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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 February 2014 (27.02.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/028974 A1

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

(21) International Application Number:
PCT/AU2013/000931

(22) International Filing Date:
22 August 2013 (22.08.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2012903671 24 August 2012 (24.08.2012) AU

(71) Applicants: **THE UNIVERSITY OF QUEENSLAND** [AU/AU]; St Lucia, Queensland 4072 (AU). **UNIVERSITY OF BRISTOL** [GB/GB]; Senate House, Tyndall Avenue, Bristol BS8 1TH (GB).

(72) Inventors: **BROWN, Matthew Arthur**; 287A Kenmore Road, Fig Tree Pocket, Queensland 4069 (AU). **EVANS, David Meurig Faenor**; 26 Woonalee Street, Kenmore, Queensland 4069 (AU).

(74) Agents: **ARGAET, Victor P.** et al.; C/- Davies Collison Cave, Level 10, 301 Coronation Drive, Milton, Queensland 4064 (AU).

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

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

Published:

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

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

WO 2014/028974 A1

(54) Title: DIAGNOSTIC MARKERS FOR SPONDYLOARTHROPATHIES AND USES THEREOF

(57) Abstract: This present invention discloses methods and agents for diagnosing the presence or risk of development of spondyloarthropathies in mammals. The methods and agents are based on the detection of a polymorphic nucleotide in a polymorphic site having reference sequence number rs116488202 on chromosome 6. The invention also discloses methods for the treatment or prevention of spondyloarthropathies based on the diagnostic methods of the invention.

TITLE OF THE INVENTION

DIAGNOSTIC MARKERS FOR SPONDYLOARTHROPATHIES AND USES THEREOF

FIELD OF THE INVENTION

[0001] This application claims priority to Australian Provisional Application No. 2012903671 entitled "Diagnostic markers and uses therefor" filed 24 August 2012, the contents of which are incorporated herein by reference in their entirety.

[0002] This invention relates generally to methods and agents for determining the presence or risk of development of spondyloarthropathies in mammals. The methods and agents are based on the detection of a polymorphic nucleotide in a polymorphic site having reference sequence number rs116488202 on chromosome 6. The invention also features methods for the treatment or prevention of spondyloarthropathies based on the diagnostic methods of the present invention.

BACKGROUND OF THE INVENTION

[0003] The spondyloarthropathies (SpA) are a group of clinically and aetiologically related forms of inflammatory arthritis that include ankylosing spondylitis (AS), psoriatic arthritis, colitic arthritis (complicating inflammatory bowel disease (IBD)) and reactive arthritis. Collectively they are often termed the 'seronegative' arthropathies, to differentiate them from 'seropositive' diseases , as they are not characterized by the presence of rheumatoid factor. Spondyloarthropathies share in common that they typically affect the sacroiliac joints and spine, are associated with *HLA-B*27*, involve inflammation in the enthesis (attachment of tendons and ligaments to bone), and by the development of new bone at the site of inflammation, rather than erosion. These features are shared to a varying extent by the different forms of SpA (see Table 1).

TABLE 1

FEATURES OF DIFFERENT SPONDYLOARTHOPATHIES.

	Ankylosing Spondylitis	Reactive arthritis	Psoriatic Arthritis	Colitic arthritis (associated with IBD)
Male:Female ratio	2-3:1	5:1 (sexually acquired), 1:1 (post-diarrhoeal)	1:1	1:1

Peripheral arthritis	Uncommon, asymmetric lower limb oligoarthritis	Common, asymmetric lower limb oligoarthritis	Common, small and large joint, asymmetric	Common, asymmetric lower limb oligoarthritis
Enthesitis	Common	Very common	Common	Uncommon
Skin lesions	Nil (10% have psoriasis)	Circinate balinitis, keratoderma blennorrhagica	Psoriasis	Erythema nodosum, pyoderma gangrenosum
Sacroiliitis	Universal, symmetric	40-60%, often asymmetric	~20%, often asymmetric	~20%, often asymmetric
HLA-B*27 prevalence	90%	20-80%	80% with sacroiliitis, 8% without	80% with sacroiliitis, 8% without

[0004] A recent US study suggests that up to 2.4 million Americans suffer from SpA, compared with 1.3 million suffering from rheumatoid arthritis, making it as a group probably the most common form of inflammatory arthritis.

[0005] AS is the prototypic SpA. The condition principally affects the axial skeleton including the spine and sacroiliac joints, causing pain, stiffness, and eventually bony ankylosis. Peripheral joints and tendon insertions (entheses) are commonly affected, and approximately one-third of patients develop acute anterior uveitis. The prevalence of AS in people of white European ancestry is ~0.5%, and the condition is also common in most Asian populations. AS is rare in Africans and Australian aboriginals.

[0006] Although most frequently developing in young men, the male:female ratio is only 2-3:1, somewhat less extreme than originally thought. The typical age of onset is in the early 20s, and onset over 45 years of age is uncommon. It can develop in childhood, where the disease is more likely to affect feet and hips in addition to the pelvis and spine. Whilst AS can follow an infection where reactive arthritis has occurred, in most cases there is no precipitating factor.

[0007] AS runs very strongly in families, and most of the risk of developing AS is inherited (heritability >90%). Siblings or children of AS cases have about 12% likelihood of developing AS, and the likelihood of recurrent disease in more distant relatives is much lower (≤1%).

[0008] The major gene for AS, *HLA-B*27* (also referred to herein as "*B27*"), has been known for over 35 years, but we don't yet know how it causes the disease. Theories

range from differences in its capacity to present antigens, to problems caused by the formation of homodimers of B27 itself, or by its propensity to misfold in the endoplasmic reticulum.

[0009] Between 1-5% of *HLA-B*27* carriers develop AS, and in most populations world-wide, variation within B27 makes no difference to the likelihood of these individuals developing the condition (the exceptions are discussed below). There are now over 100 subtypes of *B27* known, but most are very rare. Of those that are sufficiently common to study, all are associated with disease except B*2709 which is found in Sardinia, but has not been identified in any primary AS case to-date in the absence of any other AS-associated HLA-B allele. B*2706, which is found in South-East Asia, is also under-represented in AS-cases, suggesting a protective effect. AS is endemic in nearly all populations where *B27* occurs, the exception being West Africa, where in some ethnic groups *B27* is as common as in Caucasians, but AS does not occur. *HLA-B*27* is rare in most African populations, where the development of seronegative arthritis like AS appears more closely related to HIV infection than genetic risk factors. In white European populations the two most common subtypes, HLA-B*2702 and –B*2705 are equally disease associated.

[0010] The strength of association of *HLA-B*27* with reactive arthritis is considerably lower than with primary AS, with most series reporting *B27*-carriage rates of 20-60%. Environmental triggers for the disease are primarily either sexually acquired bacterial infections (*Chlamydia*, ureoplasma) or enteric (*Salmonella*, *Shigella*, *Yersinia*, *Campylobacter*, *Clostridium* and *Vibrio* species). HIV-carriers have a higher risk of developing the condition, though they also have greater exposure to triggering infections. Whilst there is a male gender bias in sexually-acquired disease, enteric-acquired disease occurs with equal frequency amongst men and women.

[0011] It is widely believed that AS is triggered by enteric infection similar to the aetiology of enteric reactive arthritis. However the lack of epidemics of the condition, its worldwide distribution, and evidence from animal models which develop disease with exposure to normal enteric commensals, suggest that the infective trigger is likely to be ubiquitous. A variety of environmental factors is associated with greater severity of disease in AS, including cigarette smoking, socio-economic status and educational level. However these factors account for a minority of the variation in severity of disease, which is also known to be highly heritable.

[0012] Whilst *B27* is clearly the major gene in AS, twin and family studies indicate that other genes must be involved. Recently two new genes have been identified which contribute significantly to the risk of developing AS, *ERAP1* (*ARTS1*) and *IL23R*, as disclosed in WO 2008/144827. *ERAP1* encodes an aminopeptidase which is expressed in the endoplasmic reticulum, and which is involved both in preparing peptides for HLA Class I presentation to immune effector cells, and in cleaving the receptors for proinflammatory cytokines from the cell membrane, thereby downregulating their signaling. Which of these mechanisms is involved in its association with AS is unknown. *IL23R* encodes the receptor for the cytokine IL-23, which is involved in activation and differentiation of TH17 lymphocytes, a highly proinflammatory lymphocyte subset involved in mucosal immunity. Whilst a lot more research is required to determine the exact mechanisms by which IL-23R is involved in the development of AS, its association strongly suggests that disordered mucosal immunity is important in the disease, and also suggests that blockade of IL-23 and TH17 cytokines such as IL-17 are likely to be effective in AS. *IL23R* is also associated with psoriasis and both Crohn's disease and ulcerative colitis, and this genetic association at least partially explains why these conditions commonly co-occur.

[0013] The average diagnostic delay from onset of symptoms in AS is 8-11 years, even in countries with modern medical services such as Australia. The main reasons for this are that the condition is often not considered when young adults present with low back pain, and because early in the disease x-ray studies are often normal. Diagnostic delay is a major problem in the disease, as patients miss out on the potential for effective early treatments, not infrequently are treated inappropriately, and often lose faith in conventional medicine. This is at least part of the explanation why only a small proportion of AS cases seek ongoing medical care for their disease.

[0014] Studies in the United Kingdom have demonstrated that ~5% of adults attending general practitioners with low back pain have axial SpA as the underlying diagnosis. So what flags are there that should raise the question as to AS as the likely diagnosis? First, the diagnosis of AS has to be considered in all adults under the age of 45 years presenting with back or buttock pain to general practitioners. The diagnosis of AS is then more likely in the presence of:

[0015] • other features of spondyloarthritis (iritis, enthesitis (*e.g.* plantar fasciitis, Achilles tendinitis)),

[0016] • coexistent psoriasis or inflammatory bowel disease

[0017] • family history of AS

[0018] • prominent early morning stiffness (≥30 minutes duration)

[0019] • good response to NSAIDs

[0020] • lack of an obvious alternative cause, such as a significant back injury, especially where associated with sciatica.

[0021] The absence of these features is not particularly helpful, but their presence increases the likelihood of AS.

[0022] Having considered the diagnosis, confirming it involves a combination of genetic tests and diagnostic imaging. Patients at this point should be referred to a rheumatologist for a specialist opinion. Physical measures, such as spinal movement measures including the Schober's test, have little value in early disease other than to help exclude alternate potential diagnoses.

[0023] Currently the only genetic test that is publically available is *HLA-B*27*. More than 90% of AS cases are *HLA-B*27* positive, compared with 8% of healthy individuals of western European ethnicity. The likelihood of the diagnosis of AS given the findings of an *HLA-B*27* test depends on the prior likelihood of disease, and the test result. In the general practice setting of an adult patient under 45 years of age presenting with low back pain, the prior probability of AS is 5%. If the *HLA-B*27* test is negative, the probability of AS is low (1%), but not completely excluded, particularly if the patient has either psoriasis or inflammatory bowel disease, where a higher proportion of cases are *HLA-B*27* negative. The probability of AS if the test is positive is moderate, but not established (37%).

[0024] Recently, a single imputed single nucleotide polymorphism (SNP), rs4349859, was identified which closely tags *HLA-B*27*, and in turn, AS affection status (*see*, WO2011/063474). This SNP strongly associates with AS in western Europeans but it associates less so for Asian populations.

[0025] From the foregoing, there is a recognized need for effective genetic markers for determining the presence or risk of development of SpA, including AS. It would be highly advantageous to have a reliable screening method to enable better treatment and management decisions to be made in subjects with SpA or a predisposition to developing SpA.

SUMMARY OF THE INVENTION

[0026] The present invention is predicated in part on the discovery that a polymorphism in a polymorphic site having reference sequence number rs116488202 on chromosome 6 is a surrogate marker for spondyloarthropathies (also referred to herein as "a SpA marker") such as reactive arthritis, ankylosing spondylitis, anterior uveitis, colitic arthritis, psoriasis and enteropathic arthritis associated with inflammatory bowel disease. Surprisingly, this surrogate marker tags *HLA-B*27* in both western European and Asian populations better than rs4349859. Accordingly, in one aspect, the present invention provides methods for determining the presence, absence or risk of development of a spondyloarthropathy in a subject. In specific embodiments, these methods comprise analyzing a biological sample of the subject, which comprises at least a portion of chromosome 6 comprising a polymorphic site having reference sequence number rs116488202, for a polymorphism in the polymorphic site, which is indicative of the presence or risk of development of a spondyloarthropathy in the subject. In certain embodiments, the methods comprise (a) obtaining from the subject a biological sample comprising at least a portion of chromosome 6, which comprises the polymorphic site; and (b) analyzing the sample for the polymorphism. In some embodiments, the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site having reference sequence number rs116488202, wherein the presence of T (thymine) at rs116488202, indicates that the subject has a SpA or has an increased risk of developing a SpA. In other embodiments, the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site having reference sequence number rs116488202, wherein the presence of C (cytosine) at rs116488202, indicates that the subject lacks a SpA or has a lower risk of developing a SpA. Suitably, the SpA is selected from reactive arthritis, ankylosing spondylitis (AS), anterior uveitis, colitic arthritis, psoriasis and enteropathic arthritis. In specific embodiments, the SpA is AS.

[0027] In some embodiments, the methods further comprise determining the presence, absence or risk of developing a specific SpA based on one or more ancillary diagnostic parameters (*e.g.*, clinical features of the disease, blood test results such as the level of the erythrocyte sedimentation rate or C-reactive protein, and results of one or more genetic tests or gene-expression profiling tests).

[0028] In some embodiments, the SpA is AS and the methods further comprise analyzing the sample for the presence of a polymorphism in at least one AS marker (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more). In illustrative examples, the at least one AS marker may be selected from: an *ERAP1* (formerly known as *ARTS1*) gene, an *IL23R* gene, a *TNFR1* gene locus, a *TRADD* gene locus, a 2P15 chromosomal locus, a 21Q22 chromosomal locus and a *HLA-B* gene, as disclosed for example in International Application No. PCT/AU2008/000762 filed 29 May 2008 (WO 2008/144827); and/or an *IL1R1* gene locus, an *IL1R2* gene locus and the chromosome locus 2Q31.3, as disclosed for example in International Application No. PCT/AU2009/01320 filed 2 October 2009 (WO 2010/037184).

[0029] Accordingly, in some embodiments, the sample is further analyzed for the presence of a polymorphism in the *ERAP1* gene, wherein the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic site (*e.g.*, 1, 2, 3, 4, 5 polymorphic site(s)) within the *ERAP1* gene, having a reference sequence number on chromosome 5 selected from the group consisting of rs27044, rs17482078, rs10050860, rs30187 and rs2287987. In illustrative examples of this type, the presence of: G (guanine) at rs27044; T (thymine) at rs30187; T at rs2287987; C (cytosine) at rs17482078; and/or C at rs10050860 indicates that the subject has AS or has an increased risk of developing AS. In other illustrative examples, the presence of: C at rs27044; C at rs30187; C at rs2287987; T at rs17482078; and/or T at rs10050860 indicates that the subject lacks AS or has a lower risk of developing AS.

[0030] The polymorphism at rs27044 is associated with a non-synonymous variation in the coding sequence of the *ERAP1* gene, wherein the presence of G at rs27044 is associated with Gln (glutamine) at position 730 of the ERAP-1 polypeptide (as set forth for example in GenPept Accession No. NP_057526 [GI:94818901] or SEQ ID NO: 2 of WO 2008/144827), and wherein the presence of C at rs27044 is associated with Glu (glutamate) at position 730 of the ERAP-1 polypeptide. The polymorphism at rs17482078 is associated with a non-synonymous variation in the coding sequence of the *ERAP1* gene, wherein the presence of C at rs17482078 is associated with Arg (arginine) at position 725 of the ERAP-1 polypeptide, and wherein the presence of T at rs17482078 is associated with Gln at position 725 of the ERAP-1 polypeptide. The polymorphism at rs10050860 is associated with a non-synonymous variation in the coding sequence of the *ERAP1* gene, wherein the presence of C at rs10050860 is associated with Asp (aspartic acid) at position 575 of the ERAP-1 polypeptide, and wherein the presence of T at rs10050860 is associated with Asn

(asparagine). The polymorphism at rs2287987 is associated with a non-synonymous variation in the coding sequence of the *ERAP1* gene, wherein the presence of T at rs2287987 is associated with Met (methionine) at position 349 of the ERAP-1 polypeptide, and wherein the presence of C at rs2287987 is associated with Val (valine). The polymorphism at rs30187 is associated with a non-synonymous variation in the coding sequence of the *ERAP1* gene, wherein the presence of T at rs30187 is associated with Lys (lysine) at position 528 of the ERAP-1 polypeptide, and wherein the presence of C at rs30187 is associated with Arg.

[0031] Accordingly, in some embodiments, the analysis comprises determining the identity of an amino acid at one or more positions (*e.g.*, 1, 2, 3, 4, 5) of an ERAP-1 polypeptide selected from positions 730, 725, 575, 349 and 528 of the ERAP-1 polypeptide, wherein the presence of: Gln at position 730; Arg at position 725; Asp at position 575; Met at position 349; and/or Lys at position 528 indicates that the subject has AS or has an increased risk of developing AS. In other embodiments, the analysis comprises determining the identity of an amino acid at one or more positions (*e.g.*, 1, 2, 3, 4, 5) of an ERAP-1 polypeptide selected from positions 730, 725, 575, 349 or 528 of the ERAP-1 polypeptide, wherein the presence of: Glu at position 730; Gln at position 725; Asn at position 575; Val at position 349; and/or Arg at position 528 indicates that the subject lacks AS or has a lower risk of developing AS.

[0032] In some embodiments, the sample is further analyzed for the presence of a polymorphism in the *IL23R* gene, wherein the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic site (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8 polymorphic site(s)) within the *IL23R* gene having a reference sequence number on chromosome 1 selected from the group consisting of rs1004819, rs10489629, rs11465804, rs11209026, rs1343151, rs10889677, rs11209032 and rs1495965. In representative examples of this type, the presence of: T at rs11465804; T at rs10489629; G at rs11209026; G at rs1343151; C at rs1495965; A (adenine) at rs1004819; A at rs10889677; and/or A at rs11209032, indicates that the subject has AS or has an increased risk of developing AS. In other representative examples, the presence of: C at rs10489629; C at rs10889677; A at rs1343151; A at rs11209026; G at rs11209032; G at rs1004819; G at rs11465804; and/or T at rs1495965 indicates that the subject lacks AS or has a lower risk of developing AS.

[0033] The polymorphism at rs11209026 is associated with a non-synonymous variation in the coding sequence of the *IL23R* gene, wherein the presence of G at rs11209026

is associated with Arg at position 381 of the IL-23R polypeptide (as set forth for example in GenPept Accession No. NP_653302 [GI:24430212] or SEQ ID NO: 4 WO 2008/144827), and wherein the presence of A at rs11209026 is associated with Gln at position 381 of the IL-23R polypeptide. Accordingly, in some embodiments, the analysis comprises determining the identity of an amino acid at position 381 of an IL-23R polypeptide, wherein the presence of Arg at position 381 indicates that the subject has AS or has an increased risk of developing AS, and wherein the presence of Gln at position 381 indicates that the subject lacks AS or has a lower risk of developing AS.

[0034] In some embodiments, the sample is further analyzed for the presence of a polymorphism in the *TNFR1* gene locus, wherein the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site within the *TNFR1* gene locus, having reference sequence number rs4149576 on chromosome 12. In representative examples of this type, the presence of C at rs4149576 indicates that the subject has AS or has an increased risk of developing AS. In other representative examples, the presence of T at rs4149576 indicates that the subject lacks AS or has a lower risk of developing AS.

[0035] In some embodiments, the sample is further analyzed for the presence of a polymorphism in the *TRADD* gene locus, wherein the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site within that locus, having reference sequence number rs9033 on chromosome 16. In illustrative examples of this type, the presence of G at rs9033 indicates that the subject has AS or has an increased risk of developing AS. In other illustrative examples, the presence of A at rs9033 indicates that the subject lacks AS or has a lower risk of developing AS.

[0036] In some embodiments, the sample is further analyzed for the presence of a polymorphism in the 2P15 chromosomal locus. In non-limiting examples, the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic site within the 2P15 chromosome locus having a reference sequence number rs10865331 on chromosome 2. Suitably, the presence of A at rs10865331, indicates that the subject has AS or has an increased risk of developing AS. In other examples, the presence of G at rs10865331, indicates that the subject lacks AS or has a lower risk of developing AS.

[0037] In some embodiments, the sample is further analyzed for the presence of a polymorphism in the 21Q22 chromosomal locus. In illustrative examples, the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic

site within the 21Q22 chromosome locus having reference sequence number rs2242944 on chromosome 21. Suitably, the presence of G at rs2242944, indicates that the subject has AS or has an increased risk of developing AS. In other examples, the presence of A at rs2242944, indicates that the subject lacks AS or has a lower risk of developing AS.

[0038] In specific embodiments, the sample is further analyzed for the presence an *HLA-B*27* allele of the *HLA-B* gene.

[0039] In some embodiments, the sample is further analyzed for the presence of a polymorphism in the *IL1R1* gene locus. Suitably, the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site within the *IL1R1* gene locus, having reference sequence number rs949963 on chromosome 2. In illustrative examples of this type, the presence of C at rs949963, indicates that the subject has AS or has an increased risk of developing AS. In other illustrative examples, the presence of T at rs949963, indicates that the subject lacks AS or has a lower risk of developing AS.

[0040] In some embodiments, the sample is further analyzed for the presence of a polymorphism in the *IL1R2* gene locus. Generally, the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site within the *IL1R2* gene locus, having reference sequence number rs2310173 on chromosome 2. In representative examples of this type, the presence of T at rs2310173, indicates that the subject has AS or has a risk of developing AS. In other representative examples, the presence of G at rs2310173, indicates that the subject lacks AS or has a lower of developing AS.

[0041] In some embodiments, the sample is further analyzed for the presence of a polymorphism in the 2Q31.3 chromosome locus. Generally, the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site within the 2Q31.3 chromosome locus, having reference sequence number rs1018326 on chromosome 2. In illustrative examples of this type, the presence of C at rs1018326, indicates that the subject has AS or has an increased risk of developing AS. In other illustrative examples, the presence of T at rs1018326, indicates that the subject lacks AS or has a lower risk of developing AS. As disclosed in WO 2010/037184, the rs1018326 polymorphic site is considered to be in genetic linkage with the *UBE2E3* gene and consequently, in some embodiments, the analysis for the presence of a polymorphism in the 2Q31.3 chromosome locus comprises analyzing the presence of a polymorphism in the *UBE2E3* gene.

[0042] In certain embodiments, it is desirable to analyze the sample for the presence of a polymorphism in the SpA marker as broadly described above and for the presence of a polymorphism in at least 1, 2, 3, 4, 5, 6, 7, 8, 9 or all 10 AS markers as broadly described above, which are indicative of the presence, absence or risk of development of AS.

[0043] In some embodiments, the methods further comprise determining whether the subject is homozygous or heterozygous for a polymorphic nucleotide in an individual polymorphic site. In representative examples of this type, the presence of an homozygous genotype for the SpA- or AS-associating polymorphic nucleotide indicates that the subject has a higher risk of developing the SpA or AS than an heterozygous genotype for the SpA- or AS-associating polymorphic nucleotide.

[0044] The polymorphism can be detected by any method known in the art including, but not limited to: polymerase chain reaction (PCR), ligase chain reaction (LCR), hybridization analysis, digestion with nucleases, restriction fragment length polymorphism, antibody detection methods, direct sequencing or any combination thereof.

[0045] In certain embodiments, the presence, absence or risk of development of an SpA in the subject is determined from the subject's SpA marker genotype. A subject who has at least one polymorphism statistically associated with an SpA (*e.g.*, AS) possesses a factor contributing to an an increased risk of the SpA, as compared to a subject without the polymorphism.

[0046] In another aspect, the present invention contemplates the use of a nucleic acid construct comprising at least a portion of a SpA marker as broadly described above, which contains at least one SpA-associated polymorphism, for determining the presence, absence or risk of development of the SpA. In some embodiments, the at least a portion of the SpA marker is operably connected to a regulatory element, which is operable in a host cell. In certain embodiments, the construct is in the form of a vector, especially an expression vector. In illustrative examples of this type, the vector is used as a positive control.

[0047] In yet another aspect, the present invention contemplates the use of isolated host cells containing a nucleic acid construct or vector as broadly described above for diagnosing the presence or risk of development of a SpA. In certain embodiments, the host cells are selected from bacterial cells, yeast cells and insect cells. In some embodiments, the host cells are used for the production of at least one transcript comprising at least a portion of the SpA marker as broadly described above, suitably for use as a positive control.

[0048] In embodiments in which the SpA is AS and the methods of the present invention are thus directed to diagnosing the presence, absence or risk of development of AS, the present invention further encompasses the use in those methods of a nucleic acid construct comprising at least a portion of an AS marker as broadly described above, which contains at least one AS-associated polymorphism. In some embodiments, the at least a portion of the AS marker is operably connected to a regulatory element, which is operable in a host cell. In certain embodiments, the construct is in the form of a vector, especially an expression vector. In illustrative examples of this type, the vector is used as a positive control.

[0049] In certain embodiments in which the SpA is AS and the methods of the present invention are directed to diagnosing the presence or risk of development of AS, the present invention further encompasses host cells for producing at least one polypeptide selected from IL-1R1, IL-1R2, CD74, TCOF1, UBE2E3, ERAP-1 and IL-23R polypeptides for use as a positive control. Suitably, the polypeptide(s) are fragmented and analysed using mass spectrometry techniques.

[0050] Still another aspect of the present invention relates to the use of one or more oligonucleotides that hybridize to a SpA-associated polymorphic site in the SpA marker as broadly described above in the manufacture of a kit for detecting the presence or diagnosing the risk of development of a SpA (*e.g.*, reactive arthritis, AS, anterior uveitis, colitic arthritis, psoriasis and enteropathic arthritis associated with inflammatory bowel disease). The kit may comprise one or more oligonucleotides capable of detecting a polymorphism in the SpA marker of the invention as well as instructions for using the kit to detect a SpA or to diagnose the risk of developing a SpA. In some embodiments, the oligonucleotides each comprise a sequence that hybridizes under stringent hybridization conditions to a SpA-associated polymorphism in the SpA marker as broadly described above. In some embodiments, the oligonucleotides are allele-specific oligonucleotides.

[0051] In certain embodiments in which the SpA is AS, the present invention further encompasses the use of one or more oligonucleotides that hybridize to at least one AS-associated polymorphic site in an AS marker as broadly described above in the manufacture of a kit for detecting the presence or diagnosing the risk of development of AS. The kit may comprise one or more oligonucleotides capable of detecting a polymorphism in an AS marker of the invention as well as instructions for using the kit to detect AS or to diagnose the risk of developing AS. In some embodiments, the oligonucleotides each comprise a sequence that

hybridizes under stringent hybridization conditions to at least one AS-associated polymorphism in any one or more of the AS markers as broadly described above. Alternatively, or in addition, the kits may comprise an antigen-binding molecule that is immuno-interactive with an AS-associated polymorphism as broadly described above.

[0052] The invention further provides methods for treating a SpA (*e.g.*, reactive arthritis, AS, anterior uveitis, colitic arthritis, psoriasis and enteropathic arthritis associated with inflammatory bowel disease) in a subject. These methods generally comprise analysing a biological sample obtained from the subject for the presence of a SpA-associated polymorphism in the SpA marker as broadly described above and exposing the subject to a treatment that ameliorates or reverses the symptoms of the SpA on the basis that the subject tests positive for the polymorphism.

[0053] In certain embodiments in which the SpA is AS, the present invention further provides methods for treating AS in a subject, wherein the methods generally comprise analysing a biological sample obtained from the subject for the presence of a SpA-associated polymorphism in the SpA marker as broadly described above and optionally for the presence of at least one AS-associated polymorphism in an AS marker as broadly described above and exposing the subject to a treatment that ameliorates or reverses the symptoms of AS on the basis that the subject tests positive for the SpA-associated polymorphism and optionally for the AS-associated polymorphism(s).

BRIEF DESCRIPTION OF THE DRAWINGS

[0054] Figure 1 is a graphical representation showing evidence of association to AS susceptibility in samples of European ancestry in the MHC region. Associations over the MHC are shown correcting for ancestry with five principal components without and with conditioning on *HLA-B*27* tagSNP rs116488202. SNS with P-value < 10^{-320} are set P-value = 1×10^{-320}.

TABLE A

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 6, comprising the polymorphism rs116488202 (chr 6: 31,344,716-31,345,116) – NCBI reference sequence: NC_000006.11 [GI:224589818]	401 nts
SEQ ID NO: 2	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 5, comprising the polymorphism rs27044 (chr 5: 96144307-96144908) within the *ERAP1* gene - NCBI reference sequence: NC_000005.9 [GI:224589817]	601 nts
SEQ ID NO: 3	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 5, comprising the polymorphism rs17482078 (chr 5: 96144321-96144922) within the *ERAP1* gene - NCBI reference sequence: NC_000005.9 [GI:224589817]	601 nts
SEQ ID NO: 4	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 5, comprising the polymorphism rs30187 (chr 5: 96149785-96150386) within the *ERAP1* gene - NCBI reference sequence: NC_000005.9 [GI:224589817]	601 nts
SEQ ID NO: 5	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 5, comprising the polymorphism rs10050860 (chr 5: 96,147,665-96,148,266) within the *ERAP1* gene - NCBI reference sequence: NC_000005.9 [GI:224589817]	601 nts

SEQ ID NO: 6	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 5, comprising the polymorphism rs2287987 (chr 5: 96,154,990-96,155,591) within the *ERAP1* gene - NCBI reference sequence: NC_000005.9 [GI:224589817]	601 nts
SEQ ID NO: 7	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 1, comprising the polymorphism rs10489629 (chr 1: 67460937-67460937) within the *IL23R* gene - NCBI reference sequence: NC_000001.10 [GI:224589800]	201 nts
SEQ ID NO: 8	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 1, comprising the polymorphism rs1343151 (chr 1: 67491416-67492017) within the *IL23R* gene - NCBI reference sequence: NC_000001.10 [GI:224589800]	601 nts
SEQ ID NO: 9	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 1, comprising the polymorphism rs10889677 (chr 1: 67497407-67498008) within the *IL23R* gene - NCBI reference sequence: NC_000001.10 [GI:224589800]	601 nts
SEQ ID NO: 10	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 1, comprising the polymorphism rs11209032 (chr 1: 67,512,379-67,512,980) within the *IL23R* gene - NCBI reference sequence: NC_000001.10 [GI:224589800]	601 nts

SEQ ID NO: 11	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 1, comprising the polymorphism rs1004819 (chr 1: 67,442,312-67,443,004) within the *IL23R* gene - NCBI reference sequence: NC_000001.10 [GI:224589800]	692 nts
SEQ ID NO: 12	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 1, comprising the polymorphism rs11465804 (chr 14: 67,475,013-67,475,214) within the *IL23R* gene - NCBI reference sequence: NC_000001.10 [GI:224589800]	201 nts
SEQ ID NO: 13	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 1, comprising the polymorphism rs11209026 (chr 1: 67,478,245-67,478,846) within the *IL23R* gene - NCBI reference sequence: NC_000001.10 [GI:224589800]	601 nts
SEQ ID NO: 14	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 1, comprising the polymorphism rs1495965 (chr 1: 67,525,837-67,526,485) within the *IL23R* gene - [NCBI reference sequence: NC_000001.10 GI:224589800]	648 nts
SEQ ID NO: 15	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 12, comprising the polymorphism rs4149576 (chr 12: 6319075-6319676) within the *TNFR1* gene locus - NCBI reference sequence: NC_000012.11 [GI:224589803]	601 nts

SEQ ID NO: 16	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 16, comprising the polymorphism rs9033 (chr 16: 65,739,199- 65,739,799) within the *TRADD* gene locus - NCBI reference sequence: NC_000016 [GI:224589807]	600 nts
SEQ ID NO: 17	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 2, comprising the polymorphism rs10865331 (chr 2: 62,550,830-62,552,313) within the 2P15 chromosome locus - NCBI reference sequence: NT_022184.15 [GI:224515010]	1484 nts
SEQ ID NO: 18	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 21, comprising the polymorphism rs2242944 (chr 21: 40,461,078-40,462,078) within the 21Q22 chromosome locus - NCBI reference sequence: NT_022184.15 [GI:224514633]	1001 nts
SEQ ID NO: 19	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 2, comprising the polymorphism rs949963 (chr 2: 102,136,017-102,136,718) within the *IL1R1* gene locus - NCBI reference sequence: AF531102 [GI:22001412]	701 nts
SEQ ID NO: 20	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 2, comprising the polymorphism rs2310173 (chr 2: 102,029,560-102,030,511) within the *IL1R2* gene locus - NCBI reference sequence: AC007165 [GI:19033999]	951 nts

SEQ ID NO: 21	Nucleotide sequence corresponding to a portion of the genomic sequence of human chromosome 2, comprising the polymorphism rs1018326 (chr 2: 181,715,744-181,716,345) in the *UBE2E3* gene that is in genetic linkage with the 2Q31.3 chromosome locus - NCBI reference sequence: NC_000002 [GI:89161199]	601 nts

DETAILED DESCRIPTION OF THE INVENTION

1. Definitions

[0055] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, preferred methods and materials are described. For the purposes of the present invention, the following terms are defined below.

[0056] The articles "a" and "an" are used herein to refer to one or to more than one (*i.e.* to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

[0057] "Allele" is used herein to refer to a variant of a gene found at the same place or locus of a chromosome.

[0058] "Amplification product" refers to a nucleic acid product generated by nucleic acid amplification techniques.

[0059] By "antigen-binding molecule" is meant a molecule that has binding affinity for a target antigen. It will be understood that this term extends to immunoglobulins, immunoglobulin fragments and non-immunoglobulin derived protein frameworks that exhibit antigen-binding activity.

[0060] The term "biological sample" as used herein refers to a sample that may be extracted, untreated, treated, diluted or concentrated from a patient. Suitably, the biological sample is selected from any part of a patient's body, including, but lot limited to hair, skin, nails, tissues or bodily fluids such as saliva and blood.

[0061] Throughout this specification, unless the context requires otherwise, the words "comprise," "comprises" and "comprising" will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements.

[0062] By "corresponds to" or "corresponding to" is meant (a) a polynucleotide having a nucleotide sequence that is substantially identical or complementary (*e.g.*, at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 97%, 88%, 89%, 90%,

91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% to all or a portion of a reference polynucleotide sequence.

[0063] The term "diagnosis" as used herein refers to methods by which the skilled artisan can estimate and/or determine the probability ("a likelihood") of whether (*e.g.*, a "high likelihood" or "increased likelihood") or not (*e.g.*, a "low likelihood" or "reduced likelihood") a subject is suffering from a given disease or condition. In the case of the present invention, "diagnosis" includes using the results of an assay/method that determines the identity of a polymorphic nucleotide in a polymorphic site, optionally together with other clinical characteristics, to arrive at a diagnosis (that is, the occurrence or nonoccurrence) for determining the presence, absence or risk of development of a spondyloarthropathy in a subject. That such a diagnosis is "determined" is not meant to imply that the diagnosis is 100% accurate.

[0064] By "effective amount," in the context of treating or preventing a condition is meant the administration of that amount of active to an individual in need of such treatment or prophylaxis, either in a single dose or as part of a series, that is effective for treatment of, or prophylaxis against, that condition. The effective amount will vary depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated, the formulation of the composition, the assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

[0065] As used herein, the terms "function" and "functional" and the like refer to a biological, enzymatic, or therapeutic function.

[0066] By "gene" is meant a unit of inheritance that occupies a specific locus on a chromosome and consists of transcriptional and/or translational regulatory sequences and/or a coding region and/or non-translated sequences (*i.e.*, introns, 5' and 3' untranslated sequences).

[0067] "Genetic linkage" refers to an association of two or more non-allelic genetic loci which do not show independent assortment, often due to physical association on the same chromosome.

[0068] As used herein, the terms "higher risk," "elevated risk" and "increased risk" are used interchangeably herein to refer to a statistically higher frequency of occurrence of a disease or condition (*e.g.*, in an individual carrying a particular polymorphic allele in

comparison to the frequency of occurrence of the disease or condition in a member of a population that does not carry the particular polymorphic allele or in the population as a whole). Alternatively, the terms "lower risk," "reduced risk" and "decreased risk" are used interchangeably herein to refer to a statistically lower frequency of occurrence of a disease or condition (*e.g.*, in an individual carrying a particular polymorphic allele in comparison to the frequency of occurrence of the disease or condition in a member of a population that does not carry the particular polymorphic allele or in the population as a whole).

[0069] "Homology" refers to the percentage number of nucleic or amino acids that are identical or constitute conservative substitutions. Homology may be determined using sequence comparison programs such as GAP (Deveraux *et al.*, 1984, Nucleic Acids Research **12**, 387-395) which is incorporated herein by reference. In this way sequences of a similar or substantially different length to those cited herein could be compared by insertion of gaps into the alignment, such gaps being determined, for example, by the comparison algorithm used by GAP.

[0070] The term "host cell" includes an individual cell or cell culture which can be or has been a recipient of any recombinant vector(s) or isolated polynucleotide of the invention. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation and/or change. A host cell includes cells transfected or infected *in vivo* or *in vitro* with a recombinant vector or a polynucleotide of the invention. A host cell which comprises a recombinant vector of the invention is a "recombinant host cell".

[0071] "Hybridization" is used herein to denote the pairing of complementary nucleotide sequences to produce a DNA-DNA hybrid or a DNA-RNA hybrid. Complementary base sequences are those sequences that are related by the base-pairing rules. In DNA, A pairs with T and C pairs with G. In RNA U pairs with A and C pairs with G. In this regard, the terms "match" and "mismatch" as used herein refer to the hybridization potential of paired nucleotides in complementary nucleic acid strands. Matched nucleotides hybridise efficiently, such as the classical A-T and G-C base pair mentioned above. Mismatches are other combinations of nucleotides that do not hybridise efficiently.

[0072] Reference herein to "immuno-interactive" includes reference to any interaction, reaction, or other form of association between molecules and in particular where one of the molecules is, or mimics, a component of the immune system.

[0073] By "isolated" is meant material that is substantially or essentially free from components that normally accompany it in its native state.

[0074] The term "locus," or "genetic locus" generally refers to a genetically defined region of a chromosome carrying a gene or any other characterized sequence. In some embodiments, the locus is genetically linked to a polymorphic site as defined herein.

[0075] The term "marker", as used herein generally refers to a genetic locus, including a gene or other characterized sequence, which is genetically linked to a trait or phenotype of interest. The term "genetically linked" as used herein refers to two or more loci that are predictably inherited together during random crossing or intercrossing.

[0076] By "obtained from" is meant that a sample such as, for example, a polynucleotide extract or polypeptide extract is isolated from, or derived from, a particular source of the subject. For example, the extract can be obtained from a tissue or a biological fluid isolated directly from the subject.

[0077] The term "oligonucleotide" as used herein refers to a polymer composed of a multiplicity of nucleotide residues (deoxyribonucleotides or ribonucleotides, or related structural variants or synthetic analogues thereof) linked via phosphodiester bonds (or related structural variants or synthetic analogues thereof). Thus, while the term "oligonucleotide" typically refers to a nucleotide polymer in which the nucleotide residues and linkages between them are naturally occurring, it will be understood that the term also includes within its scope various analogues including, but not restricted to, peptide nucleic acids (PNAs), phosphoramidates, phosphorothioates, methyl phosphonates, 2-O-methyl ribonucleic acids, and the like. The exact size of the molecule can vary depending on the particular application. An oligonucleotide is typically rather short in length, generally from about 10 to 30 nucleotide residues, but the term can refer to molecules of any length, although the term "polynucleotide" or "nucleic acid" is typically used for large oligonucleotides.

[0078] The terms "patient" and "subject" are used interchangeably and refer to patients and subjects of human or other mammal and includes any individual it is desired to examine or treat using the methods of the invention. However, it will be understood that "patient" does not imply that symptoms are present. Suitable mammals that fall within the

scope of the invention include, but are not restricted to, primates, livestock animals (*e.g.*, sheep, cows, horses, donkeys, pigs), laboratory test animals (*e.g.*, rabbits, mice, rats, guinea pigs, hamsters), companion animals (*e.g.*, cats, dogs) and captive wild animals (*e.g.*, foxes, deer, dingoes).

[0079] By "pharmaceutically acceptable carrier" is meant a solid or liquid filler, diluent or encapsulating substance that can be safely used in topical or systemic administration to a animal, preferably a mammal including humans.

[0080] The term "polymorphism", as used herein, refers to a difference in the nucleotide or amino acid sequence of a given region as compared to a nucleotide or amino acid sequence in a homologous-region of another individual, in particular, a difference in the nucleotide of amino acid sequence of a given region which differs between individuals of the same species. A polymorphism is generally defined in relation to a reference sequence. Polymorphisms include single nucleotide differences, differences in sequence of more than one nucleotide, and single or multiple nucleotide insertions, inversions and deletions; as well as single amino acid differences, differences in sequence of more than one amino acid, and single or multiple amino acid insertions, inversions, and deletions. A "polymorphic site" is the locus at which the variation occurs. It shall be understood that where a polymorphism is present in a nucleic acid sequence, and reference is made to the presence of a particular base or bases at a polymorphic site, the present invention encompasses the complementary base or bases on the complementary strand at that site.

[0081] The term "polynucleotide" or "nucleic acid" as used herein designates mRNA, RNA, cRNA, cDNA or DNA. The term typically refers to oligonucleotides greater than 30 nucleotide residues in length.

[0082] "Polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues and to variants and synthetic analogues of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues is a synthetic non-naturally occurring amino acid, such as a chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally-occurring amino acid polymers.

[0083] By "primer" is meant an oligonucleotide which, when paired with a strand of DNA, is capable of initiating the synthesis of a primer extension product in the presence of a suitable polymerizing agent. The primer is preferably single-stranded for maximum efficiency in amplification but can alternatively be double-stranded. A primer must be

sufficiently long to prime the synthesis of extension products in the presence of the polymerization agent. The length of the primer depends on many factors, including application, temperature to be employed, template reaction conditions, other reagents, and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15 to 35 or more nucleotide residues, although it can contain fewer nucleotide residues. Primers can be large polynucleotides, such as from about 200 nucleotide residues to several kilobases or more. Primers can be selected to be "substantially complementary" to the sequence on the template to which it is designed to hybridize and serve as a site for the initiation of synthesis. By "substantially complementary", it is meant that the primer is sufficiently complementary to hybridize with a target polynucleotide. In some embodiments, the primer contains no mismatches with the template to which it is designed to hybridize but this is not essential. For example, non-complementary nucleotide residues can be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the template. Alternatively, non-complementary nucleotide residues or a stretch of non-complementary nucleotide residues can be interspersed into a primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridize therewith and thereby form a template for synthesis of the extension product of the primer.

[0084] "Probe" refers to a molecule that binds to a specific sequence or sub-sequence or other moiety of another molecule. Unless otherwise indicated, the term "probe" typically refers to a polynucleotide probe that binds to another polynucleotide, often called the "target polynucleotide", through complementary base pairing. Probes can bind target polynucleotides lacking complete sequence complementarity with the probe, depending on the stringency of the hybridization conditions. Probes can be labeled directly or indirectly.

[0085] The term "sequence identity" as used herein refers to the extent that sequences are identical on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over a window of comparison. Thus, a "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, I) or the identical amino acid residue (*e.g.*, Ala, Pro, Ser, Thr, Gly, Val, Leu, Iie, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions

in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity.

[0086] "Single nucleotide polymorphism (SNP)" as used herein refers to a change in which a single base in the DNA differs (such as via substitutions, addition or deletion) from the usual base at that position. For example, a single nucleotide polymorphism is characterized by the presence in a population of one or two, three or four nucleotides (*i.e.*, adenosine, cytosine, guanosine or thymidine) at a particular locus in a genome such as the human genome. It will be recognized that while the methods of the present invention are directed to the identification of certain SNPs, the methods can be used to identify other AS-associated SNPs either alone or in combination with the exemplified SNPs, or combined with methods for determining other AS-associated polymorphisms in the polymorphic sites described herein, to increase the accuracy of the determination.

[0087] "Stringency" as used herein, refers to the temperature and ionic strength conditions, and presence or absence of certain organic solvents, during hybridization and washing procedures. The higher the stringency, the higher will be the degree of complementarity between immobilized target nucleotide sequences and the labeled probe polynucleotide sequences that remain hybridized to the target after washing. The term "high stringency" refers to temperature and ionic conditions under which only nucleotide sequences having a high frequency of complementary bases will hybridize. The stringency required is nucleotide sequence dependent and depends upon the various components present during hybridization. Generally, stringent conditions are selected to be about 10 to 20° C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a target sequence hybridizes to a complementary probe.

[0088] As used herein, the terms "treatment," "treating," and the like, refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete cure for a disease and/or adverse affect attributable to the disease. "Treatment," as used herein, covers any treatment of a disease in a mammal, particularly in a human, and includes: (a) preventing the disease from occurring in a subject which may be predisposed to the disease but has not yet been diagnosed as having it;

(b) inhibiting the disease, *i.e.*, arresting its development; and (c) relieving the disease, *i.e.*, causing regression of the disease.

[0089] By "vector" is meant a polynucleotide molecule, preferably a DNA molecule derived, for example, from a plasmid, bacteriophage, yeast or virus, into which a polynucleotide can be inserted or cloned. A vector preferably contains one or more unique restriction sites and can be capable of autonomous replication in a defined host cell including a target cell or tissue or a progenitor cell or tissue thereof, or be integrable with the genome of the defined host such that the cloned sequence is reproducible. Accordingly, the vector can be an autonomously replicating vector, *i.e.*, a vector that exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a linear or closed circular plasmid, an extra-chromosomal element, a mini-chromosome, or an artificial chromosome. The vector can contain any means for assuring self-replication. Alternatively, the vector can be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. A vector system can comprise a single vector or plasmid, two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. In the present case, the vector is preferably a viral or viral-derived vector, which is operably functional in animal and preferably mammalian cells. Such vector may be derived from a poxvirus, an adenovirus or yeast. The vector can also include a selection marker such as an antibiotic resistance gene that can be used for selection of suitable transformants. Examples of such resistance genes are known to those of skill in the art and include the nptII gene that confers resistance to the antibiotics kanamycin and G418 (Geneticin®) and the hph gene which confers resistance to the antibiotic hygromycin B.

2. Polymorphisms of the invention

[0090] The present invention is based in part on the determination that a single imputed SNP, rs116488202 , tags *HLA-B*27* as accurately as *HLA-B*27* genotyping in affected individuals and in healthy controls and thus can be used in place of *HLA-B*27* for genetic diagnosis of *HLA-B*27* associated spondyloarthropathies (SpA) such as reactive arthritis, AS, anterior uveitis, psoriatic arthritis, and enteropathic arthritis associated with inflammatory bowel disease (colitic arthritis). Surprisingly, this SNP tags *HLA-B*27* in both

western European and Asian populations better than rs4349859 (as described for example in WO2011/063474), and the SNP rs13202464, reported to tag *HLA-B*27* in Asian populations (*see*, for example, Lin, Z. *et al.*, 2011. *Nature Genetics* **44**, 73-77). Accordingly, the present invention provides methods for diagnosing the presence, absence or risk of development of a SpA in a subject, wherein the methods generally comprise analyzing a biological sample of the subject, which comprises at least a portion of a SpA marker that comprises a polymorphic site having reference sequence number rs116488202, for the presence or absence of a polymorphism in the polymorphic site, which polymorphism is indicative of the presence, absence or risk of development of the SpA. Suitably, the methods comprise obtaining the biological sample from the subject. In some embodiments, the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site having reference sequence number rs116488202, wherein the presence of T at rs116488202, indicates that the subject has the SpA or has an increased risk of developing the SpA, and wherein the presence of C at rs116488202, indicates that the subject lacks the SpA or has a decreased risk of developing the SpA.

[0091] In some embodiments the SpA is AS, and the methods further comprise analyzing the sample for the presence of a polymorphism in at least one AS marker, illustrative examples of which include the *ERAP1* gene, the *IL23R* gene, the *TNFR1* gene locus, the *TRADD* gene locus, the 2P15 chromosomal locus, the 21Q22 chromosomal locus and the *HLA-B* gene, as disclosed for example in copending International Application No. PCT/AU2008/000762 filed 29 May 2008 (WO 2008/144827), as well as the *IL1R1* gene locus, the *IL1R2* gene locus and the chromosome locus 2Q31.3, as disclosed for example in copending International Application No. PCT/AU2009/01320 filed 2 October 2009 (WO 2010/037184). The SpA/AS-associated SNP of the present invention and the AS-associated SNPs disclosed in the copending applications noted above are summarized in Table 2 below.

TABLE 2

AS-ASSOCIATED SNPs

1	rs116488202	Chr 6: [redacted]	NT_007592.15 [GI: 224514668] 31284916 SEQ ID NO: 1 201	Non-coding	Intergenic	C/T	N/A
2	rs27044	Chr 5 96144608	NC_000005 [GI:224589817] 25041 reverse complement SEQ ID NO: 2 301	Coding	*ERAP1*	C/G	Glu/Gln 730 [1]
3	rs17482078	Chr 5 96144622	NC_000005 [GI:224589817] 25027 reverse complement SEQ ID NO: 3 301	Coding	*ERAP1*	T/C	Gln/Arg 725 [1]
4	rs30187	Chr 5 96150086	NC_000005 [GI:224589817] 19563 reverse complement SEQ ID NO: 4 301	Coding	*ERAP1*	C/T	Arg/Lys 528 [1]
5	rs10050860	Chr 5 96147966	NC_000005 [GI:224589817] 21683 reverse complement SEQ ID NO: 5 301	Coding	*ERAP1*	T/C	Asn/Asp 575 [1]
6	rs2287987	Chr 5 96155291	NC_000005 [GI:224589817] 14358 reverse complement SEQ ID NO: 6 301	Coding	*ERAP1*	C/T	Val/Met 349 [1]

7	rs10489629	Chr 1 67475114	NC_000001 [GI:224589800] 56181 SEQ ID NO: 7 101	Non-coding	*IL23R*	C/T	NA
8	rs1343151	Chr 1 67497708	NC_000001 [GI:224589800] 86961 SEQ ID NO: 8 301	Non-coding	*IL23R*	A/G	NA
9	rs10889677	Chr 1 67512680	NC_000001 [GI:224589800] 92952 SEQ ID NO: 9 301	Non-coding	*IL23R*	C/A	NA
10	rs11209032	Chr 1 67526096	SEQ ID NO: 10 301	Non-coding	*IL23R*	G/A	N/A
11	rs1004819	Chr 1 67460937	NC_000001 [GI:224589800] 38045 SEQ ID NO: 11 489	Non-coding	*IL23R*	G/A	NA
12	rs11465804	Chr 1 67478546	NC_000001 [GI:224589800] 70358 SEQ ID NO: 12 101	Non-coding	*IL23R*	G/T	NA
13	rs11209026	Chr 1 67491717	NC_000001 [GI:224589800] 73790 SEQ ID NO: 13 301	Coding	*IL23R*	A/G	Gln/Arg *381* [2]
14	rs1495965	Chr 1 67442801	NC_000001 [GI:224589800] SEQ ID NO: 14 259	Non-coding	*IL23R*	T/C	NA

15	rs4149576	Chr 12 6319376	NC_000012 [GI:224589803] 2147 SEQ ID NO: 15 301	Non-coding	*TNFR1*	T/C	NA
16	rs9033	Chr 16 65739500	NC_000016 [GI:224589807] 38528 SEQ ID NO: 16 300	Non-coding	*TRADD*	A/G	NA
17	rs10865331	Chr 2 62551472	NT_022184.15 [GI:224515010] 41373359 SEQ ID NO: 17 643	Non-coding	*2P15*	G/A	NA
18	rs2242944	Chr 21 40461578	NT_011512.11 [GI:224514633] 26127049 SEQ ID NO: 18 501	Non-coding	*21Q22*	A/G	NA
19	rs949963	Chr 2: 102136218	AF531102 [GI:22001412] 680; or SEQ ID NO: 19 201	Non-coding	*IL1R1*	T/C	N/A
20	rs2310173	Chr 2: 102030060	AC007165 [GI:19033999] 45295; or SEQ ID NO: 20 500	Non-coding	*IL1R2*	G/T	N/A
21	rs1018326	Chr 2: 181716045	NC_000002 [GI:89161199] 162689; or SEQ ID NO: 22 301	Non-coding	*2Q31.3 (genetically linked to UBE2E3)*	T/C	N/A

[1] Relative to the ERAP-1 amino acid sequence set forth in GenPept Accession No. NP_057526 [GI:94818901]

[2] Relative to the IL-23R amino acid sequence set forth in GenPept Accession No. NP_653302 [GI:24430212]

* Minor Allele is indicated in bold type face

[0092] In general, if the polymorphism is located in a gene, it may be located in a non-coding or coding region of the gene. If located in the coding region the polymorphism can result in an amino acid alteration. Such alterations may or may not have an effect on the function or activity of the encoded polypeptide. For example, the polymorphisms listed in Table 1 within the *ERAP1* and *IL23R* coding regions are non-synonymous mutations which cause a change in the amino acid sequence. The other seven polymorphisms within the *IL23R* gene sequence are in the non-coding region. However, when the polymorphism is located in a non-coding region it can cause alternative splicing, which again, may or may not have an effect on the encoded protein activity or function.

[0093] The methods of the present invention comprise detecting the presence, absence or risk of development of an SpA (*e.g.*, AS) by identifying related polymorphisms in DNA or mRNA (or on other nucleic acid sequences, such as cDNA, developed there from) or protein contained in tissue, blood or other biological samples taken from a subject. The polymorphism can be detected in any manner conventionally known in the art, *e.g.*, *via* directly sequencing of the nucleotide sequences contained in the samples. Such diagnosis or prediction can also be made by identifying the nucleotide polymorphism or variant protein in samples taken from kindred or other relatives of a subject. This can be helpful, for example, in determining whether offspring are likely to be genetically predisposed to the condition, even though it has not expressed itself in the parents.

[0094] It is to be understood that although the following discussion is specifically directed to human subjects, the teachings are also applicable to any animal that contains the polymorphisms in accordance with the present invention, such that clinical manifestations such as those seen in subjects with AS are found. In specific embodiments, the subject is caucasian or caucasoid or of caucasian ancestry. In specific embodiments, the subject is other than an asian or an african.

[0095] It will be appreciated that the methods described herein are applicable to any subject suspected of developing, or having a SpA (*e.g.*, AS), whether the condition is manifest at a young age or at a more advanced age in a patient's life. The subject can be an adult, child, fetus or embryo.

[0096] The diagnostic and screening methods of the invention are especially useful for a subject suspected of being at risk of developing a SpA (*e.g.*, AS) based on family history, or a subject in which it is desired to diagnose or eliminate the presence of the SpA as a causative agent underlying a subject's symptoms.

3. Screening for specific polymorphisms within the AS markers of the invention

3.1 Amplification techniques

[0097] In some embodiments, screening or determination of a SpA (*e.g.*, AS), or a predisposition to developing a SpA in a subject is now possible by detecting one or more polymorphisms linked to that condition. For example, numerous methods are known in the art for determining the nucleotide occurrence at a particular position corresponding to a single nucleotide polymorphism in a sample. Suitably, methods of detecting point mutations may be accomplished by molecular cloning of the specified allele and subsequent sequencing of that allele using techniques well known in the art. A method according to the present invention can identify a nucleotide occurrence for either strand of DNA. Additionally, the gene sequences may be amplified directly from a DNA or mRNA (or on other nucleic acid sequences, such as cDNA) preparation from the sample using amplification techniques, and the sequence composition can then be determined from the amplified product.

[0098] The nucleic acid sample may be obtained from any part of the subject's body, including, but not limited to hair, skin, nails, tissues or bodily fluids such as saliva and blood. The subject for the methods of the present invention can be a subject of any race or national origin.

[0099] Nucleic acid isolation protocols are well known to those of skill in the art. For example, an isolated polynucleotide corresponding to a gene or allele or chromosome region (*e.g.*, as listed in Table 2) may be prepared according to the following procedure:

[0100] - creating primers which flank an allele or transcript thereof, or a portion of the allele or transcript;

[0101] - obtaining a nucleic acid extract from an individual affected with, or at risk of developing a SpA (*e.g.*, AS); and

[0102] - using the primers to amplify, *via* nucleic acid amplification techniques, at least one amplification product from the nucleic acid extract, wherein the amplification product corresponds to the allele or transcript linked to the development of the condition.

[0103] Suitable nucleic acid amplification techniques are well known to a person of ordinary skill in the art, and include polymerase chain reaction (PCR) as for example described in Ausubel *et al.*, Current Protocols in Molecular Biology (John Wiley & Sons, Inc. 1994-1998) strand displacement amplification (SDA) as for example described in U.S. Patent No 5,422,252; rolling circle replication (RCR) as for example described in Liu *et al.*, (1996, J. Am. Chem. Soc. 118: 1587-1594 and International application WO 92/01813) and Lizardi *et al.*, (International Application WO 97/19193); nucleic acid sequence-based amplification (NASBA) as for example described by Sooknanan *et al.*, (1994, Biotechniques 17: 1077-1080); ligase chain reaction (LCR); simple sequence repeat analysis (SSR); branched DNA amplification assay (b-DNA); transcription amplification and self-sustained sequence replication; and Q-β replicase amplification as for example described by Tyagi *et al.*, (1996, Proc. Natl. Acad. Sci. USA 93: 5395-5400).

[0104] Such methods can utilize one or more oligonucleotide probes or primers, including, for example, an amplification primer pair, that selectively hybridize to a target polynucleotide, which contains one or more SNPs. Oligonucleotide probes useful in practicing a method of the invention can include, for example, an oligonucleotide that is complementary to and spans a portion of the target polynucleotide, including the position of the SNP, wherein the presence of a specific nucleotide at the polymorphic site (*i.e.*, the SNP) is detected by the presence or absence of selective hybridization of the probe. Such a method can further include contacting the target polynucleotide and hybridized oligonucleotide with an endonuclease, and detecting the presence or absence of a cleavage product of the probe, depending on whether the nucleotide occurrence at the polymorphic site is complementary to the corresponding nucleotide of the probe.

[0105] Primers may be manufactured using any convenient method of synthesis. Examples of such methods may be found in "Protocols for Oligonucleotides and Analogues; Synthesis and Properties", Methods in Molecular Biology Series; Volume 20; Ed. Sudhir Agrawal, Humana ISBN: 0-89603-247-7; 1993. The primers may also be labeled to facilitate detection.

3.2 Nucleic acid polymorphism screening techniques

[0106] Various tools for the detection of polymorphisms within a target DNA are known in the art, including, but not limited to screening techniques, DNA sequencing,

scanning techniques, hybridization based techniques, extension based analysis, incorporation based techniques, restriction enzyme based analysis and ligation based techniques.

3.3 Nucleic acid sequencing techniques

[0107] In some embodiments, the polymorphism is identified through nucleic acid sequencing techniques. Specifically, amplification products which span a SNP locus can be sequenced using traditional sequence methodologies (*e.g.*, the "dideoxy-mediated chain termination method", also known as the "Sanger Method" (Sanger, F., *et al.*, 1975, *J. Molecular, Biol.* **94**: 441; Prober *et al.*, 1987, *Science*, **238**: 336-340) and the "chemical degradation method", also known as the "Maxam-Gilbert method" (Maxam, A. M., *et al.*, 1977, *Proc. Natl. Acad. Sci.* (U.S.A.) **74**: 560), both references herein incorporated by reference to determine the nucleotide occurrence at the SNP loci.

[0108] Boyce-Jacino, *et al.*, U.S. Pat. No. 6,294,336 provides a solid phase sequencing method for determining the sequence of nucleic acid molecules (either DNA or RNA) by utilizing a primer that selectively binds a polynucleotide target at a site wherein the SNP is the most 3' nucleotide selectively bound to the target. Other sequencing technologies such as Denaturing High Pressure Liquid Chromatography or mass spectroscopy may also be employed.

[0109] In other illustrative examples, the sequencing method comprises a technique known as Pyrosequencing™. The approach is based on the generation of pyrophosphate whenever a deoxynucleotide is incorporated during polymerization of DNA. The generation of pyrophosphate is coupled to a luciferase catalysed reaction resulting in light emission if the particular deoxynucleotide added is incorporated, yielding a quantitative and distinctive pyrogram. Sample processing includes PCR amplification with a biotinylated primer, isolation of the biotinylated single strand amplicon on streptavidin coated beads (or other solid phase) and annealing of a sequencing primer. Samples are then analysed by a Pyrosequencer™ which adds a number of enzymes and substrates required for the indicator reaction, including sulfurylase and luciferase, as well as apyrase for degradation of unincorporated nucleotides. The sample is then interrogated by addition of the four deoxynucleotides. Light emission can be detected by a charge coupled device camera (CCD) and is proportional to the number of nucleotides incorporated. Results are automatically assigned by pattern recognition.

[0110] Alternatively, methods of the invention can identify nucleotide occurrences at polymorphic sites within a nucleic acid sequence using a "micro-sequencing" method. Micro-sequencing methods determine the identity of only a single nucleotide at a "predetermined" site. Such methods have particular utility in determining the presence and identity of polymorphisms in a target polynucleotide. Such micro-sequencing methods, as well as other methods for determining the nucleotide occurrence at a polymorphic site are discussed in Boyce-Jacino *et al.*, U.S. Patent Number 6294336, incorporated herein by reference.

[0111] Micro-sequencing methods include the Genetic Bit Analysis.™. method disclosed by Goelet, P. *et al.* WO 92/15712. Additional, primer-guided, nucleotide incorporation procedures for assaying polymorphic sites in DNA have also been described (Komher, J. S. *et al*, 1989, *Nucl. Acids. Res.* **17**: 7779-7784; Sokolov, B. P., 1990, *Nucl. Acids Res.* **18**: 3671; Syvanen, A. C, *et al.*, 1990, *Genomics*, **8**: 684-692; Kuppuswamy, M. N. *et al.*, 1991, *Proc. Natl. Acad. Sci.* (U.S.A.) **88**: 1143-1147; Prezant, T. R. *et al*, 1992, *Hum. Mutat.* **1**: 159-164; Ugozzoli, L. *et al.*, 1992, GATA, **9**: 107-112; Nyren, P. *et al.*, 1993, *Anal. Biochem.* **208**: 171-175; and Wallace, WO89/10414). These methods differ from Genetic Bit.™. analysis in that they all rely on the incorporation of labeled deoxynucleotides to discriminate between bases at a polymorphic site. In such a format, since the signal is proportional to the number of deoxynucleotides incorporated, polymorphisms that occur in runs of the same nucleotide can result in signals that are proportional to the length of the run (Syvanen, A. C., *et al.*, 1993, *Amer. J. Hum. Genet.* **52**: 46-59).

[0112] Further micro-sequencing methods have been provided by Mundy, C. R. (U.S. Pat. No. 4,656,127) and Cohen, D. *et al* (French Patent 2,650,840; PCT Publication No. WO91/02087) which discusses a solution-based method for determining the identity of a nucleotide of a polymorphic site. As in the Mundy method of U.S. Pat. No. 4,656,127, a primer is employed that is complementary to allelic sequences immediately 3' to a polymorphic site.

[0113] In other illustrative examples, Macevicz (U.S. Pat. No. 5,002,867), for example, describes a method for determining nucleic acid sequences *via* hybridization with multiple mixtures of oligonucleotide probes. In accordance with such methods, the sequence of a target polynucleotide is determined by permitting the target to sequentially hybridize with sets of probes having an invariant nucleotide at one position, and a variant nucleotides at other

positions. The Macevicz method determines the nucleotide sequence of the target by hybridizing the target with a set of probes, and then determining the number of sites that at least one member of the set is capable of hybridizing to the target (*i.e.*, the number of "matches"). This procedure is repeated until each member of a set of probes has been tested.

[0114] Alternatively, the template-directed dye-terminator incorporation assay with fluorescence polarization detection (FP-TDI) assay (Chen *et al.*, 1999) is a version of the primer extension assay that is also called mini-sequencing or the single base extension assay (Syvanen, 1994). The primer extension assay is capable of detecting SNPs. The DNA sequencing protocol ascertains the nature of the one base immediately 3' to the SNP-specific sequencing primer that is annealed to the target DNA immediately upstream from the polymorphic site. In the presence of DNA polymerase and the appropriate dideoxyribonucleoside triphosphate (ddNTP), the primer is extended specifically by one base as dictated by the target DNA sequence at the polymorphic site. By determining which ddNTP is incorporated, the allele(s) present in the target DNA can be inferred.

3.4 Polymorphism scanning techniques

[0115] Scanning techniques contemplated by the present invention for detecting polymorphisms within a nucleotide sequence can include, but are not restricted to, chemical mismatch cleavage (CMC) (Saleeba, J. A *et al.*, 1992, *Huma. Mutat*, **1**: 63-69), mismatch repair enzymes cleavage (MREC) (Lu, A. L and Hsu, I. C., 1992, *Genomics*, **14**(2): 249-255), chemical cleavage techniques, denaturing gradient gel electrophoresis (DGGE) Wartell *et al.*, (1990, *Nucl. Acids Res.* **18**: 2699-2705 and; Sheffield *et al.*, 1989, *Proc. Natl. Acad. Sci.* USA **86**: 232-236), temperature gradient gel electrophoresis (TGGE) (Salimullah, *et al.*, 2005, *Cellular and Mol. Biol. Letts*, **10**: 237-245), constant denaturant gel electrophoresis (CDGE), single strand conformation polymorphism (SSCP) analysis (Kumar, D *et al.*, 2006, *Genet. Mol. Biol*, **29**(2): 287-289), heteroduplex analysis (HA) (Nagamine, C. M *et al.*, 1989, *Am. J. Hum. Genet*, **45**: 337-339), microsatellite marker analysis and single strand polymorphism assays (SSPA).

[0116] In some embodiments, the SNPs of the present invention are detected through CMC, wherein a radio-labeled DNA wild type sequence (probe) is hybridized to an amplified sequence containing the putative alteration to form a heteroduplex. A chemical modification, followed by piperidine cleavage, is used to remove the mismatch bubble in the heteroduplex. Gel electrophoresis of the denatured heteroduplex and autoradiography allow to

visualize the cleavage product. Osmium tetroxide is used for the modification of mispaired thymidines and hydroxylamine for mismatched cytosines. Additionally, labeling the antisense strand of the probe DNA allows the detection of adenosine and guanosine mismatches. The chemical cleavage of mismatch can be used to detect almost 100% of mutations in long DNA fragments. Moreover, this method provides the precise characterization and the exact location of the mutation within the tested fragment. Recently, the method has been amended to make CMC more suitable for automation by using fluorescent primers also enabling multiplexing and thereby reducing the number of manipulations. Alternatively, fluorescently labeled dUTPs incorporated via PCR allow the internal labeling of both target and probe DNA strands and therefore labeling of each possible hybrid, doubling the chances of mutation detection and virtually guaranteeing 100% detection.

[0117] In other embodiments, the mismatch repair enzymes cleavage (MREC) assay is used to identify single base substitutions within an AS marker of the present invention. MREC relies on nicking enzyme systems specific for mismatch-containing DNA. The sequence of interest is amplified by PCR and homo- and heteroduplex species may be generated at the end of the PCR, by denaturing and allowing to re-anneal the amplified products. These hybrids are treated with mismatch repair enzymes and then analysed by denaturing gel electrophoresis. The MREC assay makes use of three mismatch repair enzymes. The MutY endonuclease removes adenines from the mismatches and is useful to detect both A/T and C/G transversions and G/C and T/A transitions. Mammalian thymine glycosylase removes thymines from T/G, T/C, and T/T mismatches and is useful to detect G/C and A/T transitions as well as A/T and G/C and T/A and A/T transversions. The all-type endonuclease or topoisomerase I from human or calf thymus can recognize all eight mismatches and can be used to scan any nucleotide substitution. MREC can use specific labels which can be incorporated into both DNA strands, thus allowing all four possible nucleotide substitutions in a give site to be identified.

[0118] In some embodiments, chemical cleavage analysis as described in U.S. Pat. No. 5,217,863 (by R. G. H. Cotton) is used for identifying SNPs within nucleotide sequences. Like heteroduplex analysis, chemical cleavage detects different properties that result when mismatched allelic sequences hybridize with each other. Instead of detecting this difference as an altered migration rate on a gel, the difference is detected in altered susceptibility of the hybrid to chemical cleavage using, for example, hydroxylamine, or osmium tetroxide, followed by piperidine.

[0119] Among the cleavage methods contemplated by the present invention, RNAse A relies on the principle of heteroduplex mismatch analysis. In the RNAse A cleavage method, RNA-DNA heteroduplex between radiolabeled wild-type riboprobe and a mutant DNA, obtained by PCR amplification, is enzymatically cleaved by RNAse A, by exploiting the ability of RNAse A to cleave single-stranded RNA at the points of mismatches in RNA:DNA hybrids. This is followed by electrophoresis and autoradiography. The presence and location of a mutation are indicated by a cleavage product of a given size (Meyers, R. M *et al.*, 1985, Science, 230: 1242-1246 and; Gibbs, R. A and Caskey, T, 1987, Science, 236: 303-305).

[0120] The riboprobe need not be the full-length of an AS marker sequences of the present invention (*e.g.*, as listed in Tables 1 and 2). However, a number of probes can be used to screen the whole mRNA sequence for mismatches. In a similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, *e.g.*, Cotton, *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85: 4397; Shenk *et al.*, 1975, Proc. Natl. Acad. Sci. USA 72: 989; and Novack *et al.*, 1986, Proc. Natl. Acad. Sci. USA 83: 586.

[0121] In some embodiments, the Invader® assay (Third Wave™ Technology) is employed to scan for polymorphisms within the AS marker sequences of the present invention. For example, the Invader® assay is based on the specificity of recognition, and cleavage, by a Flap endonuclease, of the three dimensional structure formed when two overlapping oligonucleotides hybridize perfectly to a target DNA (Lyamichev, V *et al.*, 1999, Nat Biotechnol, 17: 292-296).

[0122] Alternatively, denaturing gradient gel electrophoresis (DGGE) is a useful technique to separate and identify sequence variants. DGGE is typically performed in constant-concentration polyacrylamide gel slabs, cast in the presence of linearly increasing amounts of a denaturing agent (usually formamide and urea, cathode to anode). A variant of DGGE employs temperature gradients along the migration path and is known as TGGE. Separation by DGGE or TGGE is based on the fact that the electrophoretic mobility in a gel of a partially melted DNA molecule is greatly reduced as compared to an unmelted molecule.

[0123] In some embodiments, constant denaturant gel electrophoresis (CDGE) is useful for detecting SNPs within a nucleotide sequence, as described in detail in Smith-Sorenson *et al.*, 1993, Human Mutation 2: 274-285 (see also, Anderson & Borreson, 1995, Diagnostic Molecular Pathology 4: 203-211). A given DNA duplex melts in a predetermined,

characteristic fashion in a gel of a constant denaturant. Mutations alter this movement. An abnormally migrating fragment is isolated and sequenced to determine the specific mutation.

[0124] In other embodiments, single-strand conformation polymorphism (SSCP) analysis provides a method for detecting SNPs within the AS marker sequences of the present invention. SSCP is a method based on a change in mobility of separated single-strand DNA molecules in non-denaturing polyacrylamide gel electrophoresis. Electrophoretic mobility depends on both size and shape of a molecule, and single-stranded DNA molecules fold back on themselves and generate secondary structures which are determined by intra-molecular interactions in a sequence dependent manner. A single nucleotide substitution can alter the secondary structure and, consequently, the electrophoretic mobility of the single strands, resulting in band shifts on autoradiographs. The ability of a given nucleotide variation to alter the conformation of the single strands is not predictable on the basis of an adequate theoretical model and base changes occurring in a loop or in a long stable stem of the secondary structure might not be detected by SSCP. Standard SSCP reaches maximal reliability in detecting sequence alterations in fragments of 150-200 bp. More advanced protocols, allowing the detection of mutations at sensitivity equal to that of the radioactively-based SSCP analysis, have been developed. These methods use fluorescence-labeled primers in the PCR and analyze the products with a fluorescence-based automated sequencing machine. Multi-colour fluorescent SSCP also allows to include an internal standard in every lane, which can be used to compare data from each lane with respect to each other. Other variants to increase the detection rate includes a dideoxy sequencing approach based on dideoxy fingerprinting (ddF) and restriction endonuclease fingerprinting (REF).

[0125] The ddF method is a combination of SSCP and Sanger dideoxy sequencing, which involves non-denaturing gel electrophoresis of a Sanger sequencing reaction with one dideoxynucleotide. In this way, for example, a 250-bp fragment can be screened to identify a SNP. REF is a more complex modification of SSCP allowing the screening of more than 1 kb fragments. For REF, a target sequence is amplified with PCR, digested independently with five to six different restriction endonucleases and analyzed by SSCP on a non-denaturing gel. In the case of six restriction enzymes being used, a sequence variation will be present in six different restriction fragments, thus generating 12 different single-stranded segments. A mobility shift in any one of these fragments is sufficient to pinpoint the presence of a sequence variation within a portion of at least one of the AS marker sequences of the

invention. The restriction pattern obtained enables localization of an alteration in the region examined.

[0126] In some embodiments, heteroduplex analysis (HA) detects single base substitutions in PCR products or nucleotide sequences. HA can be rapidly performed without radioisotopes or specialized equipment. The HA method takes advantage of the formation of heteroduplexes between wild-type and mutated sequences by heating and renaturing of PCR products. Due to a more open double-stranded configuration surrounding the mismatched bases, heteroduplexes migrate slower than their corresponding homoduplexes, and are then detected as bands of reduced mobility compared to normal and mutant homoduplexes on polyacrylamide gels. The ability of a particular single base substitution to be detected by the HA method cannot be predicted merely by knowing the mismatched bases since the adjacent nucleotides have a substantial effect on the configuration of the mismatched region and length-based separation will clearly miss nucleotide substitutions. Optimization of the temperature, gel cross-linking and concentration of acrylamide used as well as glycerol and sucrose enhance the resolution of mutated samples. The HA method can be rapidly performed without radioisotopes or specialized equipment and screens large numbers of samples from subjects for known mutations and polymorphisms in sequenced genes. When HA is used in combination with SSCP, up to 100% of all alterations in a DNA fragment can be easily detected.

[0127] In some embodiments, the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein can be used to detect an AS-associated polymorphism within at least one of the AS marker sequences of the present invention (Modrich, 1991, Ann. Rev. Genet. 25: 229-253). In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

[0128] In further embodiments, polymorphism detection can be performed using microsatellite marker analysis. Microsatellite markers with an average genome spacing, for example of about 10 centimorgans (cM) can be employed using standard DNA isolation methods known in the art.

[0129] SSPA analysis and the closely related heteroduplex analysis methods described above may be used for screening for single-base polymorphisms (Orita, M. *et al.*, 1989, Proc Natl Acad Sci USA, 86: 2766). In these methods, the mobility of PCR-amplified

test DNA from subjects with AS or at risk of developing AS is compared with the mobility of DNA amplified from normal sources by direct electrophoresis of samples in adjacent lanes of native polyacrylamide or other types of matrix gels. Single-base changes often alter the secondary structure of the molecule sufficiently to cause slight mobility differences between the normal and mutant PCR products after prolonged electrophoresis. The presence of polymorphisms, including mutations, in nucleic acids by using mass spectrometry may be used as discussed in U.S. Pat. No: 5,869,242.

3.5 Polymorphism hybridization based techniques

[0130] Hybridization techniques for detecting polymorphisms within a nucleotide sequence can include, but are not restricted to the TaqMan® assay (*Applied Biosystems*), dot blots, reverse dot blot, Multiplex-allele-specific diagnostic assays (MASDA), Dynamic allele-specific hybridization (DASH) Jobs *et al.*, (2003, *Genome Res* **13**: 916-924), molecular beacons and Southern blots.

[0131] The TaqMan® assay for identifying SNPs within a nucleotide sequence is based on the nuclease activity of *Taq* polymerase that displaces and cleaves the oligonucleotide probes hybridized to the target DNA, generating a fluorescent signal. Two TaqMan® probes that differ at the polymorphic site are required; one probe is complementary to the wild-type allele and the other to the variant allele. The probes have different fluorescent dyes attached to the 50 end and a quencher attached to the 30 end. When the probes are intact, the quencher interacts with the fluorophore by fluorescence resonance energy transfer (FRET), quenching their fluorescence. During the PCR annealing step, the TaqMan® probes hybridize to the target DNA. In the extension step, the fluorescent dye is cleaved by the nuclease activity of the *Taq* polymerase, leading to an increase in fluorescence of the reporter dye. Mismatch probes are displaced without fragmentation. The genotype of a sample is determined by measuring the signal intensity of the two different dyes.

[0132] In some embodiments, a biological sample from a subject can be probed in a standard dot blot format. Each region within the test sample that contains a nucleotide sequence corresponding to a SpA marker or AS marker sequence or a portion thereof is individually applied to a solid surface, for example, as an individual dot on a membrane. Each individual region can be produced, for example, as a separate PCR amplification product using methods well-known in the art (see, for example, the experimental embodiment set forth in Mullis, K. B., 1987, U.S. Pat. No. 4,683,202).

[0133] In a related embodiment, a reverse dot blot format is employed, wherein oligonucleotide or polynucleotide probes having known sequence are immobilized on the solid surface, and are subsequently hybridized with the labeled test polynucleotide sample.

[0134] Another useful SNP identification method includes DASH (dynamic allele-specific hybridization), which encompasses dynamic tracking of probe (oligonucleotide) to target (PCR product) hybridization as the reaction temperature is steadily increased to identify polymorphisms (Prince, J. A *et al.*, 2001, *Genome Res*, **11**(1): 152-162).

[0135] In some embodiments, multiplex-allele-specific diagnostic assays (MASDA) can be used for the analysis of a large number of samples (> 500). MASDA utilizes oligonucleotide hybridization to interrogate DNA sequences. Multiplex DNA samples are immobilized on a solid support and a single hybridization is performed with a pool of allele-specific oligonucleotide (ASO) probes. Any probes complementary to specific mutations present in a given sample are in effect affinity purified from the pool by the target DNA. Sequence-specific band patterns (fingerprints), generated by chemical or enzymatic sequencing of the bound ASO(s), easily identify the specific mutation(s).

[0136] There are several alternative hybridization-based techniques, including, among others, molecular beacons, and Scorpion® probes (Tyagi, S and Kramer, F. R., 1996, *Nat. Biotechnol*, **14**: 303-308; Thelwell *et al.*, 2000, *Nucleic Acid Res.* **28**(19): 3752-3761). Molecular beacons are comprised of oligonucleotides that have a fluorescent reporter and quencher dyes at their 5' and 3' ends. The central portion of the oligonucleotide hybridizes across the target sequence, but the 5' and 3' flanking regions are complementary to each other. When not hybridised to their target sequence, the 5' and 3' flanking regions hybridise to form a stem-loop structure, and there is little fluorescence because of the proximity of the reporter and quencher dyes. However, upon hybridization to their target sequence, the dyes are separated and there is a large increase in fluorescence. Mismatched probe-target hybrids dissociate at substantially lower temperature than exactly complementary hybrids. There are a number of variations of the "beacon" approach. Scorpion® probes are similar but incorporate a PCR primer sequence as part of the probe. A more recent "duplex" format has also been developed.

[0137] In some embodiments, a further method of identifying an SNP comprises the SNP-IT™ method (Orchid BioSciences, Inc., Princeton, N.J.). In general, SNP-IT™ is a 3-step primer extension reaction. In the first step a target polynucleotide is isolated from a

sample by hybridization to a capture primer, which provides a first level of specificity. In a second step the capture primer is extended from a terminating nucleotide trisphosphate at the target SNP site, which provides a second level of specificity. In a third step, the extended nucleotide trisphosphate can be detected using a variety of known formats, including: direct fluorescence, indirect fluorescence, an indirect colorimetric assay, mass spectrometry, fluorescence polarization, etc. Reactions can be processed in 384 well format in an automated format using a SNPstream.™ instrument (Orchid BioSciences, Inc., Princeton, N.J.).

[0138] In these embodiments, the amplification products can be detected by Southern blot analysis with or without using radioactive probes. In one such method, for example, a small sample of DNA containing a very low level of the nucleic acid sequence of the polymorphic locus is amplified, and analyzed *via* a Southern blotting technique or similarly, using dot blot analysis. The use of non-radioactive probes or labels is facilitated by the high level of the amplified signal. Alternatively, probes used to detect the amplified products can be directly or indirectly detectably labeled, for example, with a radioisotope, a fluorescent compound, a bioluminescent compound, a chemiluminescent compound, a metal chelator or an enzyme.

[0139] Hybridization conditions, such as salt concentration and temperature can be adjusted for the nucleotide sequence from a subject suspected of having AS or being at risk of developing AS, to be screened. Southern blotting and hybridizations protocols are described in Current Protocols in Molecular Biology (Greene Publishing Associates and Wiley-Interscience), pages 2.9.1-2.9.10. Probes can be labeled for hybridization with random oligomers and the Klenow fragment of DNA polymerase. Very high specific activity probes can be obtained using commercially available kits such as the Ready-To-Go DNA Labeling Beads (Pharmacia Biotech), following the manufacturer's protocol. Possible competition of probes having high repeat sequence content, and stringency of hybridization and wash down will be determined individually for each probe used. Alternatively, fragments of a candidate sequence may be generated by PCR, the specificity may be verified using a rodent-human somatic cell hybrid panel, and sub-cloning the fragment. This allows for a large prep for sequencing and use as a probe. Once a given gene fragment has been characterized, small probe preparations can be achieved by gel or column purifying the PCR product.

[0140] Suitable materials that can be used in the dot blot, reverse dot blot, multiplex, and MASDA formats are well-known in the art and include, but are not limited to nylon and nitrocellulose membranes.

3.6 Nucleotide Arrays and gene Chips for polymorphism analysis

[0141] The invention further contemplates methods of identifying SNPs through the use of an array of oligonucleotides, wherein discrete positions on the array are complementary to one or more of the provided polymorphic sequences, *e.g.* oligonucleotides of at least 12 nt, at least about 15 nt, at least about 18 nt, at least about 20 nt, or at least about 25 nt, or longer, and including the sequence flanking the polymorphic position. Such an array may comprise a series of oligonucleotides, each of which can specifically hybridize to a different polymorphism. For examples of arrays, see Hacia *et al.* (1996, *Nat. Genet.* **14**: 441-447 and De Risi *et al.*, (1996, *Nat. Genet.* **14**: 457-460).

[0142] A nucleotide array can include all or a subset of the polymorphisms of the invention. One or more polymorphic forms may be present in the array. In some embodiments, an array includes at least 2 different polymorphic sequences, *i.e.*, polymorphisms located at unique positions within the SpA/AS marker sequences of the present invention, and may include as many of the provided polymorphisms as required. Arrays of interest may further comprise sequences, including polymorphisms, of other genetic sequences, particularly other sequences of interest for pharmacogenetic screening, including, but not limited to, other genes associated with a SpA (*e.g.*, AS). The oligonucleotide sequence on the array is generally at least about 12 nt in length, at least about 15 nt, at least about 18 nt, at least about 20 nt, or at least about 25 nt, or may be the length of the provided polymorphic sequences, or may extend into the flanking regions to generate fragments of 100 to 200 nt in length. For examples of arrays, see Ramsay (1998, *Nature Biotech.* **16**: 40-44; Hacia *et al.*, (1996, *Nature Genetics* **14**: 441-447; Lockhart *et al.*, (1996, *Nature Biotechnol.* **14**:1675-1680; and De Risi *et al.*, (1996, *Nature Genetics* **14**: 457-460).

[0143] A number of methods are available for creating micro-arrays of biological samples, such as arrays of DNA samples to be used in DNA hybridization assays. Examples of such arrays are discussed in detail in PCT Application number. WO95/35505 (1995); U.S. Patent Application number. 5,445,934, (1995); and Drmanac *et al.*, (1993, *Science* **260**:1649-1652). Yershov *et al.*, (1996, *Genetics* **93**: 4913-4918) describe an alternative construction of

an oligonucleotide array. The construction and use of oligonucleotide arrays is reviewed by Ramsay (1998) *supra.*

[0144] Methods of using high density oligonucleotide arrays for identifying polymorphisms within nucleotide sequences are known in the art. For example, Milosavljevic *et al.*, (1996, *Genomics* **37**: 77-86) describe DNA sequence recognition by hybridization to short oligomers. See also, Drmanac *et al.*, (1998, *Nature Biotech.* **16**: 54-58; and Drmanac and Drmanac, 1999, *Methods Enzymol.* **303**: 165-178). The use of arrays for identification of unknown mutations is proposed by Ginot, (1997, *Human Mutation* **10**: 1-10).

[0145] Detection of known mutations is described in Hacia *et al.* (1996, *Nat. Genet.* **14**: 441-447; Cronin *et al.*, (1996) *Human Mut.* **7**: 244-255; and others. The use of arrays in genetic mapping is discussed in Chee *et al.*, (1996, *Science* **274**: 610-613; Sapolsky and Lishutz, 1996, *Genomics* **33**: 445-456; and Shoemaker *et al.*, 1996, *Nat. Genet.* **14**: 450-456) perform quantitative phenotypic analysis of yeast deletion mutants using a parallel bar-coding strategy.

[0146] Quantitative monitoring of gene expression patterns with a complementary DNA microarray is described in Schena *et al.*, (1995, *Science* **270**: 467; DeRisi *et al.*, 1997, *Science* **270**: 680-686) explore gene expression on a genomic scale. Wodicka *et al.*, (1997, *Nat. Biotech.* **15**: 1-15) perform genome wide expression monitoring in *S. cerevisiae.*

[0147] A DNA sample for analysis is prepared in accordance with conventional methods, *e.g.*, lysing cells, removing cellular debris, separating the DNA from proteins, lipids or other components present in the mixture and then using the isolated DNA for cleavage. See Molecular Cloning, A Laboratory Manual, 2nd ed. (eds. Sambrook *et al.*) CSH Laboratory Press, Cold Spring Harbor, N.Y. 1989. Generally, at least about 0.5 μg of DNA will be employed, usually at least about 5 μg of DNA, while less than 50 μg of DNA will usually be sufficient.

[0148] The nucleic acid samples are cleaved to generate probes. It will be understood by one of skill in the art that any method of random cleavage will generate a distribution of fragments, varying in the average size and standard deviation. Usually the average size will be at least about 12 nucleotides (nts) in length, more usually at least about 20 nts in length, and preferably at least about 35 nts in length. Where the variation in size is great, conventional methods may be used to remove the large and/or small regions of the fragment population.

[0149] It is desirable, but not essential to introduce breaks randomly, with a method which does not act preferentially on specific sequences. Preferred methods produce a reproducible pattern of breaks. Methods for introducing random breaks or nicks in nucleic acids include but are not restricted to reaction with Fenton reagent to produce hydroxyl radicals and other chemical cleavage systems, integration mediated by retroviral integrase, partial digestion with an ultra-frequent cutting restriction enzyme, partial digestion of single stranded DNA with SI nuclease, partial digestion with DNAseI in the absence or presence of Mn^{++}, etc.

[0150] The fragmented nucleic acid samples are denatured and labeled. Labeling can be performed according to methods well known in the art, using any method that provides for a detectable signal either directly or indirectly from the nucleic acid fragment. In a preferred embodiment, the fragments are end-labeled, in order to minimize the steric effects of the label. For example, terminal transferase may be used to conjugate a labeled nucleotide to the nucleic acid fragments. Suitable labels include biotin and other binding moieties; fluorochromes, *e.g.*, fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), and the like. Where the label is a binding moiety, the detectable label is conjugated to a second stage reagent, *e.g.*, avidin, streptavidin, etc., that specifically binds to the binding moiety, for example a fluorescent probe attached to streptavidin. Incorporation of a fluorescent label using enzymes such as reverse transcriptase or DNA polymerase, prior to fragmentation of the sample, is also possible.

[0151] Each of the labeled genome samples is separately hybridized to an array of oligonucleotide probes. Hybridization of the labeled sequences is accomplished according to methods well known in the art. Hybridization can be carried out under conditions varying in stringency, preferably under conditions of high stringency, *e.g.*, 6 × SSPE, at 65 °C., to allow for hybridization of complementary sequences having extensive homology, usually having no more than one or two mismatches in a probe of 25 nts in length, *i.e.*, at least 95% to 100% sequence identity.

[0152] High density microarrays of oligonucleotides are known in the art and are commercially available. The sequence of oligonucleotides on the array will correspond to a

known target sequences. The length of oligonucleotide present on the array is an important factor in how sensitive hybridization will be to the presence of a mismatch. Usually oligonucleotides will be at least about 12 nt in length, more usually at least about 15 nt in length, preferably at least about 20 nt in length and more preferably at least about 25 nt in length, and will be not longer than about 35 nt in length, usually not more than about 30 nt in length.

[0153] Methods of producing large arrays of oligonucleotides are described in U.S. Pat. No. 5,134,854 (Pirrung *et al.*), and U.S. Pat. No. 5,445,934 (Fodor *et al.*) using light-directed synthesis techniques. Using a computer controlled system, a heterogeneous array of monomers is converted, through simultaneous coupling at a number of reaction sites, into a heterogeneous array of polymers. Alternatively, microarrays are generated by deposition of pre-synthesized oligonucleotides onto a solid substrate, for example as described in International Publication WO 95/35505.

[0154] Microarrays can be scanned to detect hybridization of the labeled genome samples. Methods and devices for detecting fluorescently marked targets on devices are known in the art. Generally such detection devices include a microscope and light source for directing light at a substrate. A photon counter detects fluorescence from the substrate, while an x-y translation stage varies the location of the substrate. A confocal detection device that may be used in the subject methods is described in U.S. Pat. No. 5,631,734. A scanning laser microscope is described in Shalon *et al.* (1996, *Genome Res.* **6**: 639). A scan, using the appropriate excitation line, is performed for each fluorophore used. The digital images generated from the scan are then combined for subsequent analysis. For any particular array element, the ratio of the fluorescent signal from one Nucleic acid sample is compared to the fluorescent signal from the other Nucleic acid sample, and the relative signal intensity determined.

[0155] Methods for analyzing the data collected by fluorescence detection are known in the art. Data analysis includes the steps of determining fluorescent intensity as a function of substrate position from the data collected, removing outliers, *i.e.*, data deviating from a predetermined statistical distribution, and calculating the relative binding affinity of the targets from the remaining data. The resulting data may be displayed as an image with the intensity in each region varying according to the binding affinity between targets and probes.

[0156] Nucleic acid analysis *via* microchip technology is also applicable to the present invention. In this technique, thousands of distinct oligonucleotide probes can be applied in an array on a silicon chip. A nucleic acid to be analyzed is fluorescently labeled and hybridized to the probes on the chip. It is also possible to study nucleic acid-protein interactions using these nucleic acid microchips. Using this technique one can determine the presence of mutations, sequence the nucleic acid being analyzed, or measure expression levels of a gene of interest. The method is one of parallel processing of many, even thousands, of probes at once and can tremendously increase the rate of analysis.

[0157] Alteration of mRNA transcription can be detected by any techniques known to persons of ordinary skill in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA transcription indicates an alteration of the sequence.

[0158] The array/chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA 1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia *et al.*, 1996; Shoemaker *et al.*, 1996; Kozal *et al.*, 1996). Chips of various formats for use in detecting SNPs can be produced on a customized basis.

[0159] An array-based tiling strategy useful for detecting SNPs is described in EP 785280. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms. "Tiling" refers to the synthesis of a defined set of oligonucleotide probes that are made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, *e.g.*, substitution of one or more given positions with one or more members of the basis set of monomers, *i.e.*, nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In some embodiments, arrays are tiled for a number of specific SNPs. In particular, the array is tiled to include a number of detection blocks, each detection block being specific for a specific SNP or a set of SNPs. For example, a detection block may be tiled to include a number of probes that span the sequence segment that includes a specific SNP. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the SNP position. In addition to the probes differing at the SNP position, monosubstituted probes are also generally tiled within the detection block. Such methods can readily be applied to the SNP information disclosed herein.

[0160] These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically, the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the SNP. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artifactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the SNP are present in the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and U.S. Pat. No. 5,424,186.

[0161] Thus, in some embodiments, the chips may comprise an array of nucleic acid sequences of fragments of about 15 nucleotides in length and the sequences complementary thereto, or a fragment thereof, the fragment comprising at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides and containing a polymorphic base. In some embodiments the polymorphic base is within 5, 4, 3, 2, or 1 nucleotides from the center of the polynucleotide, more preferably at the center of the polynucleotide. In other embodiments, the chip may comprise an array containing any number of polynucleotides of the present invention.

[0162] An oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschwieler *et al.*). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number which lends itself to the efficient use of commercially available instrumentation.

[0163] Using such arrays, the present invention provides methods of identifying the SNPs of the present invention in a sample. Such methods comprise incubating a test sample with an array comprising one or more oligonucleotide probes corresponding to at least one

SNP position of the present invention, and assaying for binding of a nucleic acid from the test sample with one or more of the oligonucleotide probes. Such assays will typically involve arrays comprising oligonucleotide probes corresponding to many SNP positions and/or allelic variants of those SNP positions, at least one of which is a SNP of the present invention.

[0164] Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel SNPs disclosed herein. Examples of such assays can be found in Chard, T, An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, Techniques in Immunocytochemistry, Academic Press, Orlando, Fla. Vol. 1 (I 982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

[0165] The samples of the present invention include, but are not limited to, nucleic acid extracts, cells, and protein or membrane extracts from cells, which may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. The test sample used in the above-described methods will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods of preparing nucleic acid, protein, or cell extracts are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

[0166] Multicomponent integrated systems may also be used to analyze SNPs. Such systems miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in U.S. Pat. No. 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips.

[0167] Integrated systems can be envisaged mainly when micro-fluidic systems are used. These systems comprise a pattern of micro-channels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electro-osmotic or hydrostatic forces applied across different areas of

the microchip to create functional microscopic valves and pumps with no moving parts. Varying the voltage controls the liquid flow at intersections between the micro-machined channels and changes the liquid flow rate for pumping across different sections of the microchip.

[0168] For genotyping SNPs, the microfluidic system may integrate, for example, nucleic acid amplification, mini-sequencing primer extension, capillary electrophoresis, and a detection method such as laser induced fluorescence detection.

[0169] In a first step, the DNA samples are amplified, preferably by PCR. Then, the amplification products are subjected to automated mini-sequencing reactions using ddNTPs (specific fluorescence for each ddNTP) and the appropriate oligonucleotide mini-sequencing primers which hybridize just upstream of the targeted polymorphic base. Once the extension at the 3' end is completed, the primers are separated from the unincorporated fluorescent ddNTPs by capillary electrophoresis. The separation medium used in capillary electrophoresis can be, for example, polyacrylamide, polyethyleneglycol or dextran. The incorporated ddNTPs in the single nucleotide primer extension products are identified by laser-induced fluorescence detection. This microchip can be used to process at least 96 to 384 samples, or more, in parallel.

3.7 Extension based techniques for the detection of polymorphisms

[0170] Extension based techniques for detecting polymorphisms within a nucleotide sequence can include, but are not restricted to allele-specific amplification, also known as the amplification refractory mutation system (ARMS) as disclosed in European Patent Application Publication No. 0332435 and in Newton *et al.*, (1989, *Nucl. Acids Res.* **17**: 2503-2516), and cloning of polymorphisms (COPS) as contemplated by Gibbs *et al.*, (1989, *Nucleic Acids Research*, **17**: 2347).

[0171] The extension based technique, ARMS, uses allele specific oligonucleotide (ASO) PCR primers for genotyping. In this approach, one of the two oligonucleotide primers used for PCR is designed to bind to the mutation site, most commonly with the 3' end of the primer targeting the mutation site. Under carefully controlled conditions (annealing temperature, magnesium concentration etc.), amplification only takes place if the nucleotide at the 3' end of the PCR primer is complementary to the base at the mutation site, with a mismatch being "refractory" to amplification. If the 3' end of the primer is designed to be complementary to the normal gene, then PCR products should be formed when amplifying the

normal gene but not genes with the mutation, and vice versa. There are numerous variations of the approach, for example, one of the simplest embodiments comprises where two amplifications are carried out, one using a primer specific for the normal gene, and a second using a primer specific for the mutant gene. This is followed by gel electrophoresis and ethidium bromide staining to detect the presence of amplified products.

[0172] A variation of the ARMS approach, termed mutagenically separated PCR (MS-PCR), comprises two ARMS primers of different lengths, one specific for the normal gene and one for the mutation. This method yields PCR products of different lengths for the normal and mutant alleles. Subsequent gel electrophoresis shows at least one of the two allelic products.

[0173] In some embodiments, Cloning of polymorphisms (COPs) can be applicable to the isolation of SNPs from particular regions of the genome, *e.g.*, CpG islands, chromosomal bands, YACs or PAC contigs.ALEX. For example, Li *et al.*, (2000, *Nucleic Acid Research*, **28**(2): e1) disclose a combination of nucleic acid sequence digestion with restriction enzymes, treatment with uracil-DNA glycosylase and mung bean nuclease, PCR amplification and purification with streptavidin magnetic beads to isolate polymorphic sequences from the genomes of two human samples.

3.8 Ligation based assays for detecting polymorphisms

[0174] Another typical method of SNP detection encompasses the oligonucleotide ligation assay. A number of approaches make use of DNA ligase, an enzyme that can join two adjacent oligonucleotides hybridized to a DNA template. The specificity of the approach comes from the requirement for a perfect match between the hybridized oligonucleotides and the DNA template at the ligation site. In the oligonucleotide ligation assay (OLA), or ligase chain reaction (LCR) assay the sequence surrounding the mutation site is first amplified, and one strand serves as a template for three ligation probes, two of these are allele specific oligonucleotides (ASO) and the third a common probe. Numerous approaches can be used for the detection of the ligated products. For example, the two ASOs can be differentially labeled with fluorescent or hapten labels and ligated products detected by fluorimetric or colorimetric enzyme-linked immunosorbent assays, respectively. For electrophoresis-based systems, use of mobility modifier tags or variation in probe lengths coupled with fluorescence detection enables the multiplex genotyping of several single nucleotide substitutions in a single tube. When used on arrays, ASOs can be spotted at specific locations or addresses on a chip. PCR

amplified DNA can then be added and ligation to labeled oligonucleotides at specific addresses on the array can be measured.

3.9 Signal generating polymorphism detection assays

[0175] In some embodiments, fluorescence resonance energy transfer (FRET) is contemplated as a method to identify a polymorphism within any one or more of the AS marker sequences of the present invention. FRET occurs due to the interaction between the electronic excited states of two dye molecules. The excitation is transferred from one (the donor) dye molecule to the other (the acceptor) dye molecule without emission of a photon. This is distance-dependent, that is the donor and the acceptor dye must be in close proximity. The hybridization probe system consists of two oligonucleotides labeled with fluorescent dyes. The hybridization probe pair is designed to hybridize to adjacent regions on the target DNA. Each probe is labeled with a different marker dye. Interaction of the two dyes can only occur when both are bound to their target. The donor probe is labeled with fluorophore at the 3' end and the acceptor probe at the 5' end. During PCR, the two different oligonucleotides hybridize to adjacent regions of the target DNA such that the fluorophores, which are coupled to the oligonucleotides, are in close proximity in the hybrid structure. The donor fluorophore (F1) is excited by an external light source, and then passes part of its excitation energy to the adjacent acceptor fluorophore (F2). The excited acceptor fluorophore (F2) emits light at a different wavelength which can then be detected and measured for molecular proximity.

[0176] In other embodiments, the MagSNiPer method, based on single base extension, magnetic separation, and chemiluminescence provides a further method for SNP identification in a nucleotide sequence. Single base nucleotide extension reaction is performed with a biotinylated primer whose 3' terminus is contiguous to the SNP site with a tag-labeled ddNTP. Then the primers are captured by magnetic-coated beads with streptavidin, and unincorporated labeled ddNTP is removed by magnetic separation. The magnetic beads are incubated with anti-tag antibody conjugated with alkaline phosphatase. After the removal of excess conjugates by magnetic separation, SNP typing is performed by measuring chemiluminescence. The incorporation of labeled ddNTP is monitored by chemiluminescence induced by alkaline phosphatase.

[0177] In some embodiments, fluorescence polarization provides a method for identifying polymorphisms within a nucleotide sequence. For example, amplified DNA containing a polymorphic is incubated with oligonucleotide primers (designed to hybridize to

the DNA template adjacent to the polymorphic site) in the presence of allele-specific dye-labeled dideoxyribonucleoside triphosphates and a commercially available modified *Taq* DNA polymerase. The primer is extended by the dye-terminator specific for the allele present on the template, increasing approximately 10-fold the molecular weight of the fluorophore. At the end of the reaction, the fluorescence polarization of the two dye-terminators in the reaction mixture are analyzed directly without separation or purification. This homogeneous DNA diagnostic method is shown to be highly sensitive and specific and is suitable for automated genotyping of large number of samples.

[0178] In other embodiments, surface enhanced Raman scattering can be used as a method for detecting and identifying single base differences in double stranded DNA fragments. Chumanov, G. "Surface Enhanced Raman Scattering (SERS) for Discovering and Scoring Single Based Differences in DNA" Proc. Volume SPIE, 3608 (1999). SERS has also been used for single molecule detection. Kneipp, K, (1997, *Physical Review Letters*, **78**(9): 1667-1670). SERS results in strongly increased Raman signals from molecules which have been attached to nanometer sized metallic structures.

[0179] Illustrative examples include a genotyping method discussed by Xiao and Kwok (2003, *Genome Research*, **13**(5): 932-939) based on a primer extension assay with fluorescence quenching as the detection. The template-directed dye-terminator incorporation with fluorescence quenching detection (FQ-TDI) assay is based on the observation that the intensity of fluorescent dye R110- and R6G-labeled acycloterminators is universally quenched once they are incorporated onto a DNA oligonucleotide primer. By comparing the rate of fluorescence quenching of the two allelic dyes in real time, the frequency of SNPs in DNA samples can be measured. The kinetic FQ-TDI assay is highly accurate and reproducible both in genotyping and in allele frequency estimation.

4. Vectors

[0180] Described herein are systems of vectors and host cells that can be used for the expression of at least a portion of a SpA or AS marker sequence of the present invention. A variety of expression vectors may be used in the present invention which include, but are not limited to, plasmids, cosmids, phage, phagemids, or modified viruses. Typically, such expression vectors comprise a functional origin of replication for propagation of the vector in an appropriate host cell, one or more restriction endonuclease sites for insertion of the SpA or AS marker sequence, and one or more selection markers. The expression vector can be used

with a compatible host cell which may be derived from a prokaryotic or a eukaryotic organism including but not limited to bacteria, yeasts, insects, mammals, and humans.

[0181] Where the SpA/AS markers of the present invention contain transcribable sequences, those sequences in whole or in part are suitably rendered expressible in a host cell by operably linking them with a regulatory polynucleotide. The synthetic construct or vector thus produced may be introduced firstly into an organism or part thereof before subsequent expression of the construct in a particular cell or tissue type. Any suitable organism is contemplated by the invention, which may include unicellular as well as multi-cellular organisms. Suitable unicellular organisms include bacteria. Exemplary multi-cellular organisms include yeast, mammals and plants.

[0182] The construction of the vector may be carried out by any suitable technique as for example described in the relevant sections of Ausubel *et al.*, (*supra*) and Sambrook *et al.*, ("Molecular Cloning. A Laboratory Manual", Cold Spring Harbour Press, 1989). However, it should be noted that the present invention is not dependent on and not directed to any one particular technique for constructing the vector.

[0183] Regulatory polynucleotides which may be utilised to regulate expression of the polynucleotide include, but are not limited to, a promoter, an enhancer, and a transcriptional terminator. Such regulatory sequences are well known to those of skill in the art. Suitable promoters that may be utilised to induce expression of the polynucleotides of the invention include constitutive promoters and inducible promoters.

5. Amino acid polymorphism screening techniques

[0184] As described above, where the particular nucleotide occurrence of a SNP is such that the nucleotide occurrence results in an amino acid change in the encoded polypeptide, the nucleotide occurrence can be identified indirectly by detecting a particular mutation or variation in the sequence of the polypeptide. For example, the *ERAP1* and *IL23R* polymorphisms disclosed in Table 2 may comprise a non-synonymous base substitution in the coding region of the corresponding gene, which results in variation in the amino acid sequence. In representative examples relating to the *ERAP1* coding sequence, the polymorphism at rs27044 is associated with a non-synonymous variation in the coding sequence of the *ERAP1* gene, wherein the presence of G at rs27044 is associated with Gln (glutamine) at position 730 of the ERAP-1 polypeptide (as set forth for example in GenPept Accession No. NP_057526 [GI:94818901] or SEQ ID NO: 2 of WO 2008/144827), and

wherein the presence of C at rs27044 is associated with Glu (glutamate) at position 730 of the ERAP-1 polypeptide. In other representative examples relating to the *ERAP1* coding sequence, the polymorphism at rs17482078 is associated with a non-synonymous variation in the *ERAP1* coding sequence, wherein the presence of C at rs17482078 is associated with Arg (arginine) at position 725 of the ERAP-1 polypeptide, and wherein the presence of T at rs17482078 is associated with Gln at position 725 of the ERAP-1 polypeptide. In still other representative examples relating to the *ERAP1* coding sequence, the polymorphism at rs10050860 is associated with a non-synonymous variation in the *ERAP1* coding sequence, wherein the presence of C at rs10050860 is associated with Asp (aspartic acid) at position 575 of the ERAP-1 polypeptide, and wherein the presence of T at rs10050860 is associated with Asn (asparagine). In still other representative examples relating to the *ERAP1* coding sequence, the polymorphism at rs2287987 is associated with a non-synonymous variation in the *ERAP1* coding sequence, wherein the presence of T at rs2287987 is associated with Met (methionine) at position 349 of the ERAP-1 polypeptide, and wherein the presence of C at rs2287987 is associated with Val (valine). In still other representative examples relating to the *ERAP1* coding sequence, the polymorphism at rs30187 is associated with a non-synonymous variation in the *ERAP1* coding sequence, wherein the presence of T at rs30187 is associated with Lys (lysine) at position 528 of the ERAP-1 polypeptide, and wherein the presence of C at rs30187 is associated with Arg.

[0185] In a non-limiting example relating to the *IL23R* coding sequence, the polymorphism at rs11209026 is associated with a non-synonymous variation in the coding sequence of the *IL23R* gene, wherein the presence of G at rs11209026 is associated with Arg at position 381 of the IL-23R polypeptide (as set forth for example in GenPept Accession No. NP_653302 [GI:24430212] or SEQ ID NO: 4 WO 2008/144827), and wherein the presence of A at rs11209026 is associated with Gln at position 381 of the IL-23R polypeptide.

5.1 Protein Truncation Test (PTT)

[0186] In some embodiments, the PTT can be used to identify polymorphisms within a protein sequence. PTT uses *in vitro* transcription and translation of the cDNA generated to focus on mutations that generate proteins with an altered size; shorter proteins caused by premature translation termination. For some genes containing large exons, PTT can also be performed using a genomic DNA target (Hogervorst, F. B. L., 1997, *Promega Notes Magazine*, **62**: 7-11).

[0187] Thus, in the above embodiment, the coding region of a gene is screened for the presence of translation terminating mutations using *de novo* protein synthesis from the amplified copy. The procedure includes three important steps. The first step involves the isolation of genomic DNA and amplification of the target gene coding sequences using PCR or, alternatively, isolation of RNA and amplification of the target sequence using Reverse Transcription PCR (RT-PCR). The resulting PCR products are then used as a template for the *in vitro* synthesis of RNA, which is subsequently translated into protein. The final step is the SDS-PAGE analysis of the synthesized protein. The shorter protein products of mutated alleles are easily distinguished from the full length protein products of normal alleles.

[0188] Mutant truncated proteins can result from for example, nonsense substitution mutations, frameshift mutations, in-frame deletions, and splice site mutations.

[0189] For example, a nonsense substitution mutation occurs when a nucleotide substitution causes a codon that normally encodes an amino acid to code for one of the three stop signals (TGA, TTA, TAG). For such mutations, the protein truncation point occurs at the corresponding position in the gene at which the mutation occurs.

[0190] Frameshift mutations result from the addition or deletion of any number of bases that is not a multiple of three (*e.g.*, one or two base insertion or deletion). For such frameshift mutations, the reading frame is altered from the point of mutation downstream. A stop codon, and resulting truncation of the corresponding encoded protein product, can occur at any point from the position of the mutation downstream.

[0191] In-frame deletions result from the deletion of one or more codons from the coding sequence. The resulting protein product lacks only those amino acids that were encoded by the deleted codons.

[0192] Splice site mutations result in an improper excision and/or joining of exons. These mutations can result in inclusion of some or all of an intron in the mRNA, or deletion of some or all of an exon from the mRNA. In some instances, these insertions or deletions result in stop codon being encountered prematurely, as typically occurs with frameshift mutations. In other instances, one or more specific exons are deleted from the mature mRNA in such a manner that the proper reading frame is maintained for the remaining exons, *i.e.*, non-contiguous exons are fused in frame with each other. For such splice mutations, the encoded protein may terminate at the appropriate stop codon, but is shortened by the absence of the un-spliced internal exon.

5.2 Protein sequencing

[0193] In some embodiments, sequencing of a polypeptide may be performed by site-directed or random cleavage of the polypeptide using, for example endopeptidases or CNBr, to produce a set of polypeptide fragments and subsequent sequencing of the polypeptide fragments by, for example, Edman sequencing or mass spectrometry, as is known in the art. Alternatively, the polypeptide probes or polypeptide fragments could be sequenced by use of antibody probes as for example described by Fodor *et al* in U.S. Patent Serial No. 5,871,928. Briefly, such antibody probes specifically recognise particular subsequences (*e.g.*, at least three contiguous amino acids) found on a polypeptide. Optimally, these antibodies would not recognise any sequences other than the specific desired subsequence and the binding affinity should be insensitive to flanking or remote sequences found on a target molecule.

[0194] The Edman degradation process is commonly used, while other methods have been developed and can be used in certain instances. In the Edman degradation method, amino acid removal from the end of the protein is accomplished by reacting the N-terminal amino acid residue with a reagent which allows selective removal of that residue from the protein. The resulting amino acid derivative is converted into a stable compound which can be chemically removed from the reaction mixture and identified.

[0195] Most current chemical sequencing methods are done with an amount of protein in the 5-100 nm range. It has been reported that micro-sequencing of polypeptides by reverse phase high pressure liquid chromatography using ultraviolet light detection means has been accomplished with protein samples in the range of 50-500 pm. Other methods used in the micro-sequencing of polypeptides involves radio labeling of the peptide or reagent, intrinsic radio labeling of the polypeptide, and enhanced UV detection of sequence degradation products, and others.

[0196] It is possible to determine the C-terminus sequence of peptides and proteins using a combination of Matrix-Assisted Laser Desorption/Ionization-Time Of Flight-Mass Spectrometry (MALDI-TOF-MS) and enzymatic digestions using for example, the Applied Biosystems Sequazyme technology. In some illustrative examples, Carboxypeptidase Y is a non-specific exoprotease, which sequentially cleaves all residues, including proline, from the C-terminus. This generates a nested set of fragments that form a sequence "ladder." The masses of individual members of the set are determined by MALDI-TOF-MS, and the amino

acids are identified from the unique mass differences between peaks. Trace quantities of peptides and proteins, as little as 2 pmol, can be analyzed. Up to 20 residues can be identified in less than 30 minutes. Aminopeptidase can similarly be used to generate N-terminal ladders from the peptides.

[0197] In some embodiments, peptides can be fragmented by either post-source decay (PSD) or collision-induced dissociation (CID) for use in MS/MS studies. The process of PSD starts as the peptide is ionized using a higher than normal laser power to pump more energy into the peptide. PSD is also facilitated by the selection of a matrix that is more favorable to promoting fragmentation. The ionized peptides are extracted from the ion source and gain full kinetic energy necessary for mass analysis. As the ions travel down the flight tube, those having excess internal energy must change. If enough energy is localized in a single bond, it will break apart, producing a product ion and a neutral fragment. Product ions come in many forms which can include N-terminal, C-terminal, and internal fragments. The ion reflector separates ions based on their kinetic energy. When ions enter the reflector, they experience an electric field that reverses their direction. The product ions have kinetic energies that are directly proportional to the ratio between the product ion mass and the peptide precursor mass. For low mass product ions, those having low kinetic energy, the reflection shortens their flight path, reducing the time required to reach the detector. For higher mass ions, those having a higher kinetic energy, reflection lengthens their flight path, increasing the time of flight to the detector. Modulation of the potential applied to the ion reflector enables collection of high quality PSD spectra with good mass accuracy.

[0198] In CID, the peptide ion interacts with a collision gas to modulate the internal energy and promote fragmentation. As with PSD, fragmentation does not change the velocity of the ions once they are in the flight tube, so the peptide precursor ion and product ions only separate when they encounter the ion reflector.

5.3 Immunohistology

[0199] In some embodiments immunohistochemical analysis of a tissue sample from a subject suspected of having a SpA (*e.g.*, AS) or being at risk of developing a SpA can be employed to detect the presence of a related sequence polymorphism. For examples, antibodies specific to the region of the protein sequence suspected of containing the polymorphism can be raised and used in a visual test to identify polymorphisms. Specifically, tissue samples can be probed with an antibody of choice before detecting the level of bound

antibody and comparing it with a control sample. To enhance visual detection, the secondary antibody can be conjugated with a fluorophore such as Texas Red.

5.4 Immunoassays .

5.4.1 Antigen-binding molecules

[0200] The invention also contemplates antigen-binding molecules that bind specifically to the polypeptide encoded by any of the AS markers of the invention or to a fragment of said polypeptide for use in diagnosing the presence or risk of development of AS in a subject. For example, the antigen-binding molecules may comprise whole polyclonal antibodies. Such antibodies may be prepared, for example, by injecting a polypeptide of the invention or fragment thereof into a production species, which may include mice or rabbits, to obtain polyclonal antisera. Methods of producing polyclonal antibodies are well known to those skilled in the art. Exemplary protocols which may be used are described for example in Coligan *et al.*, 1991, *Current Protocols in Immunology*, (John Wiley & Sons, Inc) and Ausubel *et al.*, (1994-1998, *supra*), in particular Section III of Chapter 11.

[0201] In lieu of the polyclonal antisera obtained in the production species, monoclonal antibodies may be produced using the standard method as described, for example, by Köhler and Milstein (1975, *Nature* **256**, 495-497), or by more recent modifications thereof as described, for example, in Coligan *et al.*, (1991, *supra*) by immortalizing spleen or other antibody producing cells derived from a production species which has been inoculated with a polypeptide of the invention or a fragment thereof.

[0202] The invention also contemplates as antigen-binding molecules Fv, Fab, Fab' and $F(ab')_2$ immunoglobulin fragments. Alternatively, the antigen-binding molecule may comprise a synthetic stabilised Fv fragment. Exemplary fragments of this type include single chain Fv fragments (sFv, frequently termed scFv) in which a peptide linker is used to bridge the N terminus or C terminus of a V_H domain with the C terminus or N-terminus, respectively, of a V_L domain. ScFv lack all constant parts of whole antibodies and are not able to activate complement. Suitable peptide linkers for joining the V_H and V_L domains are those which allow the V_H and V_L domains to fold into a single polypeptide chain having an antigen binding site with a three dimensional structure similar to that of the antigen binding site of a whole antibody from which the Fv fragment is derived. Linkers having the desired properties may be obtained by the method disclosed in U.S. Patent No 4,946,778. However, in some cases a linker is absent. ScFvs may be prepared, for example, in accordance with methods

outlined in Kreber *et al.*, (1997, *J. Immunol. Methods*; **201(1)**: 35-55). Alternatively, they may be prepared by methods described in U.S. Patent No 5,091,513, European Patent No 239,400 or the articles by Winter and Milstein (1991, *Nature* **349**:293) and Plünckthun *et al.*, (1996, In *Antibody engineering*: *A practical approach*. 203-252).

[0203] Alternatively, the synthetic stabilised Fv fragment comprises a disulphide stabilised Fv (dsFv) in which cysteine residues are introduced into the V_H and V_L domains such that in the fully folded Fv molecule the two residues will form a disulphide bond there between. Suitable methods of producing dsFv are described for example in (Glockscuther *et al.*, *Biochem.* **29**: 1363-1367; Reiter *et al.*, 1994, *J. Biol. Chem.* **269**: 18327-18331; Reiter *et al.*, 1994, *Biochem.* **33**: 5451-5459; Reiter *et al.*, 1994. *Cancer Res.* **54**: 2714-2718; and Webber *et al.*, 1995, *Mol. Immunol.* **32**: 249-258).

[0204] Also contemplated as antigen-binding molecules are single variable region domains (termed dAbs) as for example disclosed in (Ward *et* al.,1989, *Nature* **341**: 544-546; Hamers-Casterman *et al.*, 1993, *Nature*. **363**: 446-448; and Davies & Riechmann, 1994, *FEBS Lett.* **339**: 285-290).

[0205] Alternatively, the antigen-binding molecule may comprise a "minibody". In this regard, minibodies are small versions of whole antibodies, which encode in a single chain the essential elements of a whole antibody. Suitably, the minibody is comprised of the V_H and V_L domains of a native antibody fused to the hinge region and CH3 domain of the immunoglobulin molecule as, for example, disclosed in U.S. Patent No 5,837,821.

[0206] In an alternate embodiment, the antigen binding molecule may comprise non-immunoglobulin derived, protein frameworks. For example, reference may be made to (Ku & Schultz, 1995, *Proc. Natl. Acad. Sci. USA*, **92**: 652-6556) which discloses a four-helix bundle protein cytochrome b562 having two loops randomised to create complementarity determining regions (CDRs), which have been selected for antigen binding.

[0207] The antigen-binding molecule may be multivalent (*i.e.*, having more than one antigen-binding site). Such multivalent molecules may be specific for one or more antigens. Multivalent molecules of this type may be prepared by dimerisation of two antibody fragments through a cysteinyl-containing peptide as, for example disclosed by (Adams *et al.*, 1993, *Cancer Res.* **53**: 4026-4034; Cumber *et al.*, 1992, *J. Immunol.* **149**: 120-126). Alternatively, dimerisation may be facilitated by fusion of the antibody fragments to amphiphilic helices that naturally dimerise (Pack P. Plünckthun, 1992, *Biochem.* **31**: 1579-

1584), or by use of domains (such as the leucine zippers jun and fos) that preferentially heterodimerize (Kostelny *et al.*, 1992, *J. Immunol.* **148**: 1547-1553). In an alternate embodiment, the multivalent molecule may comprise a multivalent single chain antibody (multi-scFv) comprising at least two scFvs linked together by a peptide linker. In this regard, non-covalently or covalently linked scFv dimers termed "diabodies" may be used. Multi-scFvs may be bispecific or greater depending on the number of scFvs employed having different antigen binding specificities. Multi-scFvs may be prepared for example by methods disclosed in U.S. Patent No. 5,892,020.

5.5 Protein arrays

[0208] In some embodiments, the variant or polymorphic polypeptides can be detected through the use of protein arrays. Protein arrays may comprise a surface upon which are deposited at specially defined locations at least two protein moieties characterised in that the protein moieties are those of the sequence of interest. The protein moieties can be attached to the surface either directly or indirectly. The attachment can be non-specific (*e.g.*, by physical absorption onto the surface or by formation of a non-specific covalent interaction). In some embodiments the protein moieties are attached to the surface through a common marker moiety appended to each protein moiety. In another embodiment, the protein moieties can be incorporated into a vesicle or liposome which is tethered to the surface. An example of such a protein array is described in Frank, R (2002, *Comb. Chem.* **5**: 429-440).

[0209] In an alternate embodiment, the non-synonymous SNPs of the invention can be detected through the use of antibody arrays. In a similar manner to RNA profiling on DNA chips, antibody arrays can be employed for overlay assays to identify and quantify proteins and their specific amino acids. An illustrative example of this type is the protein binding assay, wherein an antibody array is overlayed with protein complexes and specific antibodies can detect potential binding partners of the proteins bound to the array (Wang *et al.*, 2000, *Mol. Cell. Biol*, **20**: 4505-4512; and Maercker, *Bioscience Reports*, **25**(1/2): 57-70).

6. Polymorphism sequence analysis

[0210] Further contemplated by the present invention is the analysis of samples from subjects suspected or having AS or at risk of developing AS using a sequence analysis program. For example, the sequence analysis program may be in the form of a computer program for use in homology searching, mapping, haplotyping, genotyping or pharmacogenetic analysis. The information gained from the analysis can be in any computer

readable format and can comprise any composition of matter used to store information or data, including, for example, floppy disks, tapes, chips, compact disks, video disks, punch cards or hard drives to name but a few.

*7. **Kits***

[0211] All the essential materials and reagents required for detecting SpA or AS-associated polymorphisms in at least a portion of a SpA or AS marker sequence according to the invention may be assembled together in a kit. The kits may also optionally include appropriate reagents for detection of labels, positive and negative controls, washing solutions, blotting membranes, microtitre plates dilution buffers and the like. For example, a nucleic acid-based detection kit for the identification of polymorphisms may include one or more of the following: (i) a SpA marker polynucleotide (which may be used as a positive control); (ii) an AS marker polynucleotide (which may be used as a positive control), (iii) a primer or probe that specifically hybridizes to at least a portion of a SpA marker polynucleotide or an AS marker polynucleotide (*e.g.*, a polynucleotide comprising or consisting essentially of rs116488202, *IL1R1* gene locus, *IL1R2* gene locus, 2Q31.3 chromosomal locus, *ERAP1* gene, *IL23R* gene, *TNFR1* gene locus, *TRADD* gene locus, 2P15 chromosome locus or 21Q22 chromosome locus sequences), and optionally one or more other AS markers, at or around the suspected SNP site. Also included may be enzymes suitable for amplifying nucleic acids including various polymerases (Reverse Transcriptase, *Taq*, Sequenase™ DNA ligase etc. depending on the nucleic acid amplification technique employed), deoxynucleotides and buffers to provide the necessary reaction mixture for amplification. Such kits also generally will comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each primer or probe. The kit can also feature various devices and reagents for performing one of the assays described herein; and/or printed instructions for using the kit to identify the presence of a SpA- or AS-associated polymorphism as defined herein. The kit may further contain reagents (*e.g.*, primers, probes or antigen-binding molecules) for detecting the presence of other AS markers, illustrative examples of which include an *HLA-B*27* allele of the *HLA-B* gene and its expression products.

[0212] In some embodiments, the kit may comprise appropriate agents for the detection of polymorphisms within the AS- marker polypeptides defined herein by Mass Spectrometry (MS). In illustrative examples of this type, an MS polymorphism detection kit may comprise (i) a vector that expresses an AS marker polynucleotide with at least one AS-

associated polymorphism for the expression of an AS marker polypeptide in a host cell (which may be used as a positive control) (ii) enzymes for digesting the expressed polypeptide, comprising for example non-specific exoproteases; and (iii) polypeptide fragments (which may be used as positive controls). The kit can also feature various devices and reagents for performing MS or any related form of MS known in the art; and/or printed instructions for using the kit to identify the presence of an AS-associated polymorphism within the AS marker polypeptide as described, for example, above.

8. Methods of Managing AS

[0213] The present invention also extends to the management of a SpA (*e.g.*, AS), or prevention of further progression of the SpA, or assessment of the efficacy of therapies in subjects following positive diagnosis for the presence of a SpA/AS-associated polymorphism in the subjects. Generally, the management of an SpA often includes a treatment regime involving medication, exercise, physical therapy and if necessary surgery. Examples of effective medications include but are not restricted to nonsteroidal anti-inflammatory drugs (NSAIDS) such as Sulfasalazine (Azulfidine), Methotrexate (Rheumatrex or Trexall) and Corticosteroids (cortisone); TNF blockers such as etanercept (Enbrel), infliximab (Remicade) and adalimumab (Humira);

[0214] It will be understood, however, that the present invention encompasses the use of any agent or process that is useful for treating or preventing a SpA (*e.g.*, AS) and is not limited to the aforementioned illustrative management strategies and compounds.

[0215] Typically, SpA-ameliorating agents will be administered in pharmaceutical (or veterinary) compositions together with a pharmaceutically acceptable carrier and in an effective amount to achieve their intended purpose. The dose of active compounds administered to a subject should be sufficient to achieve a beneficial response in the subject over time such as a reduction in, or relief from, the symptoms of the SpA (*e.g.*, AS) and the prevention of the disease from developing further. The quantity of the pharmaceutically active compounds(s) to be administered may depend on the subject to be treated inclusive of the age, sex, weight and general health condition thereof. In this regard, precise amounts of the active compound(s) for administration will depend on the judgement of the practitioner. In determining the effective amount of the active compound(s) to be administered in the treatment or prevention of an SpA (*e.g.*, AS), the physician or veterinarian may evaluate severity of any symptom associated with the presence of the SpA. For example, symptoms

related to AS include acute, painful episodes followed by temporary periods of remission. In any event, those of skill in the art may readily determine suitable dosages of the AS-ameliorating agents and suitable treatment regimens without undue experimentation.

[0216] In order that the invention may be readily understood and put into practical effect, particular preferred embodiments will now be described by way of the following non-limiting examples.

EXAMPLES

EXAMPLE 1

HLA ASSOCIATIONS OF AS

[0217] Classical HLA-loci were imputed in white Europeans to perform a case-control study. *HLA-B*27* tagging used the SNP rs116488202, which was shown to tag *HLA-B*27* in both western European and Asian populations better than the previous reported tagSNP rs4349859 (*see*, WO2011/063474), and the SNP rs13202464, reported by Lin *et al.* (2011, *supra*) to tag *HLA-B*27* in Asian populations (*see*, Table 3). Strong association was observed with *HLA-B*27* (cases 82% HLA-B*27 positive, controls 9.1% B*27 positive, OR 46 ($P<10^{-100}$)) (Figure 1). The risk of AS was further increased in *HLA-B*27* homozygotes; B*27-homozygosity was more prevalent in B*27-positive AS cases than in B*27-positive controls (OR=2.07, P=0.001).

[0218] Controlling for the association with *HLA-B*27*, residual signal was observed in the *HLA-A* and *HLA-B* loci (Figure 1). The residual signal at *HLA-B* may reflect either imperfect *HLA-B*27* tagging by rs116488202, or association of other *HLA-B* alleles with AS. However, no other individual non-B*27 allele was associated with AS however, including HLA-B*40, previously reported to be associated with AS.

[0219] From this it can be concluded that rs116488202 is in near 100% correlation with *HLA-B*27*, and can tag *HLA-B*27* as accurately as *HLA-B*27* genotyping in cases and in healthy controls. Therefore it can be used in place of *HLA-B*27* for genetic diagnosis of *HLA-B*27* related spondyloarthropathies such as AS, reactive arthritis, anterior uveitis, and colitic arthritis.

TABLE 3

*HLA-B*27* TAGGING PERFORMANCE BY PREVIOUSLY REPORTED TAG SNPS AND RS116488202

TABLE 3A: Validation datasets used to asses *HLA-B*27* performance. Samples of British, Australia and China with PCR based *HLA-B*27* typing.

	British & Australian		China	
	Cases	Controls	Cases	Controls
Number	542	754	104	5
*HLA-B*27* positive	489 (90.2%)	67 (8.9%)	89 (85.6%)	0 (0.0%)

TABLE 3B: *HLA-B*27* tagging performance.

SNP	Position [a]	Alleles (Minor/Major) [b]	British & Australian				China			
			MAF [c]	r^2	Sensitivity	Specificity	MAF [d]	r^2	Sensitivity	Specificity
rs13202464	31,344,583	G/A	0.064	0.826	0.989	0.934	0.096	0.821	0.989	0.917
rs116488202	31,344,916	T/C	0.042	0.951	0.989	0.985	0.011	0.913	0.989	1.000
rs4349859	31,365,787	A/G	0.042	0.930	0.976	0.988	0.002	0.033	0.234	1.000

a) Chromosome 6 NCBI human genome build 37 coordinates.
b) Minor allele tags the HLA-B*27 allele.
c) Minor allele frequency from 1000 Genomes Project Phase I haplotypes (2010-11 data freeze) in European reference panel.
d) Minor allele frequency from 1000 Genomes Project Phase I haplotypes (2010-11 data freeze) in Asian reference panel.

METHODS

SAMPLES

[0220] All cases had definite AS according to the modified New York criteria (van der Linden, S., *et al.*, 1984. *Arthritis Rheum* **27**, 361-368). Written informed consent was obtained from all cases with approval from the relevant research ethics authorities at each participating centre. A total of 12,252 DNA samples passed genotyping control filters. The cases collection consisted of 10,417 individuals of European ancestry, 1,560 of Asian ancestry and 275 from Latin America (Columbia and Mexico).

[0221] A total of 12,338 controls of European ancestry, 1,570 of east Asian ancestry and 445 from Latin America were obtained. These included shared controls from the UK 1958 Birth Cohort, the UK Blood Services Common Controls and from USA; and from participating centres from Spain, France, Netherlands, Mexico, Colombia, China, Taiwan and Korea.

GENOTYPING

[0222] Samples were genotyped using the Immunochip, an Illumina Infinium platform, according to the manufacturer's recommendations. Control samples from the UK were genotyped at Sanger Centre, Hinxton, Cambridge, UK and at the University of Virginia, USA; control samples from the USA were genotyped at the Feinstein Institute, New York, USA; all other controls and all cases were genotyped at the University of Queensland Diamantina Institute, Brisbane, Queensland, Australia. Bead intensity data was processed and normalized for each sample in Genome Studio; data for successfully genotyped samples was extracted and genotypes were called within collections using optiCall (Shah, T.S. *et al.*, 2012. *Bioinformatics* **28**, 1598-1603). NCBI build 36 (hg18) mapping was used (Illumina manifest file Immuno_BeadChip_11419691_B.bpm).

DATA QUALITY CONTROL

[0223] The present inventors first excluded, for each of the collections separately, SNPs with call rate below 95% or with a Hardy-Weingberg $P<10^{-6}$ in controls, and samples with call rate below 95%. For the overlapping SNPs, pair-wise missingness tests were performed between the collections and removed all SNPs with differential missingness $P<10^{7}$. After merging datasets SNPs with call rate below 98% and samples with call rate below 98%

were removed. A total of 128,935 SNPs were analysed. Cluster plots were visually inspected for all SNPs used to inform conclusions.

[0224] Sample European and Asian ancestry was confirmed by principal component analysis. Immunochip data was merged with seven HapMap 3 populations and samples were identified as European or east Asian based on their projection onto the first 5 principal components of genetic variation. Ethnic outliers were removed by assigning samples to an ethnic group using a model-based unsupervised clustering approach. A second round of principal component analysis was performed on the European and Asian populations independently to better resolve the ethnic differences within the cohorts. In both cases only the first principal component (PC) was correlated to place of origin of the samples and only in the European collection the PC was marginally correlated to cases/control status (P=0.016).

[0225] Duplicated samples (intentional or unintentional) and cryptic relatedness was assessed for the European and the Asian cohorts separately by calculating identity-by-descent (IBD) using PLINK (v1.07) (Purcell, S. *et al.*, 2007. *American journal of human genetics* **81**, 559-575). For each pair of related samples (PI_HAT > 0.2) the sample with the lowest call rate was removed and preferentially selecting cases.

ASSOCIATION ANALYSIS

[0226] Association analysis and population stratification control was performed using linear mixed models as implemented in FaST-LMM v1.05 (Lippert, C. *et al.*, 2011. *Nature Methods* **8**, 833-835). For each chromosome a relationship matrix was constructed with all SNPs excluding those in the chromosome being analysed and in the MHC. Conditional analysis for secondary signal detection was performed by fitting the primary SNP as a fixed-effect as implemented in FaST-LMM and using the same SNP set for the relationship matrix.

LOW FREQUENCY SNP ASSOCIATION ANALYSIS

[0227] A collection of 7,447 cases and 11,479 controls were selected for low frequency SNP association analysis. These samples were of British European origin based on PC analysis and recruitment centre. Association analysis was performed with Fisher's exact test and with logistic regression conditioning on common variant association within the locus.

IMPUTATION

[0228] Genotypes were imputed in candidate regions for the European and Asian cohort using the EUR and ASN reference panels, respectively, from the 1000 Genomes Project (Phase 1, 2010-11 data freeze) (2010, *Nature* **467**, 1061-1073). Genotype data was phased with MaCH and genotypes were imputed with minimac (Li, Y., *et al.*, 2010. *Genetic epidemiology* **34**, 816-834). SNPs with low imputation quality ($r^2 < 0.5$) were excluded. Association analysis in imputed genotypes was assessed with probabilistic genotypes, correcting for population stratification with the first five principal components as covariates, using logistic regression as implemented in mach2dat.

OVERLAP WITH OTHER AUTOIMMUNE DISEASES

[0229] For all genome-wide and suggestive significant loci to AS susceptibility the present inventors searched in a chromosomal window of 0.5 cM around the lead SNP for associations to other autoimmune diseases on the NHGRI GWAS catalogue (as accessed on Jun 05, 2012). LD was computed between the AS associated SNP and the reported SNP using phased data from 1000 Genomes Project. If the AS SNP was not found the next significant SNP was searched in the locus. Positively correlated alleles were then compared to their risk or protective effect on the two diseases to determine if directionality was concordant or discordant. LD was computed with Haploxt (http://genome.sph.umich.edu/wiki/Haploxt).

HLA-B*27 EXPERIMENT WIDE INTERACTIONS

[0230] Testing for interaction between *HLA-B*27* tag SNP rs116488202 and all other non-MHC SNPs was performed in samples of European ancestry by logistic regression by fitting a model with an additive term for each SNP, including a multiplicative interaction term, and five principal components for ancestry correction:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 * x_2 + \sum_{i=1}^{5} \beta_{PC_i} PC_i$, where y is the log odds of disease, β_0 is the intercept, x_1 and x_2 are SNP variables coded so as to reflect underlying additive effects (0, 1 or 2), PC_i codes the projection of samples into the i^{th} principal component, and β terms are regression coefficients estimated from the data.

[0231] The disclosure of every patent, patent application, and publication cited herein is hereby incorporated herein by reference in its entirety.

[0232] The citation of any reference herein should not be construed as an admission that such reference is available as "Prior Art" to the instant application.

[0233] Throughout the specification the aim has been to describe the preferred embodiments of the invention without limiting the invention to any one embodiment or specific collection of features. Those of skill in the art will therefore appreciate that, in light of the instant disclosure, various modifications and changes can be made in the particular embodiments exemplified without departing from the scope of the present invention. All such modifications and changes are intended to be included within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A method of determining the presence, absence or risk of development of a spondyloarthropathy in a subject, the method comprising analyzing a biological sample of the subject, which comprises at least a portion of chromosome 6 comprising a polymorphic site having reference sequence number rs116488202, for a polymorphism in the polymorphic site, which polymorphism is indicative of the presence, absence or risk of development of the spondyloarthropathy.

2. A method according to claim 1, wherein the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site having reference sequence number rs116488202, wherein the presence of T (thymine) at rs116488202, indicates that the subject has the SpA or has an increased risk of developing the SpA.

3. A method according to claim 1, wherein the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site having reference sequence number rs116488202, wherein the presence of C (cytosine) at rs116488202, indicates that the subject lacks a SpA or has a lower risk of developing a SpA.

4. A method according to any one of claims 1 to 3, wherein the SpA is selected from reactive arthritis, ankylosing spondylitis (AS), anterior uveitis, colitic arthritis, psoriasis and enteropathic arthritis.

5. A method according to any one of claims 1 to 4, further comprising diagnosing the presence, absence or risk of development of a specific SpA based on one or more ancillary diagnostic parameters.

6. A method according to claim 5, wherein the or each ancillary diagnostic parameter is selected from clinical features of the disease, blood test results such as the level of the erythrocyte sedimentation rate or C-reactive protein, and results of one or more genetic tests or gene-expression profiling tests.

7. A method according to any one of claims 1 to 6, wherein the SpA is AS.

8. A method according to claim 7, further comprising analyzing the sample for the presence of a polymorphism in at least one AS marker.

9. A method according to claim 8, wherein the at least one AS marker is selected from an *ERAP1* gene, an *IL23R* gene, a *TNFR1* gene locus, a *TRADD* gene locus, a 2P15 chromosomal locus, a 21Q22 chromosomal locus, a *HLA-B* gene, an *IL1R1* gene locus, an *IL1R2* gene locus or the chromosome locus 2Q31.3 .

10. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in the *ERAP1* gene.

11. A method according to claim 10, wherein the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic site within the *ARTS1* gene, having a reference sequence number on chromosome 5 selected from the group consisting of rs27044, rs17482078, rs10050860, rs30187 and rs2287987.

12. A method according to claim 11, wherein the presence of: G (guanine) at rs27044; T (thymine) at rs30187; T at rs2287987; C (cytosine) at rs17482078; and/or C at rs10050860 indicates that the subject has AS or has an increased risk of developing AS.

13. A method according to claim 11, wherein the presence of: C at rs27044; C at rs30187; C at rs2287987; T at rs17482078; and/or T at rs10050860 indicates that the subject lacks AS or has a lower risk of developing AS.

14. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in the *IL23R* gene.

15. A method according to claim 14, wherein the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic site within the *IL23R* gene having a reference sequence number on chromosome 1 selected from the group consisting of rs1004819, rs10489629, rs11465804, rs11209026, rs1343151, rs10889677, rs11209032 and rs1495965.

16. A method according to claim 15, wherein the presence of: T at rs11465804; T at rs10489629; G at rs11209026; G at rs1343151; C at rs1495965; A (adenine) at rs1004819; A at rs10889677; and/or A at rs11209032, indicates that the subject has AS or has an increased risk of developing AS.

17. A method according to claim 15, wherein the presence of: C at rs10489629; C at rs10889677; A at rs1343151; A at rs11209026; G at rs11209032; G at rs1004819; G at rs11465804; and/or T at rs1495965 indicates that the subject lacks AS or has a lower risk of developing AS.

18. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in the *TNFR1* gene locus.

19. A method according to claim 18, wherein the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic site within the *TNFR1* gene locus, having reference sequence number rs4149576 on chromosome 12.

20. A method according to claim 19, wherein the presence of C at rs4149576 indicates that the subject has AS or has an increased risk of developing AS.

21. A method according to claim 19, wherein the presence of T at rs4149576 indicates that the subject lacks AS or has a lower risk of developing AS.

22. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in the *TRADD* gene locus.

23. A method according to claim 22, wherein the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic site within that locus, having reference sequence number rs9033 on chromosome 16.

24. A method according to claim 23, wherein the presence of G at rs9033 indicates that the subject has AS or has an increased risk of developing AS.

25. A method according to claim 23, wherein the presence of A at rs9033 indicates that the subject lacks AS or has a lower risk of developing AS.

26. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in the 2P15 chromosomal locus.

27. A method according to claim 26, wherein the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic site within the 2P15 chromosome locus having a reference sequence number rs10865331 on chromosome 2.

28. A method according to claim 27, wherein the presence of A at rs10865331, indicates that the subject has AS or has an increased risk of developing AS.

29. A method according to claim 27, wherein the presence of G at rs10865331, indicates that the subject lacks AS or has a lower risk of developing AS.

30. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in the 21Q22 chromosomal locus.

31. A method according to claim 30, wherein the analysis comprises determining the identity of a polymorphic nucleotide in at least one polymorphic site within the 21Q22 chromosome locus having reference sequence number rs2242944 on chromosome 21.

32. A method according to claim 31, wherein the presence of G at rs2242944, indicates that the subject has AS or has an increased risk of developing AS.

33. A method according to claim 31, wherein the presence of A at rs2242944, indicates that the subject lacks AS or has a lower risk of developing AS.

34. A method according to claim 9, wherein the sample is further analyzed for the presence of an *HLA-B*27* allele of the *HLA-B* gene.

35. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in the *IL1R1* gene locus.

36. A method according to claim 35, wherein the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site within the *IL1R1* gene locus, having reference sequence number rs949963 on chromosome 2.

37. A method according to claim 36, wherein the presence of C at rs949963, indicates that the subject has AS or has an increased risk of developing AS.

38. A method according to claim 36, wherein the presence of T at rs949963, indicates that the subject lacks AS or has a lower risk of developing AS.

39. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in the *IL1R2* gene locus.

40. A method according to claim 39, wherein the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site within the *IL1R2* gene locus, having reference sequence number rs2310173 on chromosome 2.

41. A method according to claim 40, wherein the presence of T at rs2310173, indicates that the subject has AS or has a risk of developing AS.

42. A method according to claim 40, wherein the presence of G at rs2310173, indicates that the subject lacks AS or has a lower of developing AS.

43. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in the 2Q31.3 chromosome locus.

44. A method according to claim 43, wherein the analysis comprises determining the identity of a polymorphic nucleotide in a polymorphic site within the 2Q31.3 chromosome locus, having reference sequence number rs1018326 on chromosome 2.

45. A method according to claim 44, wherein the presence of C at rs1018326, indicates that the subject has AS or has an increased risk of developing AS.

46. A method according to claim 44, wherein the presence of T at rs1018326, indicates that the subject lacks AS or has a lower risk of developing AS.

47. A method according to claim 9, wherein the sample is further analyzed for the presence of a polymorphism in at least 2 AS markers.

48. A method for treating a SpA in a subject, the method comprising analyzing a biological sample of the subject, which comprises at least a portion of chromosome 6 comprising a polymorphic site having reference sequence number rs116488202 for a polymorphism in the polymorphic site, which polymorphism is indicative of the presence or risk of development of the spondyloarthropathy, and exposing the subject to a treatment that ameliorates or reverses the symptoms of the SpA on the basis that the subject tests positive for the polymorphism.

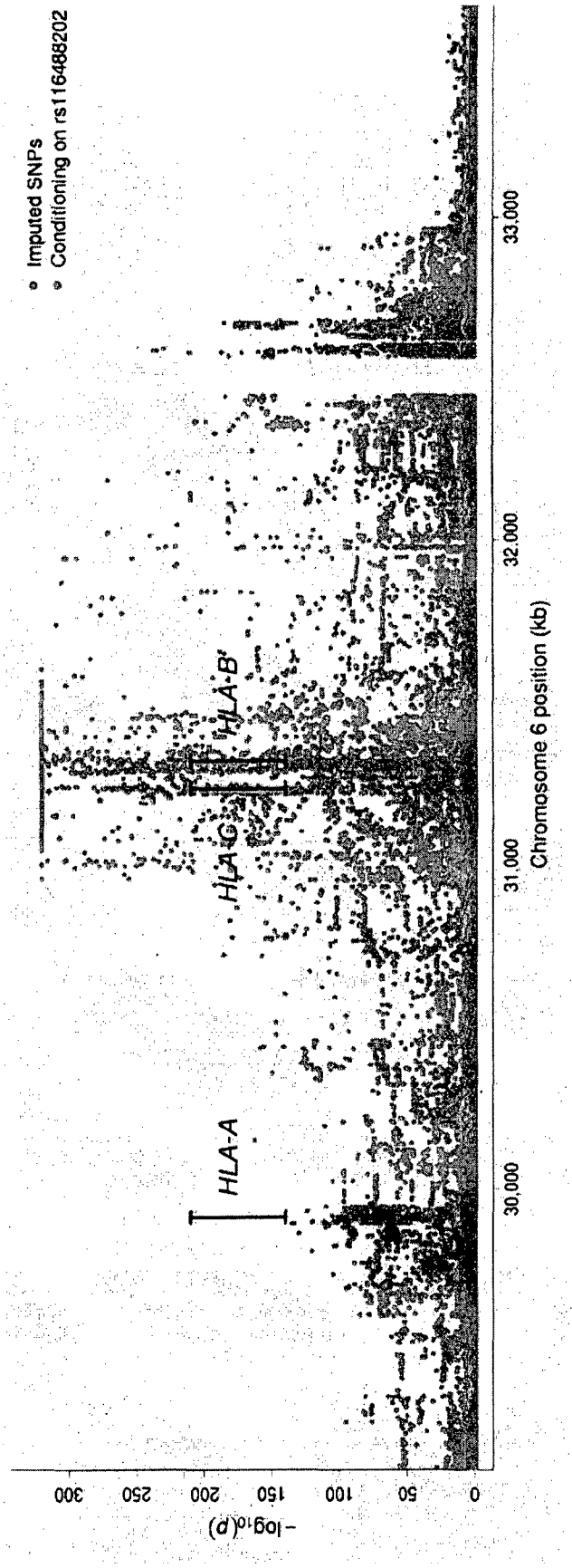

Imputed SNPs
Conditioning on rs116488202
HLA-A
HLA-G
HLA-B
30,000
31,000
32,000
33,000
Chromosome 6 position (kb)
0
50
100
150
200
250
300
$-\log_{10}(p)$

FIGURE 1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2013/000931

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2006.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, CAPLUS, BIOSIS, BIOTECHABS, EPODOC, WPI, TXTE (spondyloarthropathy, spondylarthritis, vertebral column, ankylosing spondylitis, reactive arthritis, reiter, spondylitis, psoriatic arthritis, anterior uveitis, behcet, whipple, bechterew, colitic arthritis, enteropathic arthritis, psoriasis, rs116488202, 201C/T, reference snp, refsnp, 116488202, mutation, polymorphism, substitution, addition, deletion, SNP, variant, chromosome 6; and other synonyms and like terms) GENOMEQUEST (201C/T SNP in SEQ ID NO:1)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

[X] Further documents are listed in the continuation of Box C

[X] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 October 2013	29 October 2013

Name and mailing address of the ISA/AU	**Authorised officer**
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA Email address: pct@ipaustralia.gov.au Facsimile No.: +61 2 6283 7999	Damian Triffett AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No. +61 2 6283 2845

INTERNATIONAL SEARCH REPORT C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		International application No. **PCT/AU2013/000931**
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2011/063474 A1 (THE UNIVERSITY OF QUEENSLAND) 03 June 2011 See whole document especially claims 1, 3 and 44	1-48

INTERNATIONAL SEARCH REPORT Information on patent family members	International application No. **PCT/AU2013/000931**

This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	**Publication Date**	**Publication Number**	**Publication Date**
WO 2011/063474 A1	03 Jun 2011	None	

End of Annex

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.