of the Invention

[0029] The present invention provides methods and kits for determining an individual's risk of developing an abdominal aortic aneurysm (AAA). AAAs, which are typically easily treated, often develop undetected as the patient is not aware of the presence of the aneurysm. In certain embodiments, the present invention may be used as a general screen to detect individuals at risk for the development of an AAA, allowing for that individual to be monitored and treated as necessary.

[0030] Currently, AAAs are diagnosed using ultrasound scans to detect dilation of the abdominal aorta. Ultrasound scans are usually only performed on persons who fall into the highest risk groups, namely men over 65 years of age who have ever smoked and men and women with a family history of AAAs. Most guidelines recommend repair of AAAs that are larger than 5.5 cm in men and 5 cm in women, which are considered more likely to rupture. Repair of AAAs is typically a straightforward procedure with minimal risk, while rupture of an AAA is associated with approximately an 80% mortality rate. However, as only the most significant risk groups are ever screened and as there are not other reliable tests to determine AAA risk, a significant number of cases of rupture occur in patients who were unaware that they even had an aneurysm.

[0031] In addition to surgical procedures to repair AAAs, pharmacotherapeutic agents, such as statins have recently been indicated for decreasing the expansion rate of AAAs. Guessous et al., PLoS One, 2008, 3:e 1895. As such, the ability to predict a patient's risk of developing an AAA may allow for therapeutic treatment to be provided before the AAA increases to a dangerous size.

[0032] The present invention provides genetic variants associated with risk of developing an AAA that can be used to determine which patients should be further tested for the presence of an AAA, such as through an ultrasound scan. In certain embodiments of the invention, the methods of the present invention allow for the detection of these genetic variants during routine medical visits, in order to determine whether further testing for an AAA should be undertaken. The methods of the present invention may be incorporated into standard visits to a doctor's office, such as for a physical or check-up. The genetic variants of the present invention may be used as part of a general panel of tests done in these types of appointments. Further, the methods and genetic variants of the present invention may be used in direct to consumer genetic testing procedures.

[0033] A description of preferred embodiments of the invention follows. It should be apparent to those of skill in the art in the field that there are other embodiments, which fall within the scope of the claims set forth below, that are not explicitly described herein.

[0034] Definitions

[0035] Unless otherwise indicated, nucleic acid sequences are written left to right in a 5' to 3' orientation. Numeric ranges recited within the specification are inclusive of the numbers defining the range and include each integer or any non-integer fraction within the defined range. Unless defined otherwise, all technical and scientific terms used herein

have the same meaning as commonly understood by the ordinary person skilled in the art to which the invention pertains.

[0036] The following terms shall, in the present context, have the meaning as indicated:

[0037] An "allele" refers to the nucleotide sequence of a given locus (position) on a chromosome. A polymorphic marker allele thus refers to the composition (i.e., sequence) of the marker on a chromosome. Genomic DNA from an individual contains two alleles (e.g., allele-specific sequences) for any given polymorphic marker, representative of each copy of the marker on each chromosome.

[0038] A nucleotide position at which more than one sequence is possible in a population (either a natural population or a synthetic population, e.g., a library of synthetic molecules) is referred to herein as a "polymorphic site".

[0039] A "Single Nucleotide Polymorphism" or "SNP" is a DNA sequence variation occurring when a single nucleotide at a specific location in the genome differs between members of a species or between paired chromosomes in an individual. Most SNP polymorphisms have two alleles. Each individual is in this instance either homozygous for one allele of the polymorphism (i.e. both chromosomal copies of the individual have the same nucleotide at the SNP location), or the individual is heterozygous (i.e. the two sister chromosomes of the individual contain different nucleotides). The SNP nomenclature as reported herein refers to the official Reference SNP (rs) ID identification tag as assigned to each unique SNP by the National Center for Biotechnological Information (NCBI). See, <http://www.ncbi.nlm.nih.gov/projects/SNP/>.

[0040] A "variant", as described herein, refers to a segment of DNA that differs from the reference DNA. A marker or a polymorphic marker, is a variant. Alleles that differ from the reference are referred to as "variant" alleles.

[0041] A "haplotype," as described herein, refers to a segment of genomic DNA that is characterized by a specific combination of alleles arranged along the segment. For diploid organisms such as humans, a haplotype comprises one member of the pair of alleles for each polymorphic marker or locus . In a certain embodiment, the haplotype can comprise two or more alleles, three or more alleles, four or more alleles, or five or more alleles.

[0042] "HapMap Positions," as described herein refer to enumerated positions on the chromosome as are described by the International HapMap project (www.hapmap.org) and as are well known in the art. Each HapMap position number designates a specific nucleotide position on a human chromosome.

[0043] A "computer-readable medium", is an information storage medium that can be accessed by a computer using a commercially available or custom-made interface.

Exemplary compute- readable media include memory (e.g., RAM, ROM, flash memory, etc.), optical storage media (e.g., CD-ROM), magnetic storage media (e.g., computer hard drives, floppy disks, etc.), punch cards, or other commercially available media.

Information may be transferred between a system of interest and a medium, between computers, or between computers and the computer- readable medium for storage or access of stored information. Such transmission can be electrical, or by other available methods, such as IR links, wireless connections, etc.

[0044] A "nucleic acid sample" is a sample obtained from a individual that contains nucleic acid. In certain embodiments, i.e. the detection of specific polymorphic markers and/or haplotypes, the nucleic acid sample comprises genomic DNA. Such a nucleic acid sample can be obtained from any source that contains genomic DNA, including as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or as a tissue sample from skin, muscle, buccal or conjunctival mucosa, placenta, gastrointestinal tract or other organs.

[0045] Reference may be made to different alleles at a polymorphic site without choosing a reference allele. Alternatively, a reference sequence may be referred to for a particular polymorphic site. The reference allele may be referred to as the "wild-type" allele and it usually is chosen as either the first sequenced allele or as the allele from a "non-affected" individual (e.g., an individual that does not have an increased risk for developing an AAA).

[0046] Alleles for SNP markers as referred to herein refer to the bases A, C, G or T as they occur at the polymorphic site in the SNP assay employed. However it will be apparent to one of skill in the art that by assaying or reading the opposite DNA strand, the complementary allele can in each case be measured. Thus, for a polymorphic site (polymorphic marker) containing an A/G polymorphism, the assay employed may either measure the percentage or ratio of the two bases possible, i.e. A and G. Alternatively, by designing an assay that determines the opposite strand on the DNA template, the percentage or ratio of the complementary bases T/C can be measured.

[0047] Typically, a reference sequence is referred to for a particular sequence. Alleles that differ from the reference are referred to as "variant" alleles. A variant sequence, as

used herein, refers to a sequence that differs from the reference sequence but is otherwise substantially similar. The allele that corresponds to the reference sequence, e.g. the allele that is not associated with an increased risk of AAA, may be referred to as the “major” allele.

[0048] A haplotype refers to a segment of DNA that is characterized by a specific combination of alleles arranged along the segment. For diploid organisms such as humans, a haplotype comprises one member of the pair of alleles for each polymorphic marker or locus. In a certain embodiment, the haplotype can comprise two or more alleles, three or more alleles, four or more alleles, or five or more alleles, each allele corresponding to a specific polymorphic marker along the segment. Haplotypes can comprise a combination of various polymorphic markers, e.g., SNPs and microsatellites, having particular alleles at the polymorphic sites. The haplotypes thus comprise a combination of alleles at various genetic markers. Detecting specific polymorphic markers and/or haplotypes can be accomplished by methods known in the art for detecting sequences at polymorphic sites. For example, standard techniques for genotyping for the presence of SNPs and/or microsatellite markers can be used, such as fluorescence-based techniques (e.g., Chen, X. et al., *Genome Res.* 9(5): 492-98 (1999); Kuttyavin et al., *Nucleic Acid Res.* 34:e128 (2006)), utilizing PCR, LCR, Nested PCR and other techniques for nucleic acid amplification. Specific methodologies available for SNP genotyping include, but are not limited to, TaqMan genotyping assays and SNPlex platforms (Applied Biosystems), mass spectrometry (e.g., MassARRAY system from Sequenom), minisequencing methods, real-time PCR, Bio-Plex system (BioRad), CEQ and SNPstream systems (Beckman), Molecular Inversion Probe array technology (e.g.,

Affymetrix GeneChip), and BeadArray Technologies (e.g., Illumina GoldenGate and Infinium assays). By these or other methods available to the person skilled in the art, one or more alleles at polymorphic markers, including microsatellites, SNPs or other types of polymorphic markers, can be identified.

[0049] Single Nucleotide Polymorphisms

[0050] The present invention provides SNPs which are for the first time associated with risk of developing an AAA, as well as polymorphisms in LD with these SNPs. The primary SNPs of the present invention which are associated with AAA risk are:

[0051] rs7635818 – where C is the allele associated with AAA risk. This SNP is located on chromosome 3p12.3, has HapMap position 74866166, and is represented by SEQ ID NO: 1, where position 27 of SEQ ID NO: 1 is the site of the SNP. This SNP is located upstream of the *CNTN3* gene, which encodes contactin-3.

[0052] rs12039875 – where C is the allele associated with AAA risk. This SNP is located on chromosome 1q41, has HapMap position 213336427, and is represented by SEQ ID NO: 2, where position 27 of SEQ ID NO: 2 is the site of the SNP. This SNP overlies the *KCNK2* gene, which encodes a 2-pore-domain membrane potassium channel.

[0053] SNPs of the present invention associated with AAA risk include those in the same haplotype block as rs7635818 (HapMap Positions 74762904-74998753), as shown in Table 1.

Table 1

SNP Ref.	HapMap Position
rs6549634	74762904
rs12714811	74841440
rs7642555	74852110
rs7635639	74866049

rs7635818	7486616
rs1569066	74881275
rs10222550	74902382
rs1527527	74919636
rs2178032	74943083
rs4417905	74998753

[0054] SNPs of the present invention associated with AAA risk also include those in the same haplotype block as rs12039875 (HapMap Positions 213267336 – 213447659), as shown in Table 2.

Table 2

SNP Ref.	HapMap Position
rs2841599	213267336
rs17024179	213304166
rs12039875	213336427
rs11120508	213383424
rs10494994	213428830
rs1538547	213447659

[0055] In certain methods described herein, an individual who is at an increased susceptibility (i.e., at risk) for developing an AAA is an individual in whom at least one specific allele at one or more polymorphic site or haplotype conferring increased susceptibility for developing an AAA is identified (i.e., at-risk marker alleles or haplotypes). In one aspect, the at-risk marker or haplotype is one that confers a significant increased risk (or susceptibility) of developing an AAA. In one embodiment, the significance associated with a marker or haplotype is measured by a relative risk (RR). In another embodiment, significance associated with a marker or haplotype is measured by an odds ratio (OR). In a further embodiment, the significance is measured by a percentage.

[0056] In other embodiments, a significant increased risk is measured as a risk (relative risk and/or odds ratio) of at least 1.2, including but not limited to: at least 1.2, at least 1.3, at least 1.4, at least 1.5, at least 1.6, at least 1.7, 1.8, at least 1.9, at least 2.0, at least 2.5, at least 3.0, at least 4.0, and at least 5.0. In a particular embodiment, a risk (relative risk and/or odds ratio) of at least 1.2 is significant. In another particular embodiment, a risk of at least 1.3 is significant. In yet another embodiment, a risk of at least 1.4 is significant. In a further embodiment, a relative risk of at least 1.5 is significant. In another further embodiment, a significant increase in risk is at least 1.7 is significant. However, other cutoffs are also contemplated, e.g., at least 1.15, 1.25, 1.35, and so on, and such cutoffs are also within scope of the present invention.

[0057] In other embodiments, a significant increase in risk is at least about 20%, including but not limited to about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 150%, 200%, 300%, and 500%. In one particular embodiment, a significant increase in risk is at least 20%. In other embodiments, a significant increase in risk is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% and at least 100%. Other cutoffs or ranges as deemed suitable by the person skilled in the art to characterize the invention are however also contemplated, and those are also within scope of the present invention.

[0058] In certain embodiments, a significant increase in risk is characterized by a p-value, such as a p-value of less than 0.05, less than 0.01, less than 0.001, less than 0.0001, less than 0.00001, less than 0.000001, less than 0.0000001, less than 0.00000001, or less than 0.000000001.

[0059] An at-risk polymorphic marker or haplotype of the present invention is one where at least one allele of at least one marker or haplotype is more frequently present in an individual at risk for developing an AAA (affected), compared to the frequency of its presence in a healthy individual (control), and wherein the presence of the marker or haplotype is indicative of susceptibility to developing an AAA. The control group may in one embodiment be a population sample, i.e. a random sample from the general population. In another embodiment, the control group is represented by a group of individuals who are determined to not have AAAs. Such AAA free control may in one embodiment be characterized by the absence of one or more specific clinical signs of AAAs. In another embodiment, the AAA free control group is characterized by the absence of one or more AAA specific risk factors, such as age, gender, race and a history of smoking.

[0060] As an example of a simple test for correlation would be a Fisher-exact test on a two by two table. Given a cohort of chromosomes, the two by two table is constructed out of the number of chromosomes that include both of the markers or haplotypes, one of the markers or haplotypes but not the other and neither of the markers or haplotypes. Other statistical tests of association known to the skilled person are also contemplated and are also within scope of the invention.

[0061] The person skilled in the art will appreciate that for markers with two alleles present in the population being studied (such as SNPs), and wherein one allele is found in increased frequency in a group of individuals with a trait or disease in the population, compared with controls, the other allele of the marker will be found in decreased frequency in the group of individuals with the trait or disease, compared with controls. In

such a case, one allele of the marker (the one found in increased frequency in individuals with the trait or disease) will be the allele associated with increased risk of developing an AAA.

[0062] Linkage Disequilibrium

[0063] The natural phenomenon of recombination, which occurs on average once for each chromosomal pair during each meiotic event, represents one way in which nature provides variations in sequence (and biological function by consequence). It has been discovered that recombination does not occur randomly in the genome; rather, there are large variations in the frequency of recombination rates, resulting in small regions of high recombination frequency (also called recombination hotspots) and larger regions of low recombination frequency, which are commonly referred to as Linkage Disequilibrium (LD) blocks (Myers, S. et al., *Biochem Soc Trans* 34:526-530 (2006); Jeffreys, A.J., et al., *Nature Genet* 29:217-222 (2001); May, C.A., et al., *Nature Genet* 31:272-275(2002)).

[0064] Linkage Disequilibrium (LD) refers to a non-random assortment of two genetic elements. For example, if a particular genetic element (e.g., "alleles" of a polymorphic marker) occurs in a population at a frequency of 0.50 (50%) and another occurs at a frequency of 0.50 (50%), then the predicted occurrence of a person's having both elements is 0.25 (25%), assuming a random distribution of the elements. However, if it is discovered that the two elements occur together at a frequency higher than 0.25, then the elements are said to be in linkage disequilibrium, since they tend to be inherited together at a higher rate than what their independent allele frequencies would predict. Roughly speaking, LD is generally correlated with the frequency of recombination events between the two elements. Allele or haplotype frequencies can be determined in a

population by genotyping individuals in a population and determining the frequency of the occurrence of each allele or haplotype in the population. For populations of diploids, e.g., human populations, individuals will typically have two alleles or allelic combinations for each genetic element (e.g., a marker, haplotype or gene).

[0065] Many different measures have been proposed for assessing the strength of linkage disequilibrium (LD). Most capture the strength of association between pairs of biallelic sites. Two important pairwise measures of LD are r^2 (sometimes denoted Δ^2) and $|D'|$. Both measures range from 0 (no disequilibrium) to 1 (complete disequilibrium), but their interpretation is slightly different. $|D'|$ is defined in such a way that it is equal to 1 if just two or three of the possible haplotypes are present, and it is < 1 if all four possible haplotypes are present. Therefore, a value of $|D'|$ that is < 1 indicates that historical recombination may have occurred between two sites (recurrent mutation can also cause $|D'|$ to be < 1 , but for single nucleotide polymorphisms (SNPs) this is usually regarded as being less likely than recombination). The measure r^2 represents the statistical correlation between two sites, and takes the value of 1 if only two haplotypes are present.

[0066] The r^2 measure is arguably the most relevant measure for association mapping, because there is a simple inverse relationship between r^2 and the sample size required to detect association between susceptibility loci and SNPs. These measures are defined for pairs of sites, but for some applications a determination of how strong LD is across an entire region that contains many polymorphic sites might be desirable (e.g., testing whether the strength of LD differs significantly among loci or across populations, or whether there is more or less LD in a region than predicted under a particular model). Measuring LD across a region is not straightforward, but one approach is to use the

measure r , which was developed in population genetics. Roughly speaking, r measures how much recombination would be required under a particular population model to generate the LD that is seen in the data. This type of method can potentially also provide a statistically rigorous approach to the problem of determining whether LD data provide evidence for the presence of recombination hotspots.

[0067] For the methods described herein, a significant r^2 value can be at least 0.1 such as at least 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, or at least 0.99. In one preferred embodiment, the significant r^2 value can be at least 0.2. Alternatively, linkage disequilibrium as described herein, refers to linkage disequilibrium characterized by values of $|D'|$ of at least 0.2, such as 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, or at least 0.99. Thus, linkage disequilibrium represents a correlation between alleles of distinct markers. It is measured by correlation coefficient or $|D'|$ (r^2 up to 1.0 and $|D'|$ up to 1.0). In certain embodiments, linkage disequilibrium is defined in terms of values for both the r^2 and $|D'|$ measures. In one such embodiment, a significant linkage disequilibrium is defined as $r^2 > 0.1$ and $|D'| > 0.8$. In another embodiment, a significant linkage disequilibrium is defined as $r^2 > 0.2$ and $|D'| > 0.9$. Other combinations and permutations of values of r^2 and $|D'|$ for determining linkage disequilibrium are also contemplated, and are also within the scope of the invention. Linkage disequilibrium can be determined in a single human population, as defined herein, or it can be determined in a collection of samples comprising individuals from more than one human population. In one embodiment of the invention, LD is determined in a sample from one or more of the

HapMap populations (Caucasian, African, Japanese, Chinese), as defined (<http://www.hapmap.org>).

[0068] In one such embodiment, LD is determined in the CEU population of the HapMap samples. In another embodiment, LD is determined in the YRI population. In yet another embodiment, LD is determined in samples from the Icelandic population. If all polymorphisms in the genome were independent at the population level (i.e., no LD), then every single one of them would need to be investigated in association studies, to assess all the different polymorphic states. However, due to linkage disequilibrium between polymorphisms, tightly linked polymorphisms are strongly correlated, which reduces the number of polymorphisms that need to be investigated in an association study to observe a significant association. Another consequence of LD is that many polymorphisms may give an association signal due to the fact that these polymorphisms are strongly correlated.

[0069] Genomic LD maps have been generated across the genome, and such LD maps have been proposed to serve as framework for mapping disease-genes (Risch, N. & Merkiangas, K, *Science* 273: 1516-1517 (1996); Maniatis, N., et al., *Proc Natl Acad Sci USA* 99:2228-2233 (2002); Reich, DE et al, *Nature* 411: 199-204 (2001)). It is also now established that many portions of the human genome can be broken into series of discrete haplotype blocks containing a few common haplotypes; for these blocks, linkage disequilibrium data provides little evidence indicating recombination (see, e.g., Wall, J. D. and Pritchard, J. K., *Nature Reviews Genetics* 4:587-597 (2003); Daly, M. et al., *Nature Genet.* 29:229-232 (2001); Gabriel, S. B. et al., *Science* 296:2225-2229 (2002); Patil, N. et al., *Science* 294: 1719-1723 (2001); Dawson, E. et al., *Nature* 418:544-548

(2002); Phillips, M.S. et al., *Nature Genet.* 33:382-387 (2003)). There are two main methods for defining these haplotype blocks: blocks can be defined as regions of DNA that have limited haplotype diversity (see, e.g., Daly, M. et al., *Nature Genet.* 29:229-232 (2001); Patil, N. et al., *Science* 294: 1719-1723 (2001); Dawson, E. et al., *Nature* 418:544-548 (2002); Zhang, K. et al., *Proc. Natl. Acad. Sci. USA* 99:7335-7339 (2002)), or as regions between transition zones having extensive historical recombination, identified using linkage disequilibrium (see, e.g., Gabriel, S. B. et al., *Science* 296:2225-2229 (2002); Phillips, M.S. et al., *Nature Genet.* 33:382-387 (2003); Wang, N. et al., *Am. J. Hum. Genet.* 71 : 1227-1234 (2002); Stumpf, M. P., and Goldstein, D. B., *Curr. Biol.* 13: 1-8 (2003)). More recently, a fine- scale map of recombination rates and corresponding hotspots across the human genome has been generated (Myers, S., et al., *Science* 310:321-323 (2005); Myers, S. et al., *Biochem Soc Trans* 34:526-530 (2006)). The map reveals the enormous variation in recombination across the genome, with recombination rates as high as 10-60 cM/Mb in hotspots, while closer to 0 in intervening regions, which thus represent regions of limited haplotype diversity and high LD. The map can therefore be used to define haplotype blocks/LD blocks as regions flanked by recombination hotspots. As used herein, the terms "haplotype block" or "LD block" includes blocks defined by any of the above described characteristics, or other alternative methods used by the person skilled in the art to define such regions. Haplotype blocks (LD blocks) can be used to map associations between phenotype and haplotype status, using single markers or haplotypes comprising a plurality of markers. The main haplotypes can be identified in each haplotype block, and then a set of "tagging" SNPs or markers (the smallest set of SNPs or markers needed to distinguish among the

haplotypes) can then be identified. These tagging SNPs or markers can then be used in assessment of samples from groups of individuals, in order to identify association between phenotype and haplotype. If desired, neighboring haplotype blocks can be assessed concurrently, as there may also exist linkage disequilibrium among the haplotype blocks.

[0070] It has thus become apparent that for any given observed association to a polymorphic marker in the genome, it is likely that additional markers in the genome also show association. This is a natural consequence of the uneven distribution of LD across the genome, as observed by the large variation in recombination rates. The markers used to detect association thus in a sense represent "tags" for a genomic region (i.e., a haplotype block or LD block) that is associating with a given disease or trait, and as such are useful for use in the methods and kits of the present invention. One or more causative (functional) variants or mutations may reside within the region found to be associating to the disease or trait. Such variants may confer a higher relative risk (RR) or odds ratio (OR) than observed for the tagging markers used to detect the association. The present invention thus refers to the markers used for detecting association to the disease, as described herein, as well as markers in linkage disequilibrium with the markers. Thus, in certain embodiments of the invention, markers that are in LD with the markers and/or haplotypes of the invention, as described herein, may be used as surrogate markers. The surrogate markers have in one embodiment relative risk (RR) and/or odds ratio (OR) values smaller than for the markers or haplotypes initially found to be associating with the disease, as described herein. In other embodiments, the surrogate markers have RR or OR values greater than those initially determined for the markers initially found to be

associating with the disease, as described herein. An example of such an embodiment would be a rare, or relatively rare (such as $< 10\%$ allelic population frequency) variant in LD with a more common variant ($> 10\%$ population frequency) initially found to be associating with the disease, such as the variants described herein. Identifying and using such markers for detecting the association discovered by the inventors as described herein can be performed by routine methods well known to the person skilled in the art, and are therefore within the scope of the present invention.

[0071] Determination of haplotype frequency

[0072] The frequencies of haplotypes in patient and control groups can be estimated using an expectation-maximization algorithm (Dempster A. et al., J. R. Stat. Soc. B, 39: 1-38 (1977)). An implementation of this algorithm that can handle missing genotypes and uncertainty with the phase can be used. Under the null hypothesis, the patients and the controls are assumed to have identical frequencies. Using a likelihood approach, an alternative hypothesis is tested, where a candidate at-risk-haplotype, which can include the markers described herein, is allowed to have a higher frequency in patients than controls, while the ratios of the frequencies of other haplotypes are assumed to be the same in both groups. Likelihoods are maximized separately under both hypotheses and a corresponding 1-df likelihood ratio statistic is used to evaluate the statistical significance.

[0073] To look for at-risk and protective markers and haplotypes within a linkage region, for example, association of all possible combinations of genotyped markers is studied, provided those markers span a practical region. The combined patient and control groups can be randomly divided into two sets, equal in size to the original group of patients and controls. The marker and haplotype analysis is then repeated and the most

significant p-value registered is determined. This randomization scheme can be repeated, for example, over 100 times to construct an empirical distribution of p-values. In a preferred embodiment, a p-value of <0.05 is indicative of an significant marker and/or haplotype association.

[0074] Haplotype Analysis

[0075] One general approach to haplotype analysis involves using likelihood-based inference applied to NEsted MOdels (Gretarsdottir S., et al., Nat. Genet. 35: 131-38 (2003)). The method is implemented in the program NEMO, which allows for many polymorphic markers, SNPs and microsatellites. The method and software are specifically designed for case-control studies where the purpose is to identify haplotype groups that confer different risks. It is also a tool for studying LD structures. In NEMO, maximum likelihood estimates, likelihood ratios and p-values are calculated directly, with the aid of the EM algorithm, for the observed data treating it as a missing-data problem.

[0076] Even though likelihood ratio tests based on likelihoods computed directly for the observed data, which have captured the information loss due to uncertainty in phase and missing genotypes, can be relied on to give valid p-values, it would still be of interest to know how much information had been lost due to the information being incomplete. The information measure for haplotype analysis is described in Nicolae and Kong (Technical Report 537, Department of Statistics, University of Statistics, University of Chicago; Biometrics, 60(2):368-75 (2004)) as a natural extension of information measures defined for linkage analysis, and is implemented in NEMO.

[0077] For single marker association to a disease, the Fisher exact test can be used to calculate two-sided p-values for each individual allele. Usually, all p-values are presented unadjusted for multiple comparisons unless specifically indicated. The presented frequencies (for microsatellites, SNPs and haplotypes) are allelic frequencies as opposed to carrier frequencies. To minimize any bias due the relatedness of the patients who were recruited as families for the linkage analysis, first and second-degree relatives can be eliminated from the patient list. Furthermore, the test can be repeated for association correcting for any remaining relatedness among the patients, by extending a variance adjustment procedure described in Risch, N. & Teng, J. (Genome Res., 8: 1273-1288 (1998)), DNA pooling (ibid) for sibships so that it can be applied to general familial relationships, and present both adjusted and unadjusted p-values for comparison. The differences are in general very small as expected. To assess the significance of single-marker association corrected for multiple testing we can carry out a randomization test using the same genotype data. Cohorts of patients and controls can be randomized and the association analysis redone multiple times (e.g., up to 500,000 times) and the p-value is the fraction of replications that produced a p-value for some marker allele that is lower than or equal to the p-value we observed using the original patient and control cohorts. For both single-marker and haplotype analyses, relative risk (RR) and the population attributable risk (PAR) can be calculated assuming a multiplicative model (haplotype relative risk model) (Terwilliger, J. D. & Ott, J., Hum. Hered. 42:337-46 (1992) and Falk, CT. & Rubinstein, P, Ann. Hum. Genet. 51 (Pt 3)\227-33 (1987)), i.e., that the risks of the two alleles/haplotypes a person carries multiply. For example, if RR is the risk of A relative to a, then the risk of a person homozygote AA will be RR times that of a

heterozygote Aa and RR^2 times that of a homozygote aa. The multiplicative model has a nice property that simplifies analysis and computations — haplotypes are independent, i.e., in Hardy-Weinberg equilibrium, within the affected population as well as within the control population. As a consequence, haplotype counts of the affected and controls each have multinomial distributions, but with different haplotype frequencies under the alternative hypothesis. Specifically, for two haplotypes, h_1 and h_2 , $\text{risk}(h_1)/\text{risk}(h_2) = WP_1/(f_j/P_j)$, where f and p denote, respectively, frequencies in the affected population and in the control population. While there is some power loss if the true model is not multiplicative, the loss tends to be mild except for extreme cases. Most importantly, p -values are always valid since they are computed with respect to null hypothesis.

[0078] Linkage Disequilibrium Using NEMO

[0079] LD between pairs of markers can be calculated using the standard definition of D' and r^2 (Lewontin, R., Genetics 49:49-67 (1964); Hill, W. G. & Robertson, A. Theor. Appl. Genet. 22:226- 231 (1968)). Using NEMO, frequencies of the two marker allele combinations are estimated by maximum likelihood and deviation from linkage equilibrium is evaluated by a likelihood ratio test. The definitions of D' and r^2 are extended to include microsatellites by averaging over the values for all possible allele combination of the two markers weighted by the marginal allele probabilities.

[0080] Risk assessment and Diagnostics

[0081] Within any given population, there is an absolute risk of developing a disease or trait, defined as the chance of a person developing the specific disease or trait over a specified time- period. For example, a woman's lifetime absolute risk of breast cancer is one in nine. That is to say, one woman in every nine will develop breast cancer at some

point in their lives. Risk is typically measured by looking at very large numbers of people, rather than at a particular individual. Risk is often presented in terms of Absolute Risk (AR) and Relative Risk (RR). Relative Risk is used to compare risks associating with two variants or the risks of two different groups of people. For example, it can be used to compare a group of people with a certain genotype with another group having a different genotype. For a disease, a relative risk of 2 means that one group has twice the chance of developing a disease as the other group. The Risk presented is usually the relative risk for a person, or a specific genotype of a person, compared to the population with matched gender and ethnicity. Risks of two individuals of the same gender and ethnicity could be compared in a simple manner. For example, if, compared to the population, the first individual has relative risk 1.5 and the second has relative risk 0.5, then the risk of the first individual compared to the second individual is $1.5/0.5 = 3$.

[0082] As described herein, certain polymorphic markers and haplotypes comprising such markers are found to be useful for risk assessment of AAA. Risk assessment can involve the use of the markers for diagnosing a susceptibility to developing an AAA. Particular alleles of polymorphic markers are found more frequently in individuals with an AAA, compared with individuals without diagnosis of an AAA. Therefore, these marker alleles have predictive value for determining AAA risk. Tagging markers within haplotype blocks or LD blocks comprising at-risk markers, such as the markers of the present invention, can be used as surrogates for other markers and/or haplotypes within the haplotype block or LD block. Markers with values of r^2 equal to 1 are perfect surrogates for the at-risk variants, i.e. genotypes for one marker perfectly predicts genotypes for the other. Markers with smaller values of r^2 than 1 can also be surrogates

for the at-risk variant, or alternatively represent variants with relative risk values as high as or possibly even higher than the at-risk variant. The at-risk variant identified may not be the functional variant itself, but is in this instance in linkage disequilibrium with the true functional variant. The present invention encompasses the assessment of such surrogate markers for the markers as disclosed herein. Such markers are annotated, mapped and listed in public databases, as well known to the skilled person, or can alternatively be readily identified by sequencing the region or a part of the region identified by the markers of the present invention in a group of individuals, and identify polymorphisms in the resulting group of sequences. As a consequence, the person skilled in the art can readily and without undue experimentation genotype surrogate markers in linkage disequilibrium with the markers and/or haplotypes as described herein. The tagging or surrogate markers in LD with the at-risk variants detected, also have predictive value for determining risk of developing AAAs. These tagging or surrogate markers that are in LD with the markers of the present invention can also include other markers that distinguish among haplotypes, as these similarly have predictive value for detecting susceptibility to developing an AAA.

[0083] The present invention can in certain embodiments be practiced by assessing a sample comprising genomic DNA from an individual for the presence of variants described herein to be associated with AAA risk. Such assessment includes steps of detecting the presence or absence of at least one allele of at least one polymorphic marker, using methods well known to the skilled person and further described herein, and based on the outcome of such assessment, determine whether the individual from whom the sample is derived is at increased or decreased risk (increased or decreased susceptibility)

of developing an AAA. Alternatively, the invention can be practiced utilizing a dataset comprising information about the genotype status of at least one polymorphic marker described herein to be associated with AAA (or markers in linkage disequilibrium with at least one marker shown herein to be associated with AAA). In other words, a dataset containing information about such genetic status, for example in the form of genotype counts at a certain polymorphic marker, or a plurality of markers (e.g., an indication of the presence or absence of certain at-risk alleles), or actual genotypes for one or more markers, can be queried for the presence or absence of certain at-risk alleles at certain polymorphic markers shown by the present inventors to be associated with AAA. A positive result for a variant (e.g., marker allele) associated with AAA, as shown herein, is indicative of the individual from which the dataset is derived is at increased risk of developing AAA.

[0084] As described in the above, the haplotype block structure of the human genome has the effect that a large number of variants (markers and/or haplotypes) in linkage disequilibrium with the variant originally associated with a disease or trait may be used as surrogate markers for assessing association to the disease or trait. The number of such surrogate markers will depend on factors such as the historical recombination rate in the region, the mutational frequency in the region (i.e., the number of polymorphic sites or markers in the region), and the extent of LD (size of the LD block) in the region. These markers are usually located within the physical boundaries of the LD block or haplotype block in question as defined using the methods described herein, or by other methods known to the person skilled in the art. However, sometimes marker and haplotype association is found to extend beyond the physical boundaries of the haplotype block as

defined. Such markers and/or haplotypes may in those cases be also used as surrogate markers and/or haplotypes for the markers and/or haplotypes physically residing within the haplotype block as defined. As a consequence, markers and haplotypes in LD (typically characterized by r^2 greater than 0.1, such as r^2 greater than 0.2, including r^2 greater than 0.3, also including r^2 greater than 0.4) with the markers and haplotypes of the present invention are also within the scope of the invention, even if they are physically located beyond the boundaries of the haplotype block as defined. This may include markers that are in strong LD (e.g., characterized by r^2 greater than 0.1 or 0.2 and/or $|D'| > 0.8$) with one or more of the markers of the present invention.

[0085] Study population

[0086] In a general sense, the methods and kits of the invention can be utilized from samples containing nucleic acid material (DNA or RNA) from any source and from any individual. In preferred embodiments, the individual is a human individual. The individual can be an adult, child, or fetus. The nucleic acid source may be any sample comprising nucleic acid material, including biological samples, or a sample comprising nucleic acid material derived therefrom. The present invention also provides for assessing markers and/or haplotypes in individuals who are members of a target population. Such a target population is in one embodiment a population or group of individuals at risk of developing an AAA, such as those having the most common risk factors, old age, male gender, smoking, and a family history of AAA.

[0087] Methods

[0088] Methods for risk assessment of AAA are described herein and are encompassed by the invention. The invention also encompasses methods of assessing an

individual for probability of response to a therapeutic agent for treating AAAs, methods for predicting the effectiveness of a therapeutic agent for treating AAAs, nucleic acids and computer-implemented functions. Kits for assaying a sample from a individual to detect susceptibility to developing an AAA are also encompassed by the invention.

[0089] In embodiments of the present invention, methods are provided for determining the presence or absence of a variant allele at certain polymorphic sites and correlating the presence of a variant allele with increased risk of developing an AAA. In the methods of the present invention, a nucleic acid sample, as described above, may be taken from an individual to be tested. After the nucleic acid sample comprising the SNP to be analyzed is sufficiently purified as is well known in the art, the sample may be analyzed to determine the nucleotide present at the SNP position of interest. This determination may be done by standard sequencing methods well known in the art, as well as other methods, such as those described below.

[0090] In certain methods of the present invention, the individual is analyzed to determine the nucleotide at SNP rs7635818 (position 27 of SEQ ID NO: 1). If a C allele is found to be present at rs7635818, then the individual may be considered as having an increased risk of developing an AAA, and may be recommended for ultrasound scan and subsequent follow up to determine if an AAA is present or may develop. The presence of one C allele at rs7635818 of the individual may be enough to indicate an ultrasound scan, with the presence of two C alleles at this position indicating an even greater risk.

[0091] In these embodiments of the present invention, the odds ratio for association of the presence of a C allele at rs7635818 with the risk of developing an AAA in a general sample set may be considered as having an odds ratio of 1.33. For individuals

who are smokers, the presence of a C allele at rs7635818 may be considered as having an odds ratio of 1.8.

[0092] In certain methods of the present invention, the individual is analyzed to determine the nucleotide at SNP rs12039875 (position 27 of SEQ ID NO: 2). If a C allele is found to be present at rs12039875, then the individual may be considered as having an increased risk of developing an AAA, and may be recommended for an ultrasound scan and subsequent follow up to determine if an AAA is present or may develop. The presence of one C allele at rs12039875 of the individual may be enough to indicate an ultrasound scan, with the presence of two C alleles at this position indicating an even greater risk.

[0093] In these embodiments of the present invention, the odds ratio for association of the presence of a C allele at rs12039875 with the risk of developing an AAA in a general sample set may be considered as having an odds ratio of 1.18.

[0094] In other embodiments of the present invention, more than one SNP may be analyzed to determine AAA risk. For example, the nucleotides at both rs7635818 and rs12039875 may be analyzed to give a total of four alleles (two for each SNP). The total number of variant alleles, the presence of a C for both rs7635818 and rs12039875, can then be used to determine the risk of developing an AAA, with the presence of 2 or more variant alleles indicating that the individual should have an ultrasound scan and subsequent follow up to determine if an AAA is present or may develop. The presence of 3 or 4 alleles indicates an even greater risk of developing an AAA, and the individual should also be monitored for development of an AAA as described.

[0095] In still further embodiments of the present invention, a panel of three SNPs for determining the risk of developing an AAA is provided. The panel of three SNPs may be made up of rs7635818, rs12039875 and rs10757278, as is described in International Publication Number WO2008/102380. The G allele is associated with AAA risk for rs10757278. This panel of three SNPs provides six total alleles. The presence of three or more variant alleles indicates that the individual should have an ultrasound scan and subsequent follow up to determine if an AAA is present or may develop. The presence of 4, 5 or 6 alleles indicates an even greater risk of developing an AAA, and the individual should also be monitored for development of an AAA as described.

[0096] In these embodiments of the present invention, the odds ratio for association of the presence of three or more variant alleles with the risk of developing an AAA may be considered as having an odds ratio of 2.4, while the presence of six variant alleles may be considered as having an odds ratio of 3.5.

[0097] It is also contemplated that analysis of other polymorphisms in linkage disequilibrium with rs7635818, rs12039875 and rs10757278 may be done in combination with one, two or all of these three SNPs. Examples of polymorphisms that may be used include those in Tables 1 and 2 above, as well as other polymorphisms in the same linkage disequilibrium block as the SNPs in Tables 1 and 2. The presence of variant alleles at these other polymorphisms suggests a further increased risk of developing an AAA.

[0098] It is still further contemplated that the analysis of the polymorphisms of the present invention may be combined with other risk factors to further refine the risk of

developing an AAA. This risk factors include old age, male gender, history of smoking, and family history of AAAs.

[0099] In embodiments of the invention, the significance of association of the at least one marker allele or haplotype is characterized by a p value < 0.05 . In other embodiments, the significance of association is characterized by smaller p-values, such as < 0.01 , < 0.001 , < 0.0001 , < 0.00001 , < 0.000001 , < 0.0000001 or < 0.000000001 .

[00100] In these embodiments, the presence of the at least one marker allele or haplotype is indicative of a susceptibility to developing an AAA. These diagnostic methods involve detecting the presence or absence of at least one marker allele or haplotype that is associated with AAA risk. The haplotypes described herein include combinations of alleles at various genetic markers (e.g., SNPs, microsatellites). The detection of the particular genetic marker alleles that make up the particular haplotypes can be performed by a variety of methods described herein and/or known in the art. The marker alleles or haplotypes of the present invention correspond to fragments of a genomic DNA sequence associated with development of AAAs. Such fragments encompass the DNA sequence of the polymorphic marker or haplotype in question, but may also include DNA segments in strong LD (linkage disequilibrium) with the marker or haplotype (e.g., as determined by particular values of r^2 and/or $|D'|$).

[00101] In one embodiment, diagnosis of a susceptibility to developing an AAA can be accomplished using hybridization methods, including, but not limited to, Southern analysis, Northern analysis, and/or In situ hybridizations (see Current Protocols in Molecular Biology, Ausubel, F. et al., eds., John Wiley & Sons, including all supplements). The presence of a specific marker allele can be indicated by sequence-

specific hybridization of a nucleic acid probe specific for the particular allele. The presence of more than one specific marker allele or a specific haplotype can be indicated by using several sequence-specific nucleic acid probes, each being specific for a particular allele. In one embodiment, a haplotype can be indicated by a single nucleic acid probe that is specific for the specific haplotype (i.e., hybridizes specifically to a DNA strand comprising the specific marker alleles characteristic of the haplotype). A sequence-specific probe can be directed to hybridize to genomic DNA, RNA, or cDNA.

[00102] A "nucleic acid probe", as used herein, can be a DNA probe or an RNA probe that hybridizes to a complementary sequence. One of skill in the art would know how to design such a probe so that sequence specific hybridization will occur only if a particular allele is present in a genomic sequence from a test sample.

[00103] In a further embodiment, a method utilizing a detection oligonucleotide probe comprising a fluorescent moiety or group at its 3' terminus and a quencher at its 5' terminus, and an enhancer oligonucleotide, is employed, as described by Kuttyavin et al. (Nucleic Acid Res. 34:e128 (2006)). The fluorescent moiety can be Gig Harbor Green or Yakima Yellow, or other suitable fluorescent moieties. The detection probe is designed to hybridize to a short nucleotide sequence that includes the SNP polymorphism to be detected. Preferably, the SNP is anywhere from the terminal residue to -6 residues from the 3' end of the detection probe. The enhancer is a short oligonucleotide probe which hybridizes to the DNA template 3' relative to the detection probe. The probes are designed such that a single nucleotide gap exists between the detection probe and the enhancer nucleotide probe when both are bound to the template. The gap creates a synthetic abasic site that is recognized by an endonuclease, such as Endonuclease IV. The

enzyme cleaves the dye off the fully complementary detection probe, but cannot cleave a detection probe containing a mismatch. Thus, by measuring the fluorescence of the released fluorescent moiety, assessment of the presence of a particular allele defined by nucleotide sequence of the detection probe can be performed.

[00104] The detection probe can be of any suitable size, although preferably the probe is relatively short. In one embodiment, the probe is from 5-100 nucleotides in length. In another embodiment, the probe is from 10-50 nucleotides in length, and in another embodiment, the probe is from 12-30 nucleotides in length. Other lengths of the probe are possible and within scope of the skill of the average person skilled in the art.

[00105] In a preferred embodiment, the DNA template containing the SNP polymorphism is amplified by Polymerase Chain Reaction (PCR) prior to detection. In such an embodiment, the amplified DNA serves as the template for the detection probe and the enhancer probe.

[00106] Certain embodiments of the detection probe, the enhancer probe, and/or the primers used for amplification of the template by PCR include the use of modified bases, including modified A and modified G. The use of modified bases can be useful for adjusting the melting temperature of the nucleotide molecule (probe and/or primer) to the template DNA, for example for increasing the melting temperature in regions containing a low percentage of G or C bases, in which modified A with the capability of forming three hydrogen bonds to its complementary T can be used, or for decreasing the melting temperature in regions containing a high percentage of G or C bases, for example by using modified G bases that form only two hydrogen bonds to their complementary C base in a double stranded DNA molecule. In a preferred embodiment, modified bases are

used in the design of the detection nucleotide probe. Any modified base known to the skilled person can be selected in these methods, and the selection of suitable bases is well within the scope of the skilled person based on the teachings herein and known bases available from commercial sources as known to the skilled person.

[00107] In another hybridization method, Northern analysis (see Current Protocols in Molecular Biology, Ausubel, F. et al., eds., John Wiley & Sons, supra) is used to identify the presence of a polymorphism associated with AAA risk. For Northern analysis, a test sample of RNA is obtained from the individual by appropriate means. As described herein, specific hybridization of a nucleic acid probe to RNA from the individual is indicative of a particular allele complementary to the probe. For representative examples of use of nucleic acid probes, see, for example, U.S. Patent Nos. 5,288,611 and 4,851,330.

[00108] Additionally, or alternatively, a peptide nucleic acid (PNA) probe can be used in addition to, or instead of, a nucleic acid probe in the hybridization methods described herein. A PNA is a DNA mimic having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen, P., et al., Bioconjug. Chem. 5:3-7 (1994)). The PNA probe can be designed to specifically hybridize to a molecule in a sample suspected of containing one or more of the marker alleles or haplotypes that are associated with risk of developing an AAA. Hybridization of the PNA probe is thus diagnostic for AAA risk.

[00109] In one embodiment of the invention, a test sample containing genomic DNA obtained from the individual is collected and the polymerase chain reaction (PCR) is used to amplify a fragment comprising one or more markers or haplotypes of the present

invention. As described herein, identification of a particular marker allele or haplotype associated with AAA risk, can be accomplished using a variety of methods (e.g., sequence analysis, analysis by restriction digestion, specific hybridization, single stranded conformation polymorphism assays (SSCP), electrophoretic analysis, etc.). In another embodiment, diagnosis is accomplished by expression analysis using quantitative PCR (kinetic thermal cycling). This technique can, for example, utilize commercially available technologies, such as TaqMan[®] (Applied Biosystems, Foster City, CA). The technique can assess the presence of an alteration in the expression or composition of a polypeptide or splicing variant(s) that is encoded by a nucleic acid associated with an increased risk of AAA. Further, the expression of the variant(s) can be quantified as physically or functionally different.

[00110] In another embodiment of the methods of the invention, analysis by restriction digestion can be used to detect a particular allele if the allele results in the creation or elimination of a restriction site relative to a reference sequence. A test sample containing genomic DNA is obtained from the individual. PCR can be used to amplify particular regions that are associated with risk of AAA development (e.g. the polymorphic markers recited herein) nucleic acid in the test sample from the test individual. Restriction fragment length polymorphism (RFLP) analysis can be conducted, e.g., as described in Current Protocols in Molecular Biology, supra. The digestion pattern of the relevant DNA fragment indicates the presence or absence of the particular allele in the sample.

[00111] Sequence analysis can also be used to detect specific alleles at polymorphic sites associated with risk of AAA development. Therefore, in one embodiment, determination of the presence or absence of a particular marker alleles or haplotypes

comprises sequence analysis. For example, a test sample of DNA or RNA can be obtained from the test individual. PCR or other appropriate methods can be used to amplify a portion of a nucleic acid associated with AAA risk, and the presence of a specific allele can then be detected directly by sequencing the polymorphic site (or multiple polymorphic sites) of the genomic DNA in the sample. In another embodiment, arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from a individual, can be used to identify polymorphisms in a nucleic acid associated with AAA risk. For example, an oligonucleotide array can be used.

Oligonucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. These arrays can generally be produced using mechanical synthesis methods or light directed synthesis methods that incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis methods, or by other methods known to the person skilled in the art (see, e.g., Fodor, S. et al., *Science*, 251 :767-773 (1991); Pirrung et al., U.S. Patent No. 5,143,854 (see also published PCT Application No. WO 90/15070); and Fodor, S. et al., published PCT Application No. WO 92/10092 and U.S. Patent No. 5,424,186, the entire teachings of each of which are incorporated by reference herein). Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, e.g., U.S. Patent No. 5,384,261; the entire teachings of which are incorporated by reference herein.

[00112] In another example, linear arrays can be utilized. Additional descriptions of use of oligonucleotide arrays for detection of polymorphisms can be found, for example,

in U.S. Patent Nos. 5,858,659 and 5,837,832, the entire teachings of both of which are incorporated by reference herein.

[00113] Other methods of nucleic acid analysis that are available to those skilled in the art can be used to detect a particular allele at a polymorphic site associated with risk of developing an AAA. Representative methods include, for example, direct manual sequencing (Church and Gilbert, Proc. Natl. Acad. Sci. USA, 81 : 1991-1995 (1988); Sanger, F., et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467 (1977); Beavis, et al., U.S. Patent No. 5,288,644); automated fluorescent sequencing; single-stranded conformation polymorphism assays (SSCP); clamped denaturing gel electrophoresis (CDGE); denaturing gradient gel electrophoresis (DGGE) (Sheffield, V., et al., Proc. Natl. Acad. Sci. USA, 86:232-236 (1989)), mobility shift analysis (Orita, M., et al., Proc. Natl. Acad. Sci. USA, 86:2766-2770 (1989)), restriction enzyme analysis (Flavell, R., et al., Cell, 15:25-41 (1978); Geever, R., et al., Proc. Natl. Acad. Sci. USA, 78:5081-5085 (1981)); heteroduplex analysis; chemical mismatch cleavage (CMC) (Cotton, R., et al., Proc. Natl. Acad. Sci. USA, 85:4397-4401 (1985)); RNase protection assays (Myers, R., et al., Science, 230: 1242- 1246 (1985); use of polypeptides that recognize nucleotide mismatches, such as E. coli mutS protein; and allele-specific PCR.

[00114] Thus, in another embodiment, the variants (markers or haplotypes) of the invention showing association to cardiovascular disease affect the expression of a nearby gene (e.g., *CNTN3* for rs7635818 or *KCNK2* for rs12039875). It is well known that regulatory element affecting gene expression may be located far away, even as far as tenths or hundreds of kilobases away, from the promoter region of a gene. By assaying for the presence or absence of at least one allele of at least one polymorphic marker of the

present invention, it is thus possible to assess the expression level of such nearby genes.

It is thus contemplated that the detection of the markers or haplotypes of the present invention can be used for assessing expression for one or more of the *CNTN3* and/or *KCNK2* genes. A variety of methods can be used for detecting protein expression levels, including enzyme linked immunosorbent assays (ELISA), Western blots, immunoprecipitations and immunofluorescence.

[00115] Kits

[00116] Kits useful in the methods of the invention comprise components useful in any of the methods described herein, including for example, primers for nucleic acid amplification, hybridization probes, restriction enzymes (e.g., for RFLP analysis), allele-specific oligonucleotides, a genomic segment comprising at least one polymorphic marker and/or haplotype of the present invention) or to a non-altered (native) polypeptide encoded by a nucleic acid associated with the risk of developing an AAA, means for amplification of a nucleic acid associated with AAA risk, means for analyzing the nucleic acid sequence of a nucleic acid associated with AAA risk, means for analyzing the amino acid sequence of a polypeptide encoded by a nucleic acid associated with AAA risk, etc. The kits can for example include necessary buffers, nucleic acid primers for amplifying nucleic acids of the invention (e.g., a nucleic acid segment comprising one or more of the polymorphic markers as described herein), and reagents for allele-specific detection of the fragments amplified using such primers and necessary enzymes (e.g., DNA polymerase). Additionally, kits can provide reagents for assays to be used in combination with the methods of the present invention, e.g., reagents for use with other diagnostic assays as described herein.

[00117] In one embodiment, the invention pertains to a kit for assaying a sample from a individual to detect the risk of developing an AAA, wherein the kit comprises reagents necessary for selectively detecting at least one allele of at least one polymorphism of the present invention in the genome of the individual. In a particular embodiment, the reagents comprise at least one contiguous oligonucleotide that hybridizes to a fragment of the genome of the individual comprising at least one polymorphism of the present invention. In another embodiment, the reagents comprise at least one pair of oligonucleotides that hybridize to opposite strands of a genomic segment obtained from a individual, wherein each oligonucleotide primer pair is designed to selectively amplify a fragment of the genome of the individual that includes one polymorphism, wherein the polymorphism is selected from the group consisting of the polymorphisms listed above, and polymorphic markers in linkage disequilibrium therewith. In yet another embodiment the fragment is at least 20 base pairs in size. Such oligonucleotides or nucleic acids (e.g., oligonucleotide primers) can be designed using portions of the nucleic acids flanking polymorphisms (e.g., SNPs or microsatellites) that are indicative of a risk of developing an AAA. In another embodiment, the kit comprises one or more labeled nucleic acids capable of detecting one or more specific polymorphic markers or haplotypes associated with AAA risk, and reagents for detection of the label. Suitable labels include, e.g., a radioisotope, a fluorescent label, an enzyme label, an enzyme co-factor label, a magnetic label, a spin label, an epitope label.

[00118] In particular embodiments, the polymorphic marker or haplotype to be detected by the reagents of the kit comprises one or more markers, two or more markers, or three or more markers, selected from the group consisting of the rs7635818,

rs12039875 and rs10757278, and markers in linkage disequilibrium therewith. The fluorescent moiety can be Gig Harbor Green or Yakima Yellow, or other suitable fluorescent moieties. The detection probe is designed to hybridize to a short nucleotide sequence that includes the SNP polymorphism to be detected. Preferably, the SNP is anywhere from the terminal residue to -6 residues from the 3' end of the detection probe. The enhancer is a short oligonucleotide probe which hybridizes to the DNA template 3' relative to the detection probe. The probes are designed such that a single nucleotide gap exists between the detection probe and the enhancer nucleotide probe when both are bound to the template. The gap creates a synthetic abasic site that is recognized by an endonuclease, such as Endonuclease IV. The enzyme cleaves the dye off the fully complementary detection probe, but cannot cleave a detection probe containing a mismatch. Thus, by measuring the fluorescence of the released fluorescent moiety, assessment of the presence of a particular allele defined by nucleotide sequence of the detection probe can be performed.

[00119] The detection probe can be of any suitable size, although preferably the probe is relatively short. In one embodiment, the probe is from 5-100 nucleotides in length. In another embodiment, the probe is from 10-50 nucleotides in length, and in another embodiment, the probe is from 12-30 nucleotides in length. Other lengths of the probe are possible and within scope of the skill of the average person skilled in the art.

[00120] In a preferred embodiment, the DNA template containing the SNP polymorphism is amplified by Polymerase Chain Reaction (PCR) prior to detection, and primers for such amplification are included in the reagent kit. In such an embodiment, the amplified DNA serves as the template for the detection probe and the enhancer probe.

[00121] Certain embodiments of the detection probe, the enhancer probe, and/or the primers used for amplification of the template by PCR include the use of modified bases, including modified A and modified G. The use of modified bases can be useful for adjusting the melting temperature of the nucleotide molecule (probe and/or primer) to the template DNA, for example for increasing the melting temperature in regions containing a low percentage of G or C bases, in which modified A with the capability of forming three hydrogen bonds to its complementary T can be used, or for decreasing the melting temperature in regions containing a high percentage of G or C bases, for example by using modified G bases that form only two hydrogen bonds to their complementary C base in a double stranded DNA molecule. In a preferred embodiment, modified bases are used in the design of the detection nucleotide probe. Any modified base known to the skilled person can be selected in these methods, and the selection of suitable bases is well within the scope of the skilled person based on the teachings herein and known bases available from commercial sources as known to the skilled person.

[00122] In one such embodiment, the presence of the marker or haplotype is indicative of an increased risk of developing an AAA. In another embodiment, the presence of the marker or haplotype is indicative of response to a therapeutic agent for treatment of an AAA. In another embodiment, the presence of the marker or haplotype is indicative of prognosis of an AAA. In still further embodiments, the presence of the marker or haplotype may indicate that the AAA should be repaired even if it is smaller than aneurysms that are typically repaired, e.g. it is smaller than 5.5 cm for men and 5 cm for women.

[00123] In certain embodiments, the kit further comprises a set of instructions for using the reagents comprising the kit.

[00124] Computer-implemented aspects

[00125] The present invention also relates to computer-implemented applications of the polymorphic markers and haplotypes described herein to be associated with AAA risk. Such applications can be useful for storing, manipulating or otherwise analyzing genotype data that is useful in the methods of the invention. One example pertains to storing genotype information derived from an individual on readable media, so as to be able to provide the genotype information to a third party (e.g., the individual), or for deriving information from the genotype data, e.g., by comparing the genotype data to information about genetic risk factors contributing to increased susceptibility to developing an AAA, and reporting results based on such comparison.

[00126] One such aspect relates to computer-readable media. In general terms, such medium has capabilities of storing (i) identifier information for at least one polymorphic marker or a haplotype; (ii) an indicator of the frequency of at least one allele of said at least one marker, or the frequency of a haplotype, in individuals with increased AAA risk, and an indicator of the frequency of at least one allele of said at least one marker, or the frequency of a haplotype, in a reference population. The reference population can be a population of individuals without clinical evidence of an AAA. Alternatively, the reference population is a random sample from the general population, and is thus representative of the population at large. The frequency indicator may be a calculated frequency, a count of alleles and/or haplotype copies, or normalized or otherwise manipulated values of the actual frequencies that are suitable for the particular medium.

[00127] Additional information about the individual can be stored on the medium, such as ancestry information, information about sex, physical attributes or characteristics (including height and weight), biochemical measurements (such as blood pressure, blood lipid levels, lipid levels, such as cholesterol levels) or other useful information that is desirable to store or manipulate in the context of the genotype status of a particular individual.

[00128] The invention furthermore relates to an apparatus that is suitable for determination or manipulation of genetic data useful for determining a susceptibility to develop an AAA in a human individual. Such an apparatus can include a computer-readable memory, a routine for manipulating data stored on the computer-readable memory, and a routine for generating an output that includes a measure of the genetic data. Such measure can include values such as allelic or haplotype frequencies, genotype counts, sex, age, phenotype information, values for odds ratio (OR) or relative risk (RR), population attributable risk (PAR), or other useful information that is either a direct statistic of the original genotype data or based on calculations based on the genetic data.

[00129] The markers and haplotypes shown herein to be associated with increased risk of developing an AAA, are in certain embodiments useful for interpretation and/or analysis of genotype data. Thus in certain embodiments, an identification of an at-risk allele for AAA, as shown herein, or an allele at a polymorphic marker in LD with any one of the markers shown herein to be associated with AAA risk, indicates the increased risk of the individual to developing an AAA. In one such embodiment, genotype data is generated for at least one polymorphic marker shown herein to be associated with AAA risk, or a marker in linkage disequilibrium therewith. The genotype data is subsequently

made available to the individual from whom the data originates, for example via a user interface accessible over the internet, together with an interpretation of the genotype data, e.g., in the form of a risk measure (such as an absolute risk (AR), risk ratio (RR) or odds ratio (OR)) for the development of an AAA. In another embodiment, at-risk markers identified in a genotype dataset derived from an individual are assessed and results from the assessment of the risk conferred by the presence of such at-risk variants in the dataset are made available to the individual, for example via a secure web interface, or by other communication means. The results of such risk assessment can be reported in numeric form (e.g., by risk values, such as absolute risk, relative risk, and/or an odds ratio, or by a percentage increase in risk compared with a reference), by graphical means, or by other means suitable to illustrate the risk to the individual from whom the genotype data is derived. In particular embodiments, the results of risk assessment is made available to a third party, e.g., a physician, other healthcare worker or genetic counselor.

[00130] Nucleic acids and polypeptides

[00131] The nucleic acids and polypeptides described herein can be used in methods and kits of the present invention. An "isolated" nucleic acid molecule, as used herein, is one that is separated from nucleic acids that normally flank the gene or nucleotide sequence (as in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (e.g., as in an RNA library). For example, an isolated nucleic acid of the invention can be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. In some instances, the isolated material will form part of a composition (for example, a

crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material can be purified to essential homogeneity, for example as determined by polyacrylamide gel electrophoresis (PAGE) or column chromatography (e.g., HPLC). An isolated nucleic acid molecule of the invention can comprise at least about 50%, at least about 80% or at least about 90% (on a molar basis) of all macromolecular species present. With regard to genomic DNA, the term "isolated" also can refer to nucleic acid molecules that are separated from the chromosome with which the genomic DNA is naturally associated. For example, the isolated nucleic acid molecule can contain less than about 250 kb, 200 kb, 150 kb, 100 kb, 75 kb, 50 kb, 25 kb, 10 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of the nucleotides that flank the nucleic acid molecule in the genomic DNA of the cell from which the nucleic acid molecule is derived.

[00132] The nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. Thus, recombinant DNA contained in a vector is included in the definition of "isolated" as used herein. Also, isolated nucleic acid molecules include recombinant DNA molecules in heterologous host cells or heterologous organisms, as well as partially or substantially purified DNA molecules in solution. "Isolated" nucleic acid molecules also encompass in vivo and in vitro RNA transcripts of the DNA molecules of the present invention. An isolated nucleic acid molecule or nucleotide sequence can include a nucleic acid molecule or nucleotide sequence that is synthesized chemically or by recombinant means. Such isolated nucleotide sequences are useful, for example, in the manufacture of the encoded polypeptide, as probes for isolating homologous sequences (e.g., from other mammalian

species), for gene mapping (e.g., by in situ hybridization with chromosomes), or for detecting expression of the gene in tissue (e.g., human tissue), such as by Northern blot analysis or other hybridization techniques.

[00133] The invention also pertains to nucleic acid molecules that hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (e.g., nucleic acid molecules that specifically hybridize to a nucleotide sequence containing a polymorphic site associated with a haplotype described herein). In one embodiment, the invention includes variants that hybridize under high stringency hybridization and wash conditions (e.g., for selective hybridization) to a nucleotide sequence that comprises the nucleotide sequence of SEQ ID Nos: 1, 2, or 3, or a fragment thereof, wherein the nucleotide sequence comprises at least one at-risk allele of at least one polymorphic marker, or at least one haplotype, as described herein.

[00134] The percent identity of two nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first sequence). The nucleotides or amino acids at corresponding positions are then compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions x 100). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%, of the length of the reference sequence. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a

mathematical algorithm. A non-limiting example of such a mathematical algorithm is described in Karlin, S. and Altschul, S., Proc. Natl. Acad. Sci. USA, 90:5873-5877 (1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0), as described in Altschul, S. et al., Nucleic Acids Res., 25:3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., NBLAST) can be used. See ncbi.nlm.nih.gov.

[00135] The nucleic acid fragments of the invention are used as probes or primers in assays such as those described herein. "Probes" or "primers" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of a nucleic acid molecule. In addition to DNA and RNA, such probes and primers include polypeptide nucleic acids (PNA), as described in Nielsen, P. et al., Science 254: 1497-1500 (1991). A probe or primer comprises a region of nucleotide sequence that hybridizes to at least about 15, typically about 20-25, and in certain embodiments about 40, 50 or 75, consecutive nucleotides of a nucleic acid molecule comprising a contiguous nucleotide sequence from a haplotype block of the invention, and comprising at least one allele of at least one polymorphic marker or at least one haplotype described herein, or the complement thereof. In particular embodiments, a probe or primer can comprise 100 or fewer nucleotides; for example, in certain embodiments from 6 to 50 nucleotides, or, for example, from 12 to 30 nucleotides. In other embodiments, the probe or primer is at least 70% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence. In another embodiment, the probe or primer is capable of selectively hybridizing to the contiguous nucleotide sequence or to the complement of

the contiguous nucleotide sequence. Often, the probe or primer further comprises a label, e.g., a radioisotope, a fluorescent label, an enzyme label, an enzyme co-factor label, a magnetic label, a spin label, an epitope label.

[00136] The nucleic acid molecules of the invention, such as those described above, can be identified and isolated using standard molecular biology techniques and the sequence information provided by the nucleotide sequence of SEQ ID Nos: 1, 2, and 3. See generally PCR Technology: Principles and Applications for DNA Amplification (ed. H. A. Erlich, Freeman Press, NY, NY, 1992); PCR Protocols: A Guide to Methods and Applications (Eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila, P. et al., Nucleic Acids Res., 19:4967- 4973 (1991); Eckert, K. and Kunkel, T., PCR Methods and Applications, 1: 17-24 (1991); PCR (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202, the entire teachings of each of which are incorporated herein by reference.

[00137] In general, the isolated nucleic acid sequences of the invention can be used as molecular weight markers on Southern gels, and as chromosome markers that are labeled to map related gene positions. The nucleic acid sequences can also be used to compare with endogenous DNA sequences in patients to identify a susceptibility to development of an AAA, and as probes, such as to hybridize and discover related DNA sequences or to subtract out known sequences from a sample (e.g., subtractive hybridization). The nucleic acid sequences can further be used to derive primers for genetic fingerprinting, to raise anti-polypeptide antibodies using immunization techniques, and/or as an antigen to raise anti-DNA antibodies or elicit immune responses.

[00138] The following Examples are meant to illustrate specific embodiments of the present invention, and are not intended to limit the scope of the invention as described above and as claimed herein.

Examples

Example 1

[00139] Methods

[00140] **Study populations**

[00141] Geisinger vascular clinic (GVC) samples. The primary case and control groups consisted of patients enrolled through the Geisinger Clinic Department of Vascular Surgery. Patients had a diagnosis of AAA (defined as an aortic diameter of 3 cm or greater) as determined by abdominal imaging study and confirmed by an attending vascular_surgeon. Controls were patients of the same clinic that were AAA-free, based on evaluation by a vascular surgeon of an abdominal imaging study carried out within the preceding 5 years. At enrollment, all participants provided a blood sample for DNA extraction, completed a questionnaire that collected information about smoking history, family history of AAA, and other information potentially related to AAA risk factors. Individuals also gave permission to extract clinical and demographic information from their electronic medical record (EMR). Characteristics of the GVC individuals are provided in Table III.

Table 3

Characteristics of Study Samples

Geisinger Vascular Clinic			
<u>Cases</u>		<u>Controls</u>	
<u>Female</u>	<u>Male</u>	<u>Female</u>	<u>Male</u>

Number	101	401	114	182
Age (mean $\pm$ sd)	73 $\pm$ 9	74 $\pm$ 8	69 $\pm$ 10	69 $\pm$ 10
(range)	42-92	47-95	39-89	44-88
Smoking	86%	89%	83%	94%

Geisinger MyCode

	<u>Controls</u>	
	<u>Female</u>	<u>Male</u>
Number	98	384
Age (mean $\pm$ sd)	72 $\pm$ 10	73 $\pm$ 8
(range)	44-99	47-96

Replication sample set

	<u>Cases</u>		<u>Controls</u>	
	<u>Female</u>	<u>Male</u>	<u>Female</u>	<u>Male</u>
Number	87	366	202	214

[00142] Geisinger MyCode samples. A secondary control group was obtained through the Geisinger MyCode Project, a population cohort of Geisinger Clinic primary care patients recruited for genomic studies. Enrolled participants agreed to provide blood samples for genomic research and permission to extract clinical and demographic data from their EMR. For this study, DNA samples from 442 MyCode participants were randomly selected from a collection of DNA samples from more than 5000 patients and were matched for age distribution and gender to the GVC cases, and excluded patients with a diagnosis of AAA in their EMR. Characteristics of the MyCode individuals are provided in Table III.

[00143] Replication sample set. These samples consisted of 453 AAA cases and 418 controls that were recruited from Belgium, Canada, and the USA. These samples have been used in other genetic studies of AAA, and have been described previously. (12) Appropriate institutional approvals were obtained for use of human individuals in this research. Characteristics of the replication sample set individuals are provided in Table III.

[00144] DNA purification

[00145] DNA was purified from EDTA anti- coagulated blood using a Qiagen BioRobot M48 Workstation and MagAttract DNA Blood Midi M48 kit (Qiagen, Valencia, Calif). The yield and purity of DNA was determined by measuring the absorbance at 260 and 280 nm using a Nanodrop spectrophotometer (ThermoFisher Scientific, Waltham, Mass).

[00146] Genome wide association study. We carried out an initial GWAS of AAA cases and controls using a DNA pooling strategy similar to that used in previous studies. 13-16 A flow-chart that summarizes the GWAS strategy is in Fig 1. Three identical DNA pools were generated independently from DNA samples of 123 cases or 112 controls (a total of 6 pools) from the GVC samples. Cases and controls were matched for male/female ratio and age; all had a smoking history of ≥ 20 years. The characteristics of the individuals used for the GWAS are shown in Table IV. To create the DNA pools, 100 ng of DNA from each individual was combined and then precipitated at 20 °C by adding 0.1 volume of 3 M sodium acetate, pH 5.2, and 2.5 volumes of ethanol. The DNA was collected by centrifugation, washed with 70% ethanol, and air-dried. The dried pellets were dissolved in water at a final concentration of 50 ng/mL.

Table IV

Characteristics of GWAS Samples

	<u>Cases</u>		<u>Controls</u>	
	<u>Female</u>	<u>Male</u>	<u>Female</u>	<u>Male</u>
Number	51	72	47	65
Age (mean $\pm$ sd)	71 $\pm$ 9	65 $\pm$ 5	71 $\pm$ 7	73 $\pm$ 6
(range)	42-85	51-70	60-83	65-86
Smoking > 20 yr	100%	100%	100%	100%

[00147] The six DNA pools were hybridized separately to Affymetrix SNP arrays using the GeneChip Mapping 500K Manual Protocol (Affymetrix, Inc, Santa Clara, Calif). Pooled DNA (250 ng/array) was digested with Nsp I or Sty I and ligated to adapters that recognize overhangs generated by the restriction enzymes. Oligonucleotides that recognize the adapter sequences were used as primers to amplify the adapter-ligated DNA fragments. The amplified DNA was fragmented, labeled, and hybridized to GeneChip Human Mapping 250K Nsp and 250K Sty arrays (Affymetrix, Inc) for 18 hours at 47 °C. After hybridization, the chips were washed and stained using an FS-450 mapping 500Kv1-450 fluidics workstation (Affymetrix, Inc). The arrays were scanned using the GeneChip Scanner Model 3000 7G. The raw data were analyzed using the Affymetrix GeneChip Operating Software (GCOS), version 1.4.0.036, and GeneChip Genotyping Analysis Software, version 4.1.026 (GTYPE).

[00148] The relative frequencies of the two alleles for each SNP were determined from the hybridization signals. Only the perfect match probe signals were used in the allele frequency calculation. To identify SNPs with a high measurement variance, the intra-group coefficient of variation (CV) of the allele frequency for each SNP was calculated.

SNPs with a $CV \geq 0.1$ in either the case or control group were discarded. For the remaining 306,330 SNPs, the difference in mean allele frequency between cases and controls was calculated. The t test P values for the allele frequency difference between cases and controls for each SNP were also calculated.

[00149] Candidate SNPs for follow-up studies were chosen from genomic regions that contained multiple SNPs with highest case-control allele frequency differences. This was based on the assumption that the high density of SNPs on the arrays (median distance between SNPs on the arrays is 2.5 kb) would reveal multiple SNPs in LD with the genetic variants contributing to the disease. To identify such regions, the SNP allele frequency difference data were sorted by physical position on each autosomal chromosome. Rolling sums of the relative allele frequency differences of five adjacent SNPs or sums of the $-\log_{10}$ transformed P values of the SNP allele frequency differences of five adjacent SNPs along each chromosome were calculated.

[00150] Individual genotype analysis.

[00151] Individual genotype assays were performed using the Applied Biosystems (Foster City, Calif) 7500 Fast Real-Time polymerase chain reaction (PCR) System and the standard protocol for the MicroAmp Fast Optical 96-well reaction plate format (Affymetrix, Inc). PCR amplification was carried out in a 10 μ L reaction using 10 ng of genomic DNA, TaqMan Genotyping Master Mix, TaqMan SNP Genotyping Assay probes (Applied Biosystems), and DNase-free water. Assay conditions were as follows: pre-read hold at 60°C for 30 seconds, fluorescence recording; 1 cycle at 95°C for 10 minutes to activate AmpliTaq Gold DNA Polymerase (Affymetrix, Inc); 40 cycles of 95°C for 15 seconds and 60°C for 1 minute followed by a fluorescence recording after

each cycle; and a hold at 60°C for 30 seconds followed by an endpoint fluorescence recording. Data were analyzed using the 7500 Fast Sequence Detection System Software version 1.4 (Applied Biosystems), with background fluorescence subtraction and automatic allele calling. For SNP rs7635818, the genotype call rates were 99.9% with the Geisinger Clinic DNA samples and 98.8% with the replication samples.

[00152] Contactin-3 transcript assay.

[00153] Aortic tissue was homogenized in Trizol reagent (Invitrogen, Carlsbad, Calif) and total ribonucleic acid (RNA) was extracted following the standard protocol.

Mononuclear cells were obtained from blood samples collected in Cell Preparation Tubes with Citrate (BD Diagnostics, Franklin Lakes, NJ). RNA was isolated using QIAshredder spin columns and RNeasy kits (Qiagen, Valencia, Calif). Yield and quality of RNA were determined by measuring the absorbance at 260 nm and 280 nm.

[00154] Total RNA was used as a template for cDNA synthesis in a reaction that contained the following: 2 µg RNA, 4 µl of 10X Mg free buffer (Promega, Madison, Wis), 5 mM MgCl₂(Promega, Madison, Wis), 5 µM random hexamers, 1 mM dNTPs, 2 µl Superscript II RT (Invitrogen), and PCR grade water to a final volume of 40 µl.

Reaction conditions were 25°C for 10 minutes, 50°C for 50 minutes, 70°C for 15 minutes followed by a 4°C hold.

[00155] cDNAs were used as templates for PCR amplification of *CNTN3* transcripts using JumpStart Taq DNA Polymerase (Sigma, St Louis, Mo). The standard protocol was used, but with the addition of MgCl₂ to a final concentration of 15 mM. Primers for the *CNTN3* gene were designed using PrimerQuest (Integrated DNA Technologies, Coralville, Iowa). Primer pairs used were: forward 5' –

TGCCCTTGGAATCCCATACCTCA (SEQ ID NO:4) – 3' and reverse 5' –
TGTCCTCCACGGCTATTTCCACAT – 3' (SEQ ID NO:5) (predicted product size of
251 bp); and forward 5' – ACAGACACAACAGCCCAACTCTCT – 3' (SEQ ID NO:6)
and reverse 5' – GGCGGAAAGCAACAACATACCCAA – 3' (SEQ ID NO:7)
(predicted product size of 397 bp). PCR amplification conditions were 1 minute at 94°C
followed by 35 cycles of 94°C for 1 minute, 55°C for 1 minute, and 72°C for 30 seconds,
followed by a final extension at 72°C for 5 minutes and a 4°C hold. Products were
analyzed on 1% agarose gels and stained with ethidium bromide.

[00156] Statistical analyses.

[00157] Genotype data were analyzed using Helix Tree Software, version 6.4 (Golden Helix, Inc, Bozeman, Mont). Analyses included calculations of deviation from Hardy-Weinberg equilibrium, linkage disequilibrium, and genetic association. *P* values for genetic association were calculated using recursive partitioning with case/ control selected as a categorical dependent variable and genotype as the independent variable. The data were also analyzed assuming a dominant model of genetic association, which tests the association of having at least one minor allele vs having no copies of the minor allele. Both approaches yielded essentially identical results. Uncorrected *P* values are reported.

[00158] Power calculations were performed using the Genetic Power Calculator. 17 We assumed that the polymorphism and the disease locus were in complete LD and had the same allele frequencies. Power was calculated for different model parameters. Assuming a disease prevalence of 0.05 and risk allele frequency of 0.4, the Geisinger Clinic sample size of 502 cases and 748 controls had an 80% power to detect a

susceptibility locus with a genotypic relative risk (GRR) of 1.42 for a dominant genetic model at a significance level of 0.1. The combined data set of 995 cases and 1154 controls could detect a dominant disease locus with a GRR of 1.34 with 80% power at a significance level of 0.05.

[00159] Results

[00160] **GWAS identifies a variant on 3p12.3 associated with AAA.**

[00161] As the first part of a multi-stage approach to identify AAA-associated genetic variants, we carried out a GWAS using a subset of GVC case and control DNA samples. Relative SNP allele frequencies in triplicate pools of 123 case and 112 control samples were determined using Affymetrix 500K SNP arrays. (See Fig. 1 for additional details on the GWAS.) Criteria for prioritizing candidate SNPs included a combination of absolute case-control SNP allele frequency differences, *t* test *P* values of SNP allele frequency differences, and physical clustering of SNPs with allele frequency differences.

[00162] Twenty genomic regions on 13 autosomal chromosomes that contained clusters of SNPs with significant allele frequency differences between the case and control groups (Fig 2) were selected for initial analysis by SNP genotype assay in a sample consisting of 550 GVC cases and controls. A region on chromosome 3p12.3 showed the strongest association with AAA, and was selected for further analysis, as described in the following sections.

[00163] **Validation of results through individual SNP genotyping.**

[00164] To conditionally validate the association of this genomic region with AAA, 10 SNPs in this region were genotyped in a total of 502 GVC cases and 296 control samples. Four SNPs, including rs7635818, were significantly associated with AAA (Fig. 3). The

AAA-associated SNPs are contained within a haplotype block in the HapMap CEU (Caucasians of European descent from Utah) population (Fig. 3) and were in strong LD in the GVC case and control samples ($D' \geq 0.98$) (Fig. 4). SNP rs7635818 was genotyped in an additional set of 442 controls obtained from the Geisinger MyCode population that were matched to the GVC case samples for age and gender. In the combined Geisinger samples, (502 cases and 748 controls) SNP rs7635818 genotype was significantly associated with AAA ($P = .017$) (Table V). The AAA-associated allele was C (with G as the major allele) and its frequency was 0.48 in cases and 0.42 in controls.

Table V

<i>Sample set</i>	<i>N</i>	<i>Genotype frequency</i>				<i>P value</i>
		<i>CC/CG</i>		<i>GG</i>		
		<i>Case</i>	<i>Control</i>	<i>Case</i>	<i>Control</i>	
Geisinger	502 cases	.73	.67	.27	.33	.017
Clinic	736 controls					
Replication	448 cases	.73	.67	.27	.33	.013
set	410 controls					
All samples	950 cases	.73	.67	.27	.34	.0028
	1146 controls					

SNP, Single nucleotide polymorphism; *AAA*, abdominal aortic aneurysm; *N*, number.

[00165] Replication of association in an independent sample set.

[00166] To further substantiate the genetic association with AAA, SNP rs7635818 was genotyped in an independent set of 453 case and 416 control samples obtained from Belgium, Canada, and the USA. SNP rs7635818 was significantly associated with AAA ($P = .013$) in this replication sample (Table V), providing additional evidence that this

region contains AAA-associated genetic variants. When the data were analyzed in all samples combined (955 cases and 1154 controls), SNP rs7635818 genotype was significantly associated with AAA ($P = .0028$) (Table V). A dominant genetic model yielded the most significant association between the SNP genotype and AAA.

[00167] Interaction with smoking.

[00168] Smoking is a strong environmental risk factor for AAA disease. Because we had detailed smoking history data on the GVC case and control samples, we calculated the association of SNP rs7635818 with AAA in the subset of 391 cases and 241 controls with a smoking history of ≥ 20 years, and compared these data to those for the unstratified GVC cases and controls and the combined set of all cases and controls. In the combined sample set, the odds ratio (OR) for association of SNP rs7635818 with AAA was 1.33 (95% confidence interval [CI] 1.10-1.61, $P = .0014$) (Fig. 5). In the unstratified GVC samples, the OR was 1.56 (95% CI 1.15-2.13, $P = .0044$), and in the GVC smokers-only samples the OR was 1.80 (95% CI 1.3-2.5, $P = .00041$). Thus, the association of SNP rs7635818 with AAA appears to be stronger in individuals with a history of smoking.

[00169] *CNTN3* is expressed in AAA and other vascular tissues. SNP rs7635818 is upstream of the *CNTN3* gene, which encodes a lipid-anchored cell adhesion protein (contactin-3). (18) The *CNTN3* gene has not been reported previously to be expressed in vascular tissue. To assay *CNTN3* transcripts, RNA extracted from AAA tissue, control aortic tissue, aortic-occlusive disease (AOD) tissue, and peripheral blood mononuclear cells from AAA patients was used to generate cDNAs that were used as templates for PCR amplification using *CNTN3*-specific primers. Products derived from *CNTN3*

transcripts were detected in all the tissues sampled (Fig. 6), demonstrating that the gene is actively expressed in aortic tissue, including AAA specimens.

[00170] Discussion

[00171] We carried out a genome wide association study to identify genomic variants associated with AAA. A DNA pooling strategy in conjunction with high density allele frequency determinations using Affymetrix 500K SNP arrays identified a candidate genomic region on chromosome 3p12.3. To validate the association of this region with AAA, several SNPs in this region were individually genotyped in additional case and control samples. Genetic association was also confirmed in an independent replication sample set. One SNP in this region, rs7635818, was genotyped in a total of 2096 case and control samples. SNP rs7635818 genotype was significantly associated with AAA, with a χ^2 *P* value of .0028 and a calculated OR of 1.33 (1.10-1.61 95% CI). This is comparable to the OR of the previously reported AAA-associated SNP on chromosome 9. (10)

[00172] The Affymetrix chips we used contained probes for approximately 500,000 common SNPs, which are located throughout the genome; most do not overlap known genes. These differ from other gene chips that are used to measure gene activity or expression. An advantage of a genome wide association study of the type reported here is that the outcome is not dependent on a priori assumptions about the underlying biological mechanisms. SNP rs7635818 is probably not the genetic variant that is functionally responsible for increased AAA risk. Rather, it is likely that this SNP serves as a marker for a functional genetic variant that increases AAA risk that is in LD with the marker SNP.

[00173] SNP rs7635818 is contained within a haplotype block that shows strong LD in the HapMap CEU population and in the case and control groups used for this study. Our genetic mapping studies showed the AAA associated haplotype block that contains SNP rs7635818 to span approximately 30 kb of genomic DNA sequence. This region does not overlap any known protein coding genes, but is approximately 200 kb upstream of the transcription start site for the *CNTN3* gene. (The next nearest genes are PDZRN3 and ROBO2, which are approximately 1.3 and 2.2 million bp away, respectively.) There are no non-synonymous SNPs in *CNTN3* reported in public databases that change the predicted structure of activity of the *CNTN3* gene product. Together, these results suggest that the functional genetic variant has a regulatory function. Because it lies upstream of the *CNTN3* gene, where sequence motifs that regulate *CNTN3* gene activity are expected to be located, our working hypothesis is that the functional variant is in the upstream regulatory region of the *CNTN3* gene. The product of the *CNTN3* gene is a lipid-anchored cell adhesion protein that had been previously shown to be expressed in the central nervous system. (18) In support of this hypothesis, we found that the *CNTN3* transcript is readily detected in aortic tissue, including both normal and AAA tissue. Although this genomic region contains consensus DNA binding sites for known transcription factors, no studies to date have examined the effects of sequence variants in the AAA-associated haplotype block marked by SNP rs7635818 on *CNTN3* expression. Unfortunately, the limited number of high quality tissue samples that were available from AAA and risk-matched normal aortas did not allow a quantitative analysis of contactin-3 mRNA levels to determine whether the genetic variant altered expression of the gene. Moreover, since the cellular and biochemical changes that initiate AAA formation, the events most likely

to be influenced by genes that increase AAA risk, occur well before the lesions are diagnosed clinically and repaired, it is not feasible to obtain biological samples from humans to test this hypothesis directly.

[00174] A dominant genetic model yielded the most significant association between SNP rs7635818 genotype and AAA. This would be consistent with the existence of a genetic variant in the *CNTN3* promoter that affects contactin-3 expression, either constitutively or in response to regulatory signals. Regulation of contactin-3 expression in vascular tissue has not been investigated.

[00175] A potentially important suggestive finding from the study was that the genetic risk associated with SNP rs7635818 was greater in individuals with a history of smoking. Smoking is a strong non-genetic risk factor for AAA. These findings suggest a gene-environment interaction that influences AAA risk. Because smoking history data were only available on the individuals recruited from the Geisinger Vascular Clinic, there were not sufficient numbers of non-smoking cases or controls to carry out an independent calculation of genetic risk in non-smokers. It is also interesting to note that the initial GWAS was carried out using only cases and controls with a history of smoking. This might explain why a relatively underpowered GWAS identified this genomic region. The matching for smoking history in the GWAS and demonstration of genetic association in the smoking sub-group also rules out the possibility that the observed association is related to smoking or a related characteristic (e.g., smoking addiction).

[00176] These considerations also highlight the value of DNA samples from thoroughly characterized individuals for genetic association studies. A sufficient breadth of data allows for careful matching of cases and controls for known risk factors and the

ability to examine interactions among genetic and non-genetic risk factors. It is also important to note that the GVC controls were screened to eliminate patients with undiagnosed AAA, which has been estimated to occur in up to 6% of individuals in the highest risk group, i.e., male smokers over the age of 70 years. This could explain why the *P* value for association of SNP rs7635818 with AAA was not as robust in the replication sample set compared to the Geisinger samples, as these were not as closely matched for the other risk factors (e.g., male gender) and were not screened for undiagnosed AAA.

[00177] The genetic variant reported here has a moderate but statistically significant association with AAA disease. This variant and the previously reported SNP on chromosome 9p21 account for only a portion of the genetic risk for AAA. Studies are ongoing to discover additional AAA-associated genetic variants. A long-term goal of this research is to use genetic information to quantify AAA disease risk, probably through the use of a panel of AAA-associated genetic variants. This would enable high-risk individuals to be identified for closer monitoring to detect undiagnosed AAAs, so that appropriate treatment can be provided in a timely manner. Knowledge of disease-associated genetic variants can also provide new insight into disease mechanisms and provide targets for development of novel therapies. (19) At the present time, there are no medical treatments for AAAs. The proximity of this AAA-associated genetic variant to the *CNTN3* gene, for example, should provide the impetus to study that gene and its product, contactin-3, in the context of vascular disease.

Example 2

[00178] Identification of a AAA-associated genetic variant on chromosome 1q41

[00179] Individuals for a genomic study of AAA were recruited from Geisinger Clinic (see Methods). The characteristics of the study cohorts are summarized in Table VI.

Table VI
Characteristics of Study Samples

	<u>Cases</u>		<u>Controls</u>	
	<u>Female</u>	<u>Male</u>	<u>Female</u>	<u>Male</u>
GWAS Samples				
Number	51	72	47	65
Age, years (mean $\pm$ sd)	71 $\pm$ 9	65 $\pm$ 5	71 $\pm$ 7	73 $\pm$ 6
(range)	42-85	51-70	60-83	65-86
Smoking > 20 yr (%)	100	100	100	100
Geisinger Vascular Clinic				
Number	141	512	137	217
Age, years (mean $\pm$ sd)	73 $\pm$ 9	74 $\pm$ 8	70 $\pm$ 10	69 $\pm$ 10
(range)	43-92	47-96	39-89	44-88
Smoking (%)	85	89	83	94
Geisinger MyCode				
Number	2	38	96	344
Age, years (mean $\pm$ sd)	86 $\pm$ 9	77 $\pm$ 8	72 $\pm$ 10	73 $\pm$ 8
(range)	83-90	57-93	44-99	47-96
Replication sample set				
Number	87	366	202	214

[00180] As the first step in a multi-stage process to identify AAA-associated genetic variants we carried out a GWAS using pooled DNA samples from 123 AAA case and 112 control individuals recruited through the Geisinger Clinic Department of Vascular

Surgery. Controls were matched to the cases for age, gender, and smoking history.

Relative SNP allele frequencies were analyzed to identify candidate AAA-associated SNPs for subsequent validation.

[00181] One genomic region that was selected for further study was within chromosome 1q41. This region contained a cluster of SNPs with statistically significant allele frequency differences between the case and control samples in the pooled DNA GWAS (Fig. 7). As an initial validation step, six SNPs within this region were individually genotyped in a subset of 451 cases and 279 controls from the Geisinger vascular surgery clinic (Fig. 8). Four of these SNPs had genotypes that were significantly associated with AAA (P values ranging from 0.001 to 0.014). These AAA-associated SNPs are in a region of strong LD in the HapMap CEU population (Fig. 8), and were in strong LD in the genotyped samples from the Geisinger vascular clinic (Fig. 9).

[00182] These results are consistent with a AAA-associated haplotype in this region of chromosome 1q41. Genotype data for these 6 SNPs were used to estimate haplotype frequencies. The most common haplotype was present at frequencies of 0.42 ± 0.03 and 0.36 ± 0.03 in cases and controls, respectively. Haplotype analysis revealed a significant association with AAA (Table VII). The peak of genetic association was centered over SNPs rs12039875 and rs11120508, which are in near perfect linkage disequilibrium (LD correlation $R = 0.96$) and separated by a distance of ~47 kbp. In subsequent experiments rs12039875 was used as a tag SNP for this AAA-associated haplotype.

Table VII

Haplotype Analysis of chromosome 1q41 SNPs

SNPs								
rs2841599	rs17024179	rs12039875	rs11120508	rs10494994	rs1538547	haplotype frequency*		
C/G	T/C	C/A	A/G	G/A	A/C	case	control	P value
C/G	T	C	A/G	G/A	A/C	0.72	0.61	0.00022
C/G	T	C	A	G/A	A/C	0.70	0.60	0.00089
C/G	T	C	A	G	A/C	0.68	0.58	0.0029
C/G	T	C	A	G	A	0.53	0.45	0.0011
C	T	C	A	G	A	0.42	0.36	0.028

*The most significant haplotype for window sizes of 2-6 are shown; individual alleles in each haplotype are highlighted.

[00183] To further validate the association of this genomic region with AAA disease SNP rs12039875 was genotyped in additional case and control samples from the Geisinger Vascular Clinic and from the Geisinger MyCode Project. The genotype of SNP rs12039875 was significantly associated with AAA in the combined Geisinger Clinic sample set ($P = 0.0085$, 690 cases, 792 controls) (Table VIII).

Table VIII

Association of SNP rs12039875 on chromosome 1q41 with AAA

Samples	N	Allele Frequency		P value*
		C	A	
Geisinger Clinic	690 cases	0.71	0.29	0.0085
	792 controls	0.67	0.33	
Replication	452 cases	0.70	0.30	0.029
	412 controls	0.67	0.33	
All	1,142 cases	0.71	0.29	0.0072

1,204 controls	0.67	0.33
----------------	------	------

*P values were calculated by recursive partitioning (Helix Tree Software) using case/control as the categorical dependent variable and genotype as the independent variable.

[00184] This SNP was also genotyped in case and control samples from an independent replication sample set obtained from Belgium, Canada, and the USA. SNP rs12039875 was significantly associated with AAA in the replication sample set ($P = 0.029$, 452 cases, 412 controls) (Table VIII).

[00185] In all samples combined (1,142 cases, 1,204 controls) the association P value for SNP rs12039875 was 0.0072. The AAA-associated allele was C. The odds ratio for AAA calculated using a basic allelic genetic model was 1.18 (95% CI 1.05 – 1.34).

[00186] KCNK2 is expressed in AAA tissue

[00187] The AAA-associated haplotype overlies the *KCNK2* gene, which encodes a 2-pore-domain membrane potassium channel. *KCNK2* transcripts were assayed by PCR amplification of cDNAs generated by reverse transcription of total RNA isolated from AAA tissue, normal aortic tissue, and tissue from aortic occlusive lesions. *GAPDH* was used as a positive control. Three *KCNK2* isoforms are produced from this gene that differ in their 5'-untranslated and N-terminal amino acid sequences (isoform a (SEQ ID NO: 8), isoform b (SEQ ID NO: 9) and isoform c (SEQ ID NO: 10). Products corresponding to *KCNK2* isoforms b and c were detected in all AAA samples tested ($N = 4$). These *KCNK2* isoforms were detected at much lower levels in the occlusive lesions ($N = 2$) and normal aortic tissue ($N = 5$) (Fig. 10). The mean *KCNK2* level in AAA tissue was approximately 4 times the level in normal aortic tissue. *KCNK2* transcripts were not detected in RNA samples isolated from circulating mononuclear cells obtained

from individuals with unrepaired or surgically repaired AAA lesions or individuals without AAA (data not shown).

[00188] Association of SNP rs10757278 with AAA disease and non-AAA vascular disease

[00189] SNP rs10757278 on chromosome 9p21 was identified previously as a genetic variant that is associated with coronary artery disease and subsequently reported to be associated with AAA as well as intracranial aneurysm. (12) We determined the SNP rs10757278 genotype in AAA cases and controls from the Geisinger Clinic Department of Vascular Surgery and the Geisinger MyCode Project. When the AAA cases and non-AAA controls from the vascular clinic were compared no association of the SNP genotype with AAA was observed ($P = 0.36$, 650 cases, 355 controls) (Table IX). However, when the AAA cases were compared to the MyCode population controls a significant association was observed ($P = 0.007$, 650 cases, 437 controls).

Table IX

Association of SNP rs10757278 on chromosome 9p21 with AAA

<i>Cases</i>	<i>G allele frequency</i>	<i>Controls</i>	<i>G allele frequency</i>	<i>P value*</i>
AAA (N = 650)	0.52	Vascular clinic non-AAA (N = 355)	0.53	0.36
AAA (N = 650)	0.52	MyCode controls (N = 437)	0.45	0.007
Vascular disease [†] (N = 462)	0.53	MyCode controls (N = 437)	0.45	0.003

Vascular disease (no AAA) (N = 353)	0.54	MyCode controls (N = 437)	0.45	0.001
All vascular (AAA + non-AAA) (N = 1,043)	0.53	MyCode controls (N = 437)	0.45	0.002

*P values were calculated by recursive partitioning (Helix Tree Software) using case/control as the categorical dependent variable and genotype as the independent variable

†from the Geisinger Vascular Clinic; includes 109 subjects with both AAA and peripheral vascular disease

[00190] One potential explanation for this apparent discrepancy is that SNP rs10757278 is associated with AAA and with other forms of vascular disease. This would explain the lack of association when comparing AAA cases and non-aneurysm controls from the vascular clinic, all of whom have some form of vascular disease. This is supported by the finding that comparison of SNP rs10757278 genotype in vascular clinic controls (with vascular disease but without AAA) and the MyCode population controls identified a significant association ($P = 0.001$, 353 vascular disease cases and 437 population controls) (Table IX). Similarly, comparison of the SNP genotype in the entire set of vascular samples (AAA and non-AAA) with the population control group showed a highly significant association ($P = 0.002$, 1,043 vascular cases and 437 population controls).

[00191] Cumulative effects of AAA-associated genetic variants on AAA risk

[00192] The SNP variant on chromosome 9p21 (rs10757278), (12) our previously reported AAA-associated variant on chromosome 3p12.3 (rs7635818), (22) and the chromosome 1q41 variant (rs12039875) described above, each make a relatively small contribution to overall AAA risk. To determine the effect of cumulative allele burden on

AAA disease risk we analyzed genotype data for these SNPs in 938 AAA cases and 1,126 non-AAA control samples for which genotype data for all 3 SNPs were available.

[00193] A simple genetic model that assumed additive effects of risk alleles and no genetic interactions was used. As shown in Fig. 11, the distribution of total risk allele burden for the 3 genetic variants (from 0 to 6 risk alleles) was different in the AAA case and control groups, with the case group skewed toward more total risk alleles ($P < 0.0001$). Consistent with this finding, the relative odds ratio for AAA disease increased as a function of the number of risk alleles. For individuals with 3 or more risk alleles the odds ratio for AAA disease was increased 2.4-fold compared to individuals with 0-2 alleles ($P < 0.0001$). For individuals with 6 risk alleles the odds ratio was increased 3.5-fold ($P < 0.01$).

[00194] The association of allele burden with AAA was also examined by linear regression analysis (Fig. 11). The linear regression line had a positive non-zero slope ($P < 0.01$, $R^2 = 0.8$). The AAA disease odds varied by 3.7-fold from 0 to 6 risk alleles.

[00195] Discussion

[00196] This report describes a haplotype on chromosome 1q41 that is associated with AAA disease. The haplotype is identified by several SNP variants that are in strong LD. SNP rs12039875 was used as a marker for this haplotype to establish its association with AAA in 2 independent populations. The AAA-associated haplotype overlaps the *KCNK2*, which encodes a 2-pore membrane potassium channel. The *KCNK2* gene is expressed in AAA tissue at levels that are higher than normal aorta.

[00197] This represents at least the third genetic locus reported to be associated with AAA disease in a case-control study, and the second to result from our pooled sample

case-control GWAS. The identification of multiple AAA-associated genetic loci, each with a significant but relatively small contribution to disease risk, is consistent with the concept of AAA as a complex disease that is affected by an unknown number of genetic factors as well as environmental risk factors. The other AAA-associated genetic loci are SNP rs10757278 on chromosome 9p21, (12) and SNP rs7635818 on chromosome 3p12.3 that was identified by GWAS of pooled DNA samples in our laboratory. (22)

[00198] The AAA-associated haplotype on chromosome 1q41 overlies the *KCNK2* gene, which encodes a membrane potassium channel. SNP rs12039875 lies within the first intron of the 2 *KCNK2* isoforms that we show are expressed in AAA tissue. The steady state levels of *KCNK2* transcripts are significantly higher in AAA tissue than normal aorta. The fact that *KCNK2* transcripts were not detected in peripheral blood mononuclear cells of AAA patients suggests that the gene is expressed by cells of the vascular wall and not by infiltrating inflammatory or immune cells, although this needs to be confirmed by additional studies. While not previously reported in AAA tissue, the *KCNK2* potassium channel has been shown to be expressed in subsets of vascular smooth muscle cells, where it has been suggested to play a role in regulation of blood pressure in response to lipids. (23)

[00199] No common genetic variants that alter the protein coding information of *KCNK2* are present in the public databases. Therefore, the functional genetic variant in this AAA-associated haplotype is not known. The lack of known non-synonymous SNPs in this gene and data showing increased expression of *KCNK2* in AAA tissue support the hypothesis that the functional genetic variant influences AAA risk by altering *KCNK2* expression. One of the limitations of studying human AAA is that the main source of

AAA tissue is patients undergoing open surgical repair of their aneurysm. In nearly all cases, however, AAA repair is limited to aneurysmal lesions with a diameter of 5 cm or greater. These are usually advanced lesions that have undergone a number of cellular and molecular changes since the event that initiated aneurysm formation, which is likely to have occurred years earlier. Therefore, the state of gene expression at the time of surgery reflects the end stage of the disease, and probably does not accurately reflect earlier stages. If the gene-variant induced changes in *KCNK2* expression that alter AAA risk affect an early step in disease progression, it might not be evident at the time of surgery.

[00200] Our GWAS used a DNA pooling strategy, similar to other previously reported studies. (13, 16) The main advantage of DNA pooling is that the time, effort, and cost of conducting the GWAS are substantially reduced compared to individual genome-wide analysis of case and control samples. Previous studies, further supported by our results, show that Affymetrix high density SNP arrays have sufficient analytical power to reliably identify allele frequency differences between case and control pools. One of the keys is carrying out replicate independent array assays, which enables a determination of intra-group allele frequency differences. As these reflect to a large extent differences in the measurement variance of individual SNP probes on the arrays (i.e. not all SNPs are measured with equal reliability) this allows SNPs with insufficient measurement confidence to be discarded from further analysis. In our GWAS an intra-group coefficient of variation of 0.1 was selected as the threshold, which resulted in approximately 20% of the 500,000 SNPs interrogated on the arrays to be discarded.

[00201] There are also limitations of the DNA pooling strategy. One is the inability to stratify the GWAS data within the case and control groups for independent variables that

might influence the genetic association. For this reason, the case and control samples used in our pooled DNA GWAS were carefully matched for the known major risk factors, age, gender and smoking history. Another limitation is that the pooled array data yield only allele frequency data. Determination of genotypes requires subsequent assays of individual samples. Our approach is to use the pooled DNA GWAS to identify candidate genomic regions that are validated by independent genotype analysis in subsequent stages. Given the advantages and demonstrated success, we consider DNA pooling as a reasonable alternative to individual high density genotyping for GWAS discovery.

[00202] We also report results of genotyping of SNP rs10757278 in AAA cases and controls in the Geisinger Clinic population. This SNP on chromosome 9p21 was originally identified as a variant associated with coronary artery disease, and then later shown to be associated with AAA and with intracranial aneurysm. (12) In our studies the SNP genotype was strongly associated with AAA when an age and gender-matched population control sample was used, but not when the controls consisted of patients with peripheral vascular disease but confirmed to be AAA-free. From these data we conclude that this genetic variant is associated with multiple forms of vascular disease including, but not specifically, AAA.

[00203] In this paper we also show that analysis of genotype data from a panel of 3 SNPs from different AAA-associated genetic loci, including the newly reported SNP rs12039875 on chromosome 1q41, is substantially more informative with respect to AAA risk than data from individual SNPs. This suggests that it should be feasible to identify a panel of AAA-associated genetic variants that has sufficient information to stratify individuals for AAA risk in a manner that is useful for guiding clinical care. For example,

individuals with a high genetic risk could be monitored more frequently to detect incident AAA lesions, e.g. by ultrasound screening, than individuals with a low genetic risk.

Consideration of genetic factors might also increase AAA diagnosis in individuals with an elevated genetic risk of AAA who lack the major non-genetic risk factors, e.g. women and non-smokers, and might not otherwise be targeted for disease surveillance. For example, under current guidelines published by the U.S. Preventive Services Task Force of the Agency for Healthcare Research and Quality, AAA-screening by ultrasonography is recommended only for male smokers between the ages of 65 and 75 years. (24)

Because currently available surgical and endovascular procedures to repair AAA lesions are safe and effective, increased diagnosis through use of genetic testing should reduce the prevalence of rupture of undiagnosed AAAs and consequently reduce AAA-associated mortality.

[00204] The results presented are based on a simple model that assumes additive genetic risk and no interaction among genes or between genes and other risk factors. The association of SNP rs12039875 with AAA appears to be independent of other major risk factors, age, male gender, and smoking, since the genetic association was observed in cases and control groups that were matched for these factors. Stratification of cases and controls using these variables did not alter the apparent association of the SNP genotype with AAA (data not shown). This is in contrast to SNP rs7635818 on chromosome 3p12.3, which we showed to have a bigger odds ratio for AAA in smokers. (22) This suggests that more complex risk modeling that looks at genetic interactions or genetic factors in combination with other known risk factors could provide even more powerful risk prediction.

[00205] References

- [00206]** 1. Alcorn HG, Wolfson SK Jr, Sutton-Tyrrell K, O'Leary D. Risk factors for abdominal aortic aneurysms in older adults enrolled in The Cardiovascular Health Study. *Arterioscler Thromb Vasc Biol* 1996;16:963-70.
- [00207]** 2. Reilly JM, Tilson MD. Incidence and etiology of abdominal aortic aneurysms. *Surg Clin North Am* 1989;69:705-11.
- [00208]** 3. Elmore JR, Keister BF, Franklin DP, Youkey JR, Carey DJ. Expression of matrix metalloproteinases and TIMPs in human abdominal aortic aneurysms. *Ann Vasc Surg* 1998;12:221-8.
- [00209]** 4. Kuivaniemi H, Platsoucas CD, Tilson MD 3rd. Aortic aneurysms: an immune disease with a strong genetic component. *Circulation* 2008; 117:242-52.
- [00210]** 5. Lenk GM, Tromp G, Weinsheimer S, Gatalica Z, Berguer R, Kuivaniemi H. Whole genome expression profiling reveals a significant role for immune function in human abdominal aortic aneurysms. *BMC Genomics* 2007;8:237.
- [00211]** 6. McMillan WD, Pearce WH. Inflammation and cytokine signaling in aneurysms. *Ann Vasc Surg* 1997;11:540-5.
- [00212]** 7. Thompson RW, Parks WC. Role of matrix metalloproteinases in abdominal aortic aneurysms. *Ann N Y Acad Sci* 1996;800:157-74.
- [00213]** 8. Treska V, Kocova J, Boudova L, Neprasova P, Topolcan O, Pecan L, Tonar Z. Inflammation in the wall of the abdominal aortic aneurysm and its role in the symptomatology of aneurysm. *Cytokines Cell Mol Ther* 2002;7:91-7.
- [00214]** 9. Kuivaniemi H, Boddy AM, Lillvis JH, Nischan J, Lenk GM, Tromp G. Abdominal aortic aneurysms are deep, deadly and genetic. In: Sakalihasan N, Kuivaniemi

H, Michel JB, editors. Aortic aneurysms: new insights into an old problem. Liege, Belgium: Liege University Press; 2008.

[00215] 10. Shibamura H, Olsen JM, van Vlijmen-van Keulen C, Buxbaum SG, Dudek DM, Tromp G, et al. Genome scan for familial abdominal aortic aneurysm using sex and family history as covariates suggests genetic heterogeneity and identifies linkage to chromosome 19q13. *Circulation* 2004;109:2103-8.

[00216] 11. van Vlijmen-van Keulen C, Rauwerda JA, Pals G. Genome-wide linkage in three Dutch families maps a locus for abdominal aortic aneurysms to chromosome 19q13.3. *Eur J Vasc Endovasc Surg* 2005;30:29-35.

[00217] 12. Helgadóttir A, Thorleifsson G, Magnusson KP, Grétarsdóttir S, Steinhórsdóttir V, Manolescu A, et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet* 2008;40:217-24.

[00218] 13. Butcher LM, Davis OS, Craig IW, Plomin R. Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes Brain Behav* 2008;7:435-46.

[00219] 14. Craig DW, Huentelman MJ, Hu-Lince D, Zismann VL, Kruer MC, Lee AM, et al. Identification of disease causing loci using an array-based genotyping approach on pooled DNA. *BMC Genomics* 2005;6:138.

[00220] 15. Docherty SJ, Butcher LM, Schalkwyk LC, Plomin R. Applicability of DNA pools on 500 K SNP microarrays for cost-effective initial screens in genomewide association studies. *BMC Genomics* 2007;8:214.

- [00221]** 16. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, et al. Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. *Am J Hum Genet* 2007;80:126-39.
- [00222]** 17. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;19:149-50.
- [00223]** 18. Yoshihara Y, Kawasaki M, Tamada A, Nagata S, Kagamiyama H, Mori K. Overlapping and differential expression of BIG-2, BIG-1, TAG-1, and F3: four members of an axon-associated cell adhesion molecule subgroup of the immunoglobulin superfamily. *J Neurobiol* 1995;28: 51-69.
- [00224]** 19. Boddy AM, Lenk GM, Lillvis JH, Nischan J, Kyo Y, Kuivaniemi H. Basic research studies to understand aneurysm disease. *Drug News Perspect* 2008;21:142-8. Submitted Oct 29, 2008; accepted Jan 18, 2009.
- [00225]** 20. Sakalihasan N, Delvenne P, Nussgens BV, Limet R, Lapiere CM. Activated forms of MM2 and MMP9 in abdominal aortic aneurysms. *J. Vasc Surg.* 1996; 24:127-133
- [00226]** 21. Jones GT, Thompson AR, van Bockxmeer FM Hafez H, Cooper JA, Golledge J, Humphries SE, Norman PE, van Rij AM. Angiotensin II type 1 receptor 1166 polymorphism is associated with abdominal aortic aneurysm in three independent cohorts. *Arterioscler Thromb Vasc Biol.* 2008; 28:764-770.
- [00227]** 22. Elmore JR, Obmann MA, Kuivaniemi H, Tromp G, Gerhard GS, Franklin DP, Boddy AM, Carey DJ. Identification of a genetic variant associated with abdominal

aortic aneurysms on chromosome 3p12.3 by genome wide association. J Vasc Surg. 2009; 49: 1525-1531.

[00228] 23. Blondeau N, Petrault O, Manta S, Giodanengo V, Gounon P, Bordet R, Lazdunski M, Heurteaux C. Polyunsaturated fatty acids are cerebral vasodilators via the TREK-1 potassium channel. Circ Res. 2007; 101: 176-184.

[00229] 24. Screening for abdominal aortic aneurysm. Available at: <http://www.ahrq.gov/clinic/uspstf/uspsaneu.htm>.

What is claimed is:

1. A method for determining a genetic predisposition to developing an abdominal aortic aneurysm in an individual, comprising:

obtaining nucleic acid from the individual, wherein the nucleic acid comprises a sequence corresponding to or complementary to a sequence from a genomic region selected from chromosome 3p12.3 and 1q41;

determining the sequence of the nucleic acid; and

determining the presence of alleles on the nucleic acid that correlate with a risk of developing an abdominal aortic aneurysm, wherein the presence of an allele that correlates to risk of developing an abdominal aortic aneurysm indicates that the individual has a genetic predisposition of developing an abdominal aortic aneurysm.

2. The method of claim 1, wherein the genomic region is from chromosome 3p12.3 and the sequence corresponds to HapMap positions 74762904 – 74998753.

3. The method of claim 1, wherein the sequence comprises SEQ ID NO: 1.

4. The method of claim 1, wherein the allele to be determined is selected from the group consisting of: rs6549634, rs12714811, rs7642555, rs7635639, rs7635818, rs1569066, rs10222550, rs1527527, rs2178032 and rs4417905.

5. The method of claim 1, wherein the allele to be determined is rs7635818 and C is the allele that correlates to risk of developing an abdominal aortic aneurysm.

6. The method of claim 1, wherein the genomic region is from chromosome 1q41 and the sequence corresponds to HapMap positions 213267336 – 213447659.

7. The method of claim 1, wherein the sequence comprises SEQ ID NO: 2.

8. The method of claim 1, wherein the allele to be determined is selected from the group consisting of: rs2841599, rs17024179, rs12039875, rs11120508, rs10494994 and rs1538547.
9. The method of claim 1, wherein the allele to be determined is rs12039875 and C is the allele that correlates to risk of developing an abdominal aortic aneurysm.
10. A method of detecting an genetic predisposition to developing an abdominal aortic aneurysm in an individual, comprising analyzing a nucleic acid sample from the individual for a marker selected from the group consisting of rs7635818 and rs12039875, wherein the presence of a C allele in rs7635818 or a C allele in rs12039875 is indicative of a genetic predisposition to developing an abdominal aortic aneurysm.
11. A method of detecting an genetic predisposition to developing an abdominal aortic aneurysm in an individual, comprising:
- analyzing a nucleic acid sample to determine two alleles at rs7635818 and two alleles at rs12039875;
 - determining the number of risk alleles by adding the number of C alleles at rs7635818 and the number of C alleles at rs12039875;
 - wherein, if the individual is determined to have two or more risk alleles, the individual is indicated as having a genetic predisposition to developing an abdominal aortic aneurysm.
12. A method of detecting an genetic predisposition to developing an abdominal aortic aneurysm in an individual, comprising:
- analyzing a nucleic acid sample to determine two alleles at rs7635818, two alleles at rs12039875 and two alleles at rs10757278;

determining the number of risk alleles by adding the number of C alleles at rs7635818, the number of C alleles at rs12039875 and the number of G alleles at rs10757278;

wherein, if the individual is determined to have three or more risk alleles, the individual is indicated as having a genetic predisposition to developing an abdominal aortic aneurysm.

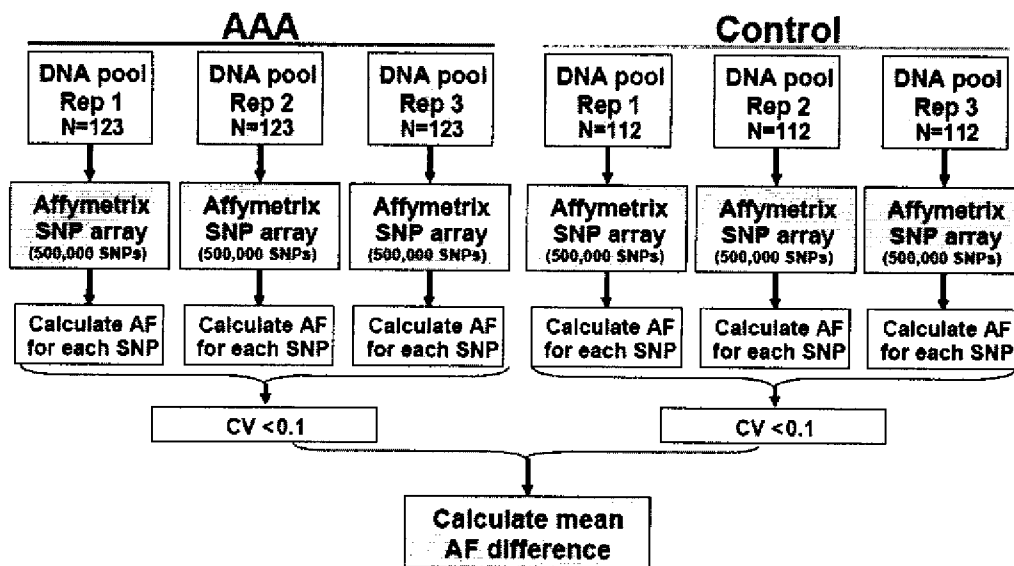


Figure 1

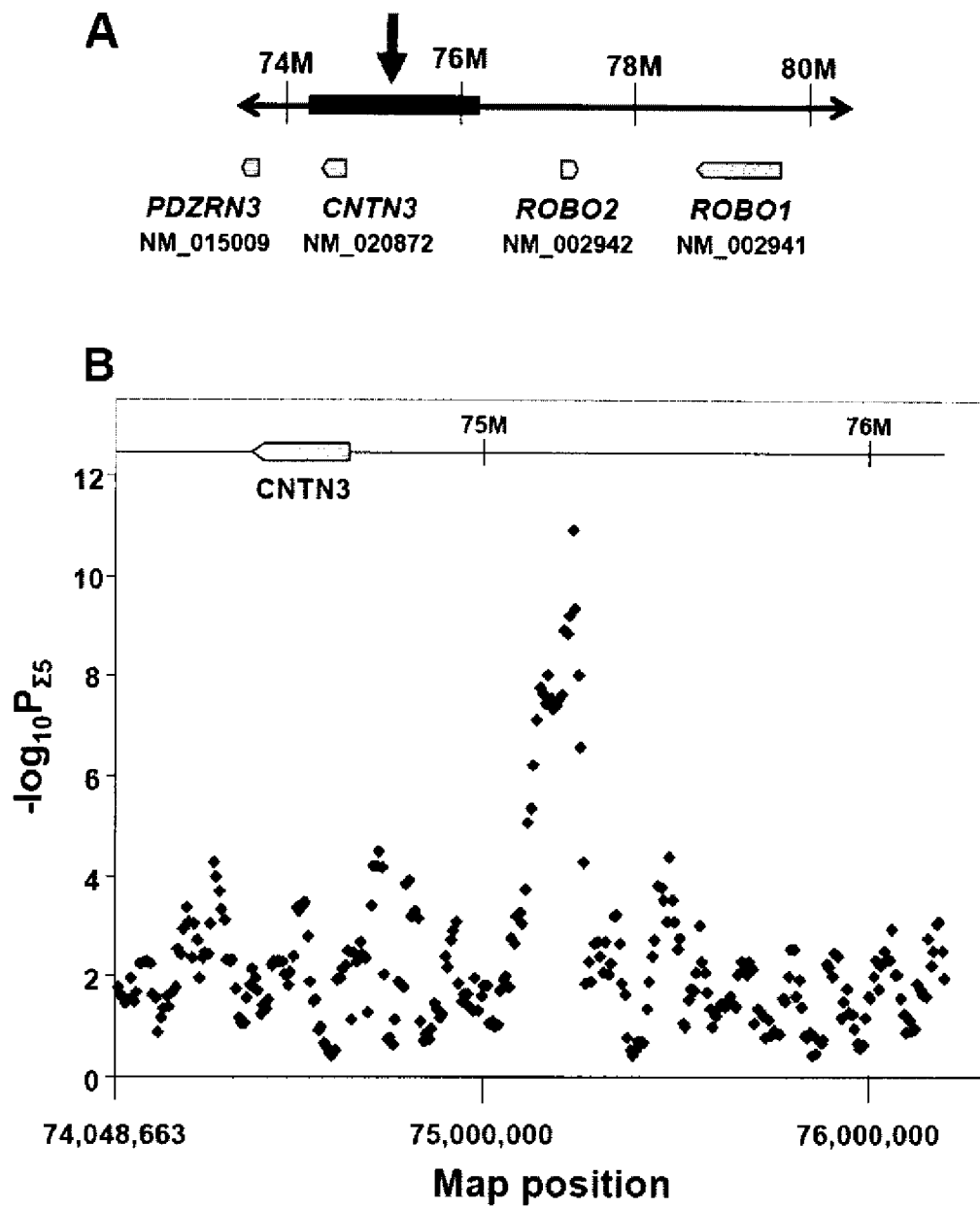
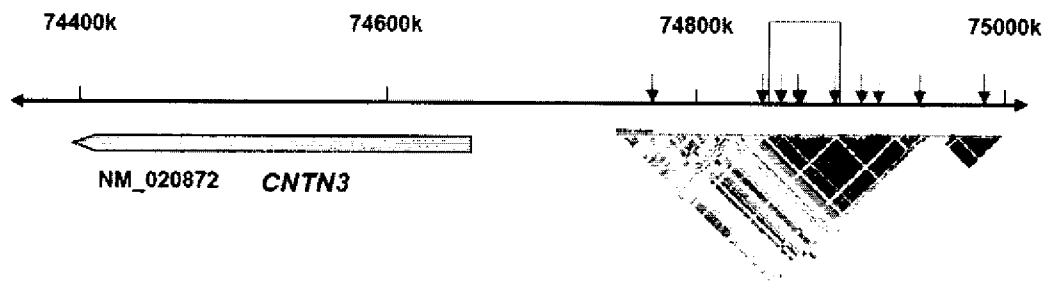


Figure 2



NCBI SNP Reference	HapMap Position	P-value
rs6549634	74762904	0.191
rs12714811	74841440	0.300
rs7642555	74852110	0.039
rs7635639	74866049	0.007
rs7635818	74866166	0.004
rs1569066	74881275	0.044
rs10222550	74902382	0.120
rs1527527	74919636	0.153
rs2178032	74943083	0.120
rs4417905	74998753	0.547

Figure 3

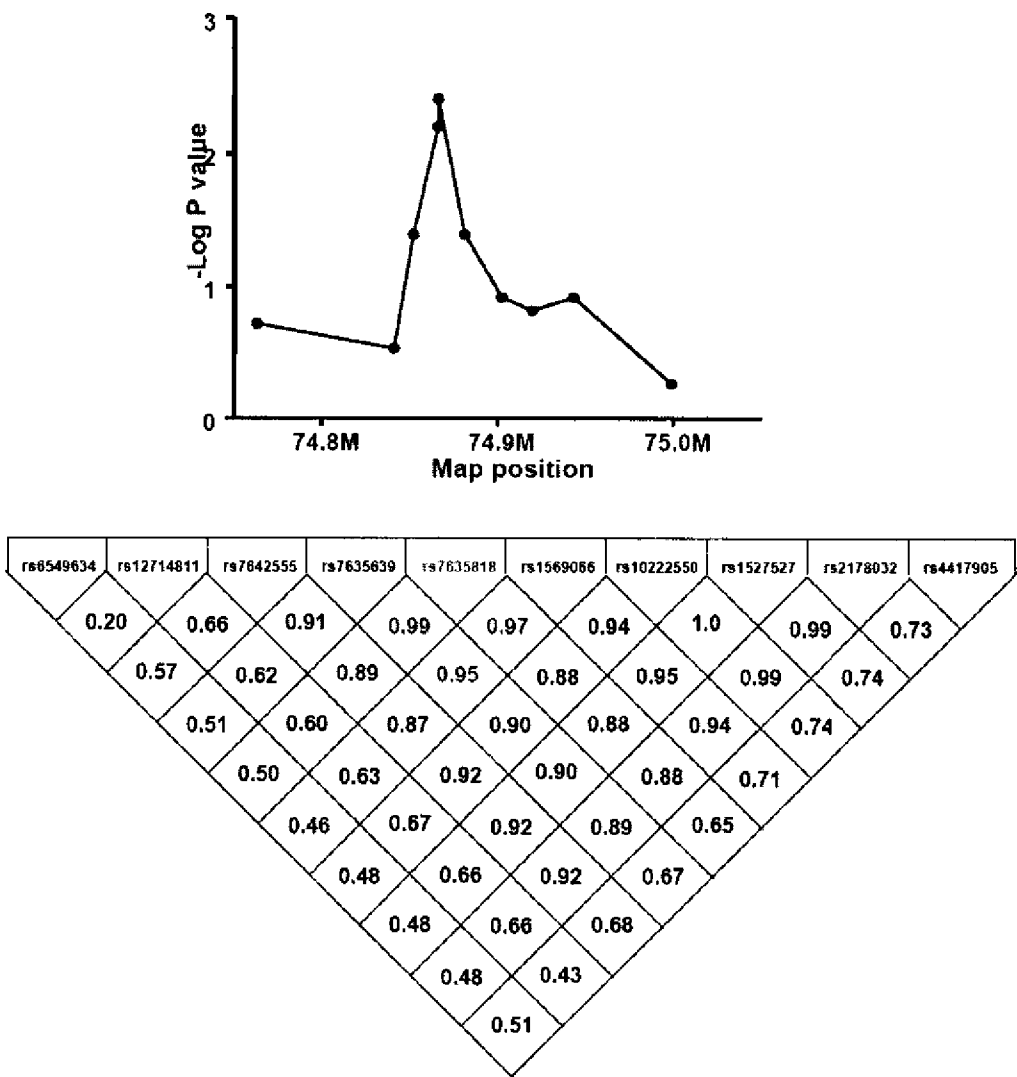


Figure 4

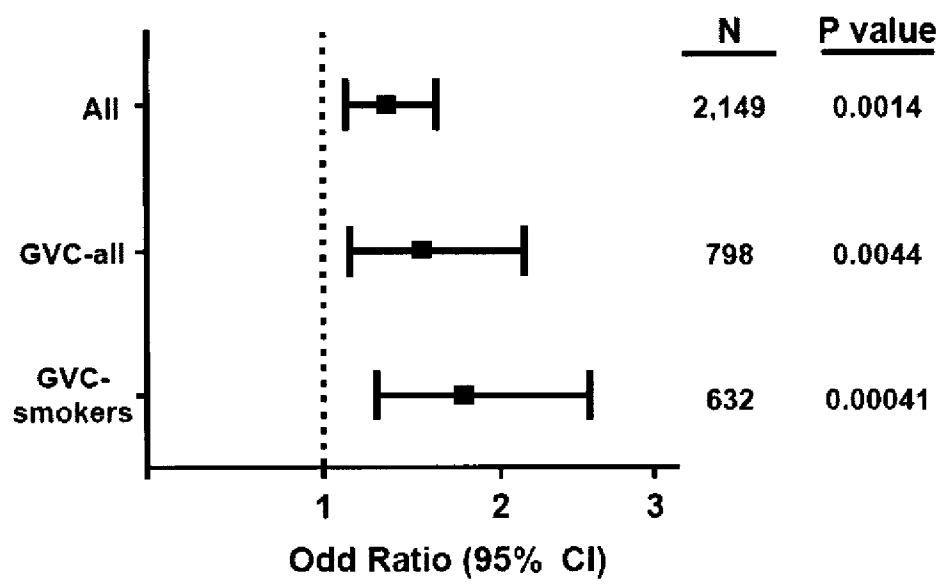


Figure 5

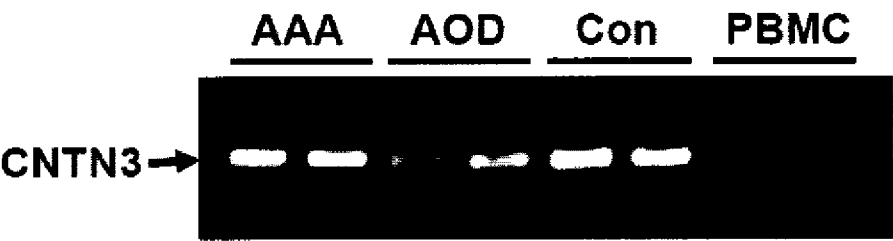


Figure 6

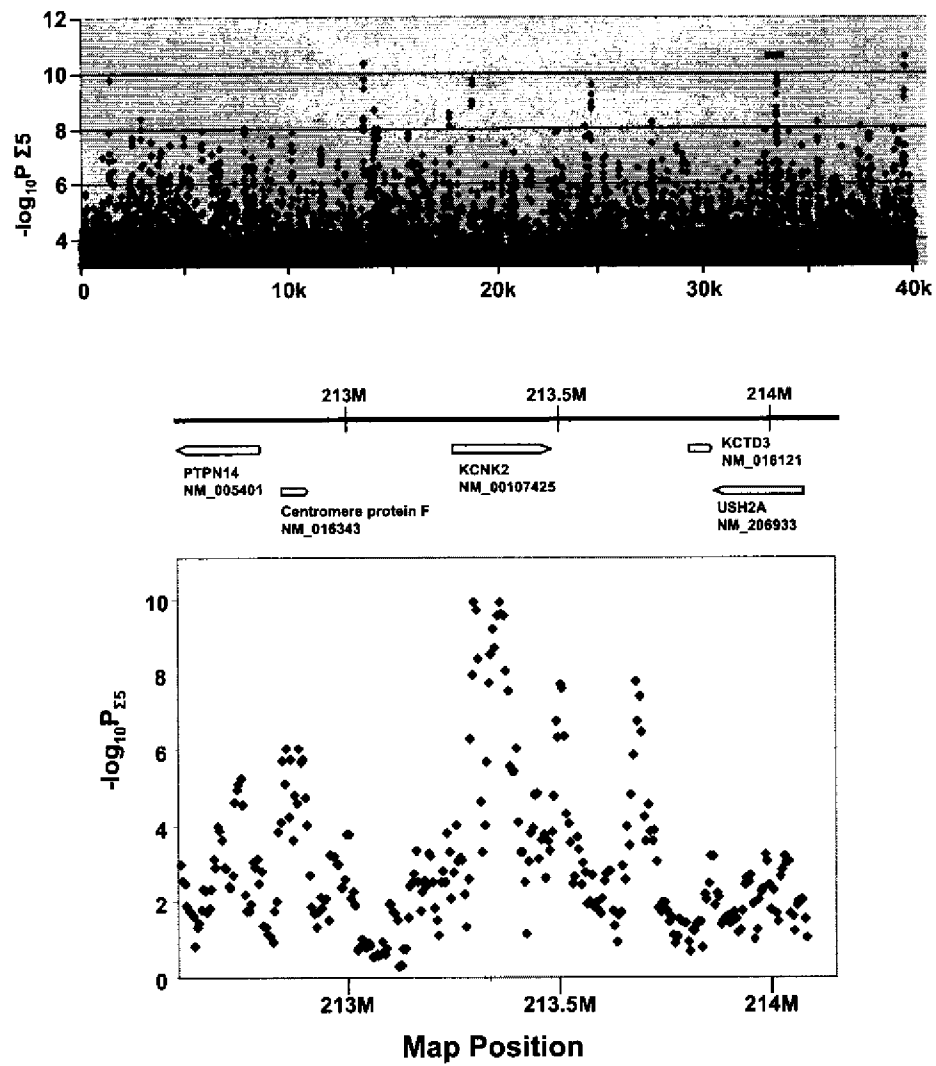


Figure 7

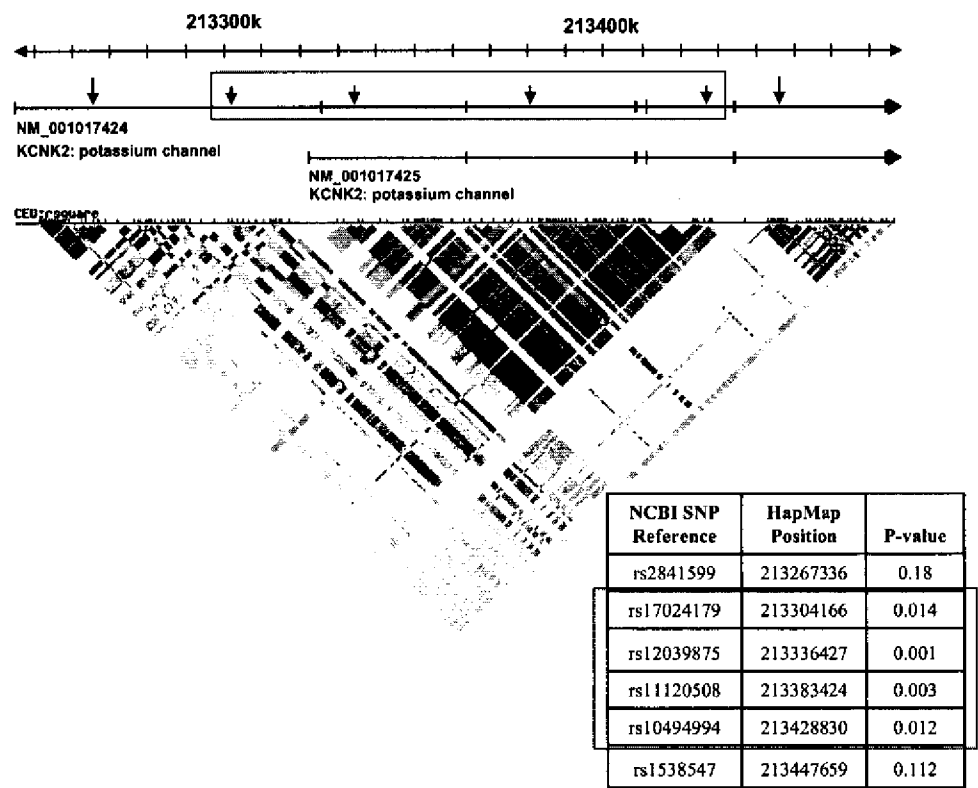


Figure 8

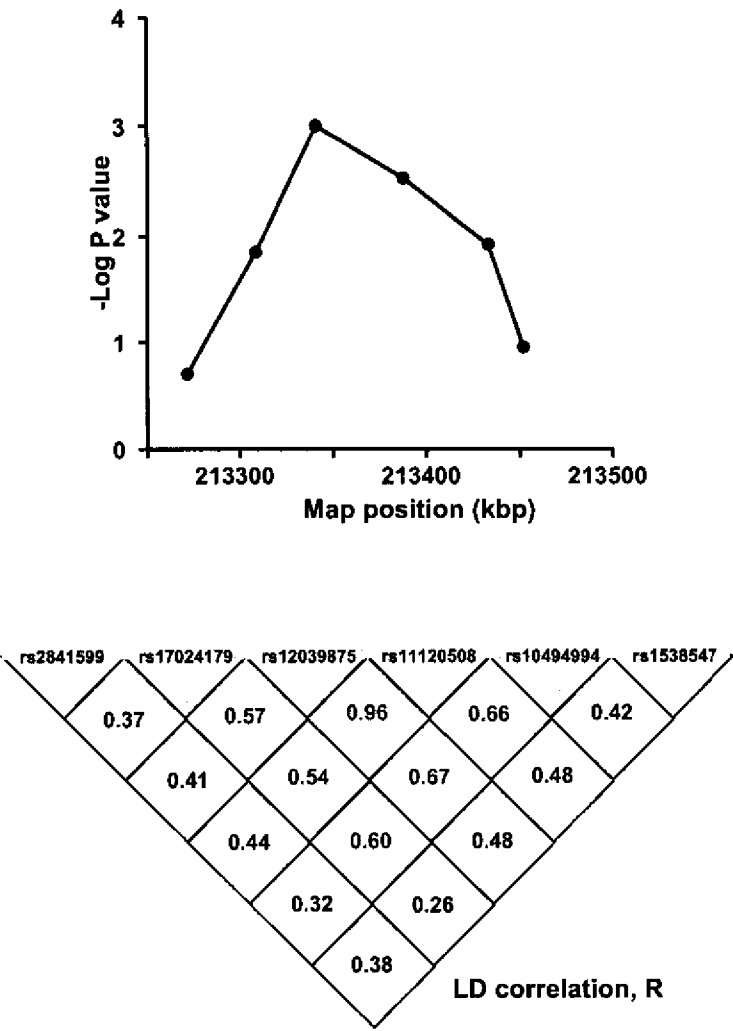
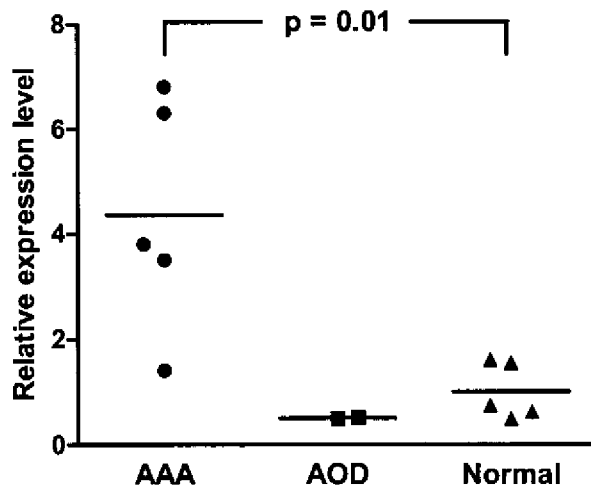
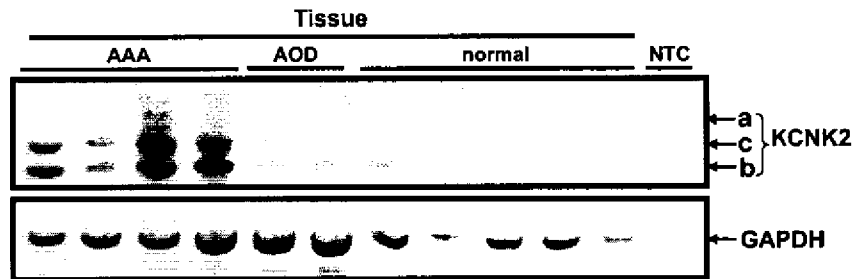


Figure 9

- a: MMNPRAKRDFYLAAPDLLDPKSAAQNSKPRLSFSTKPTVL
- b: MAAPDLLDPKSAAQNSKPRLSFSTKPTVL
- c: MLPSASRERPGYRAGVAAPDLLDPKSAAQNSKPRLSFSTKPTVL

**Figure 10**

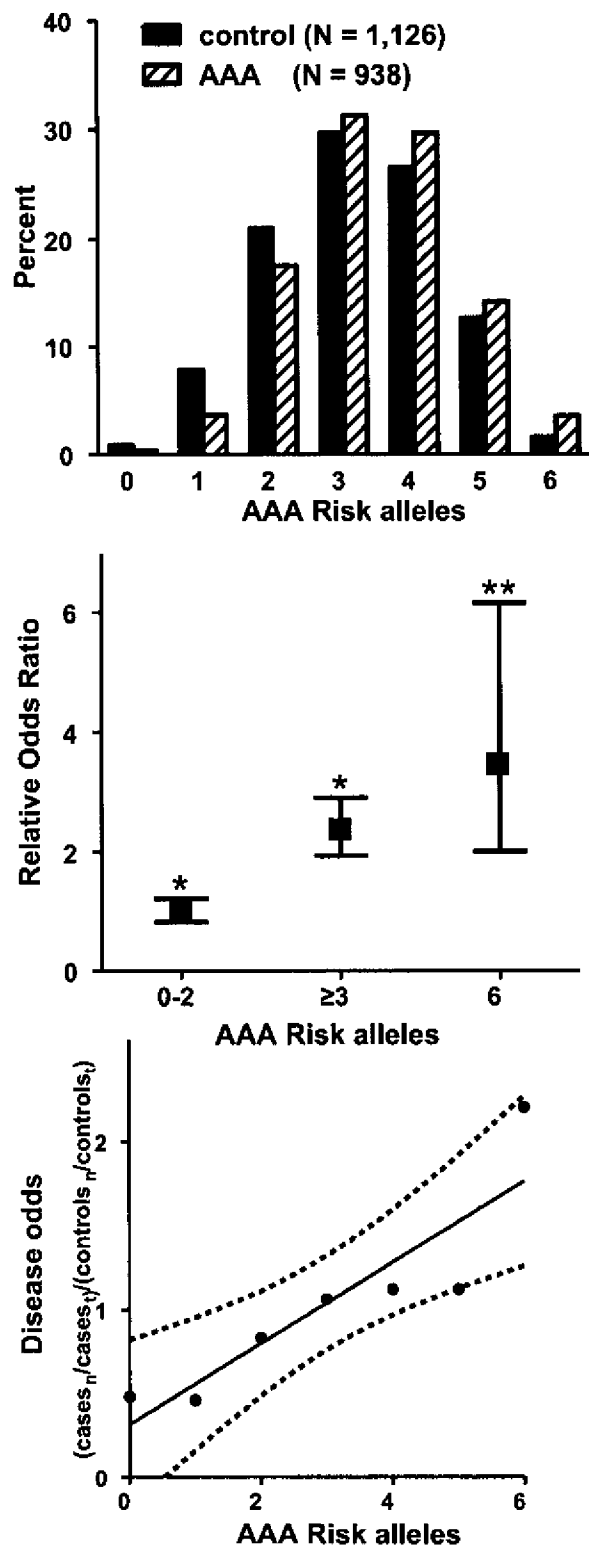


Figure 11