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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
08 March 2018 (08.03.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/045162 A1

(51) International Patent Classification:
C12Q 1/68 (2018.01)

(21) International Application Number:
PCT/US2017/049599

(22) International Filing Date:
31 August 2017 (31.08.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/382,574 01 September 2016 (01.09.2016) US

(71) Applicant: **BIOGEN MA INC.** [US/US]; 225 Binney St, Cambridge, MA 02142 (US).

(72) Inventors: **BRONSON, Paola**; c/o Biogen MA Inc., 225 Binney Street, Cambridge, MA 02142 (US). **NGUYEN, Kd**; c/o Biogen MA Inc., 225 Binney Street, Cambridge, MA 02142 (US). **DAY-WILLIAMS, Aaron**; c/o Biogen MA Inc., 225 Binney Street, Cambridge, MA 02142 (US). **GUTWIN, Karl**; c/o Biogen MA Inc., 225 Binney Street, Cambridge, MA 02142 (US). **ESTRADA, Karol**; c/o Biogen MA Inc., 225 Binney Street, Cambridge, MA 02142 (US).

(74) Agent: **SCHAFHEIMER, Steven, Nathaniel** et al.; Lando & Anastasi, LLP, Riverfront Office Park, One Main Street, Suite 1100, Cambridge, MA 02142 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: BIOMARKERS PREDICTIVE OF PRIMARY PROGRESSIVE MULTIPLE SCLEROSIS AND USES THEREOF

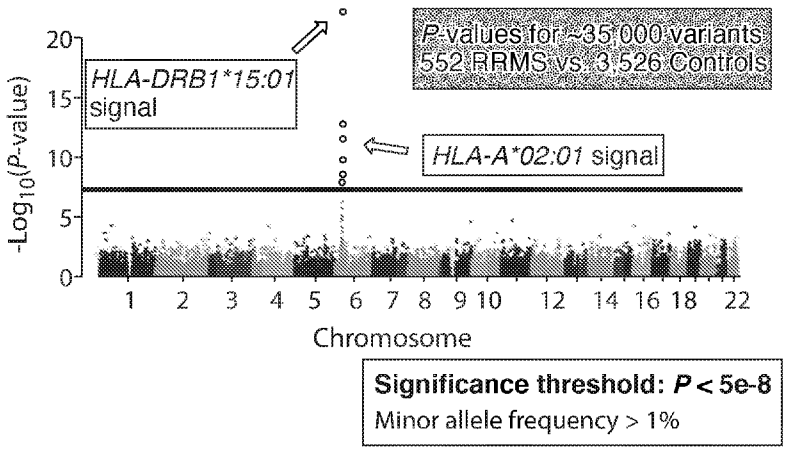

FIG. 2

(57) Abstract: Methods, assays and kits for the identification, assessment and/or treatment of a subject having multiple sclerosis (MS) (e.g., a patient with primary progressive multiple sclerosis (PPMS)) are disclosed.

WO 2018/045162 A1

BIOMARKERS PREDICTIVE OF PRIMARY PROGRESSIVE MULTIPLE SCLEROSIS AND USES THEREOF

RELATED APPLICATIONS

This application claims priority to U.S. Serial No.: 62/382,574, filed September 1, 2016, the entire contents of which is incorporated herein by reference.

BACKGROUND

Multiple sclerosis (MS) is an inflammatory disease of the brain and spinal cord characterized by recurrent foci of inflammation that lead to destruction of the myelin sheath. In many areas, nerve fibers are also damaged.

There are four types of MS. Relapsing-Remitting MS (RRMS) is the most common form of MS. About 85% of people with MS are initially diagnosed with RRMS. People with RRMS have temporary periods called relapses, flare-ups or exacerbations, when new symptoms appear. In Secondary-Progressive MS (SPMS) symptoms worsen more steadily over time, with or without the occurrence of relapses and remissions. Most people who are diagnosed with RRMS will transition to SPMS at some point. Primary-Progressive MS (PPMS) is less common, occurring in about 10% of people with MS. PPMS is characterized by slowly worsening symptoms from the beginning, with no relapses or remissions. Progressive-Relapsing MS (PRMS) is a rare form of MS, occurring in about 5% of MS patients. PRMS is characterized by a steadily worsening disease state from the beginning, with acute relapses but no remissions, with or without recovery.

MS research has primarily focused on RRMS and SPMS. For example, genetic markers associated with RRMS have been identified, including the HLA Class II allele *HLA-DRB1*15:01*. Accordingly, there is a need for the identification of genetic markers identifying susceptibility to other forms of MS, e.g., PPMS, to allow for early diagnosis and treatment.

SUMMARY

The present invention provides, at least in part, methods, assays and kits for the identification, assessment and/or treatment of a subject having multiple sclerosis (MS) (*e.g.*, a subject with Primary Progressive MS (PPMS)). In some embodiments, disclosed herein are genetic biomarkers that can be used to evaluate a subject having MS, e.g., Primary Progressive MS, or at risk of having MS. In certain embodiments, the genetic biomarker is a LIX1L gene variant, e.g., one or more genetic polymorphisms in the LIX1L gene. In other embodiments, the genetic biomarker is the HLA allele HLA-A*02:01. The gene signature can be acquired from a whole blood sample, e.g., peripheral blood, from the subject. Thus, the invention can have prognostic, diagnostic and therapeutic applications, including, for example, (i) to diagnose a subject with MS or at risk of MS; (ii) to stratify a subject, e.g., as having or not having MS (e.g., PPMS), or as being likely or unlikely to develop MS; and (iii) to treat a subject having MS or at risk of having MS.

Accordingly, in one aspect, the invention features a method, or assay, for evaluating, monitoring, stratifying, or treating, a subject. The method includes:

a) obtaining a biological sample from a subject;

b) acquiring a genotype of a nucleic acid sequence in the biological sample to identify the presence of a LIX1L gene variant, e.g., one or more genetic polymorphisms in the LIX1L gene;

c) determining from the genotype an increased likelihood of developing or having Primary Progressive Multiple Sclerosis (PPMS) when the LIX1L gene variant, e.g., one or more genetic polymorphisms in the LIX1L gene, is/are present; and

d) responsive to said determining,

i) classifying said subject,

ii) selecting said subject for treatment with a multiple sclerosis (MS) therapy, or

iii) administering an MS therapy to said subject.

In another aspect, the invention features a method of treating Primary Progressive Multiple Sclerosis (PPMS). The method includes:

a) (optionally) obtaining a biological sample from a subject;

b) acquiring a genotype of a nucleic acid sequence in the biological sample to identify the presence of an LIX1L gene variant, e.g., one or more genetic polymorphisms in the LIX1L gene;

c) determining from the genotype an increased likelihood of developing or having Primary Progressive Multiple Sclerosis (PPMS) when the LIX1L gene variant, e.g., one or more genetic polymorphisms in the LIX1L gene is/are present; and

d) responsive to said determining,

i) administering a MS therapy;

ii) selecting or altering a dosing of a MS therapy; or

iii) selecting or altering the schedule or time course of a MS therapy.

In another aspect, the invention features a method, or assay, for evaluating a subject's likelihood of developing multiple sclerosis (MS). The method includes:

a) (optionally) identifying a subject as being a blood relative of an individual with MS, e.g., a sibling, or a descendant;

b) (optionally) obtaining a biological sample from the subject;

c) acquiring a genotype of a nucleic acid sequence in the biological sample to identify the presence of an LIX1L gene variant, e.g., one or more genetic polymorphisms in the LIX1L gene,

d) determining from the genotype an increased likelihood of developing or having Multiple Sclerosis (MS) when the LIX1L gene variant, e.g., one or more genetic polymorphisms in the LIX1L gene, is/are present.

In some embodiments, wherein the subject's likelihood of developing Primary Progressive Multiple Sclerosis (PPMS) is evaluated. In some embodiments, said individual with MS had Primary Progressive Multiple Sclerosis (PPMS).

In another aspect, the invention features a method, or assay, for evaluating, monitoring, stratifying, or treating, a subject, comprising:

a) (optionally) obtaining a biological sample from a subject;

b) acquiring a genotype of a nucleic acid sequence in the biological sample to identify the presence of a HLA-A*02:01 allele;

c) determining from the genotype a decreased likelihood of developing or having Primary Progressive Multiple Sclerosis (PPMS) when the HLA-A*02:01 allele is present; and

d) responsive to said determining,

i) classifying said subject,

ii) selecting said subject for treatment with a multiple sclerosis (MS) therapy, or

iii) administering an MS therapy to said subject.

In another aspect, the invention features a kit for evaluating a subject's likelihood of developing multiple sclerosis (MS). The kit comprises:

one or more nucleic acid reagents for detecting a Primary Progressive Multiple Sclerosis (PPMS)-associated genetic polymorphism in an LIX1L gene; and

optionally, instructions for use. In embodiments, the nucleic acid reagents comprise: a) a probe complimentary to one or more genetic polymorphisms in the LIX1L gene; or b) allele-specific primers or primer pairs that provides for detection of the one or more genetic polymorphisms in the LIX1L gene.

In yet another aspect, the invention features a nucleic acid encoding an LIX1L gene variant, e.g., a genetic polymorphism in the LIX1L gene chosen from one or more of R114H, V173L, S286C, and L322F, e.g., relative to SEQ ID NO:1, or a combination of two, three, four or all of R114H, V173L, S286C, and L322F.

In yet another aspect, the invention features an LIX1L polypeptide variant, e.g., a polypeptide encoded by a genetic polymorphism in the LIX1L gene. In embodiments, the LIX1L polypeptide variant comprises an amino acid substitution chosen from one or more of R114H, V173L, S286C, and L322F, e.g., relative to SEQ ID NO:1, or a combination of two, three, four or all of R114H, V173L, S286C, and L322F.

Additionally disclosed are vectors and host cells comprising the aforesaid nucleic acid encoding the LIX1L gene variants, as well as detection reagents (e.g., probes, primers and antibodies) that specifically bind to the LIX1L gene and polypeptide variants.

Other features or embodiments of the aforesaid methods, assays, kits, nucleic acids or polypeptides include one or more of the following.

Genetic Biomarkers

In some embodiments, any of the aforesaid method, assays, or kits detect a genetic biomarker, e.g., one or more LIX1L variants, or an HLA-A*02:01 allele. In some embodiments, the LIX1L variant comprises a genetic polymorphism in the LIX1L gene.

In some embodiments, the genetic polymorphism in the LIX1L gene comprises single nucleotide polymorphisms.

In some embodiments, the genetic polymorphism in the LIX1L gene comprises one or more of: a stop gain, a splice donor, a splice acceptor, and/or a missense variants in the LIX1L sequence, or a combination thereof.

In some embodiments, the genetic polymorphism in the LIX1L gene comprises one or more of R114H, V173L, S286C, and L322F, e.g., relative to SEQ ID NO:1, or a combination of two, three, four or all of R114H, V173L, S286C, and L322F. In embodiments, the genetic polymorphism comprises R114H. In some embodiments, the genetic polymorphism comprises V173L. In certain embodiments, the genetic polymorphism comprises S286C. In further embodiments, the genetic polymorphism comprises L322F. In some embodiments, the genetic polymorphism comprises one or both of R114H and V173L. In certain embodiments, the genetic polymorphism comprises one or both of R114H and S286C. In further embodiments, the genetic polymorphism comprises one or both of R114H and L322F. In some embodiments, the genetic polymorphism comprises one or both of V173L and S286C. In certain embodiments, the genetic polymorphism comprises one or both of V173L and L322F. In further embodiments, the genetic polymorphism comprises one or both of S286C and L322F. In some embodiments, the genetic polymorphism comprises one or more of R114H, V173L, and S286C, or a combination of at least two or three thereof. In certain embodiments, the genetic polymorphism comprises one or more of R114H, V173L, and L322F, or a combination of at least two or three thereof. In further embodiments, the genetic polymorphism comprises one or more of R114H, S286C, and L322F, or a combination of at least two or three thereof. In some embodiments, the genetic polymorphism comprises one or more of V173L, S286C, and L322F, or a combination of at least two or three thereof. In certain embodiments, the genetic polymorphism comprises one or more of R114H, V173L, S286C, and L322F, or a combination of at least two, three or all thereof.

Detection Methods

In another aspect, the invention features a method of determining the presence of a polymorphism in the LIX1L gene. The method includes acquiring a genotype, or acquiring

knowledge, e.g., directly acquiring knowledge, that a nucleic acid molecule or polypeptide having the genetic polymorphism in the LIX1L gene (e.g., one or more of the polymorphisms described herein), is present in a sample from a subject.

In some embodiments of any of the aforesaid methods or kits, acquiring a genotype comprises acquiring a sequence (e.g., by sequencing) for a position in the LIX1L gene, thereby determining if the polymorphism in the LIX1L gene is present. In some embodiments, the sequence acquired is compared to a reference sequence, or a wild type reference sequence.

In some embodiments, acquiring a genotype comprises hybridizing one or more probes to the nucleic acid sequence in the biological sample, wherein the probes are complimentary to the genetic polymorphism in the LIX1L gene, and detecting the polymorphism, e.g., in the biological sample. In some embodiments, the genetic polymorphism comprises one or more single nucleotide polymorphisms.

In some embodiments, the genetic polymorphism in the LIX1L gene is detected in a nucleic acid molecule. In some embodiments, the method comprises one or more of: sequencing (e.g., next generation sequencing), nucleic acid hybridization assay, amplification-based assays, PCR-RFLP assay, real-time PCR, screening analysis, FISH, spectral karyotyping or MFISH, comparative genomic hybridization, *in situ* hybridization, SSP, HPLC or mass-spectrometric genotyping.

In some embodiments, a polypeptide having the genetic polymorphism in the LIX1L gene is detected.

In some embodiments, the method comprises contacting a protein sample with a reagent which specifically binds to a polypeptide having the genetic polymorphism in the LIX1L gene; and detecting the formation of a complex of the polypeptide and the reagent. In some embodiments, the reagent is labeled with a detectable group to facilitate

detection of the bound and unbound reagent. In some embodiments, the reagent is an antibody molecule.

Sample

In certain embodiments, the method, or assay, further includes the step of acquiring, e.g., obtaining, a sample, *e.g.*, a biological sample, from the subject. In one embodiment, the method, or assay, includes the step of acquiring, e.g., obtaining, whole blood, e.g., peripheral blood, from the subject. In such embodiments, the sample comprises whole blood. In other embodiments, the sample comprises a predominantly non-cellular fraction of a body fluid from the subject. The non-cellular fraction can be plasma, serum, or other non-cellular body fluid. In other embodiments, the sample is a serum sample. In certain embodiments, the blood can be further processed to obtain plasma or serum. In another embodiment, the sample contains a tissue, cells (*e.g.*, peripheral blood mononuclear cells (PBMC)). For example, the sample can be a fine needle biopsy sample, an archival sample (*e.g.*, an archived sample with a known diagnosis and/or treatment history), a histological section (*e.g.*, a frozen or formalin-fixed section, *e.g.*, after long term storage), among others. A sample can include any material obtained and/or derived from a biological sample, including a nucleic acid (*e.g.*, genomic DNA, cDNA, RNA) purified or processed from the sample. Purification and/or processing of the sample can include one or more of extraction, concentration, sorting, concentration, fixation, addition of reagents and the like. In one embodiment, the sample contains the biomarkers described herein.

In certain embodiments, the acquiring step of the methods described herein include acquiring a genetic biomarker, e.g., genotype, e.g., of a LIX1L gene, from a sample, e.g., a sample of whole or peripheral blood, plasma, serum, or other non-cellular body fluid; or a cellular sample (*e.g.*, a PBMC sample). In certain embodiments, the acquired genotype is compared to a specified parameter (*e.g.*, a reference sequence; a control sample; or a sample obtained from a healthy subject), to thereby diagnose,

evaluate, or identify a patient. In certain embodiments, methods of the invention relate to acquiring a genetic biomarker e.g., genotype, e.g., of a LIX1L gene, from whole blood or peripheral blood, or plasma or serum of the subject, wherein the plasma or serum is obtained from the blood of the subject, for example.

In certain embodiments of the methods or assays, a change in the genotype (*e.g.*, sequence) of the genetic biomarker relative to a reference value (*e.g.*, a relative or absolute reference value compared to a value from a normal sample) is indicative of the presence of MS, e.g., Primary Progressive MS.

Treatment

In certain embodiments, the method of treatment includes an MS therapy, *e.g.*, an MS therapy that comprises one or more of an IFN-β 1 molecule; a polymer of glutamic acid, lysine, alanine and tyrosine, *e.g.*, glatiramer (*e.g.*, Copaxone®); an antibody or fragment thereof against alpha-4 integrin, *e.g.*, natalizumab (*e.g.*, Tysabri®); an anthracenedione molecule, *e.g.*, mitoxantrone (*e.g.*, Novantrone®); a fingolimod, *e.g.*, FTY720 (*e.g.*, Gilenya®); a dimethyl fumarate, *e.g.*, an oral dimethyl fumarate (*e.g.*, Tecfidera®); an antibody to the alpha subunit of the IL-2 receptor of T cells (CD25), *e.g.*, daclizumab; an antibody against CD52, *e.g.*, alemtuzumab (*e.g.*, CAMPATH); an inhibitor of a dihydroorotate dehydrogenase, *e.g.*, leflunomide or an active metabolite thereof, *e.g.*, teriflunomide (*e.g.*, AUBAGIO); an antibody to CD20, *e.g.*, rituximab, or ocrelizumab; an anti-LINGO-1 antibody; or a corticosteroid.

In certain embodiments, the method further includes the use of a symptom management therapies, including one or more of an antidepressant, analgesic, anti-tremor agent, among others.

In some embodiments, the subject has been administered a MS therapy, *e.g.*, prior to, or at the time of, acquiring the value.

For any of the methods or assays disclosed herein, the subject treated, or the subject from which the sample is obtained, is a subject having, or at risk of having MS,

e.g. PPMS, at any stage of treatment. In one embodiment, the subject is an MS patient (*e.g.*, a patient with PPMS) prior to administration of an MS therapy described herein. In one embodiment, the subject is a newly diagnosed PPMS patient. In another embodiment, the subject is an MS patient (*e.g.*, an PPMS patient) after administration of an MS therapy described herein. In other embodiments, the subject is an MS patient after administration of the MS therapy for one, two weeks, one month, two months, three months, four months, six months, one year or more.

In some embodiments, the methods described herein further comprises treating, or preventing in, the subject having MS one or more symptoms associated with MS by administering to a subject an MS therapy, in an amount sufficient to reduce one or more symptoms associated with MS. In some embodiments, said treating or preventing comprises reducing, retarding or preventing, a relapse, or the worsening of the disease, in the MS subject. In some embodiments, the methods further comprise one or more of: performing a neurological examination, evaluating the subject's status on the Expanded Disability Status Scale (EDSS), or detecting the subject's lesion status as assessed using an MRI.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the detailed description, drawings, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee. The accompanying drawings are not intended to be drawn to scale. In the drawings, each identical or nearly identical component that is illustrated in various figures is represented by a like numeral. For purposes of clarity, not every component may be labeled in every drawing. In the drawings:

FIG. 1 is a Manhattan plot of p-values from PPMS cases vs. controls.

FIG. 2 is a Manhattan plot of p-values from RRMS cases vs. controls.

FIG. 3 is a graphical representation of a comparison of the HLA signal in PPMS cases and RRMS cases.

FIG. 4 is a diagram showing the location of rare, probably damaging variants in *LIX1L* that were over-represented in PPMS.

DETAILED DESCRIPTION

The present invention provides, at least in part, methods, assays, and kits for the identification, assessment and/or treatment of a subject having MS (*e.g.*, a subject with PPMS). In some embodiments, disclosed herein are genetic biomarkers that can be used to evaluate a subject having MS, e.g., primary progressive MS, or at risk of having MS. In certain embodiments, the genetic biomarker evaluated comprises a LIX1 variant, e.g., genetic polymorphisms in the LIX1L gene. In some embodiments, the genetic polymorphisms in the LIX1L gene comprise one or more of R114H, V173L, S286C, and L322F. In some embodiments, the biomarker evaluated comprises an HLA allele, e.g., HLA-A*02:01. Said biomarkers can discriminate a subject with PPMS from a healthy subject. The genetic biomarkers can be acquired from a whole blood sample, e.g., peripheral blood, from the subject. Thus, the invention can have prognostic, diagnostic and therapeutic applications.

Definitions

As used herein, each of the following terms has the meaning associated with it in this section.

As used herein, the articles "a" and "an" refer to one or to more than one (*e.g.*, to at least one) of the grammatical object of the article, unless indicated to the contrary or otherwise evident from the context.

The term "or" is used herein to mean, and is used interchangeably with, the term "and/or", unless context clearly indicates otherwise.

"Acquire" or "acquiring" as the terms are used herein, refer to obtaining possession of a physical entity, or a value, *e.g.*, a numerical value, by "directly acquiring" or "indirectly acquiring" the physical entity or value. "Directly acquiring" means performing a process (*e.g.*, performing a synthetic or analytical method) to obtain the physical entity or value. "Indirectly acquiring" refers to receiving the physical entity or value from another party or source (*e.g.*, a third party laboratory that directly acquired the physical entity or value). Directly acquiring a physical entity includes performing a process that includes a physical change in a physical substance, *e.g.*, a starting material. Exemplary changes include making a physical entity from two or more starting materials, shearing or fragmenting a substance, separating or purifying a substance, combining two or more separate entities into a mixture, performing a chemical reaction that includes breaking or forming a covalent or non-covalent bond. Directly acquiring a value includes performing a process that includes a physical change in a sample or another substance, *e.g.*, performing an analytical process which includes a physical change in a substance, *e.g.*, a sample, analyte, or reagent (sometimes referred to herein as "physical analysis"), performing an analytical method, *e.g.*, a method which includes one or more of the following: separating or purifying a substance, *e.g.*, an analyte, or a fragment or other derivative thereof, from another substance; combining an analyte, or fragment or other derivative thereof, with another substance, *e.g.*, a buffer, solvent, or reactant; or changing the structure of an analyte, or a fragment or other derivative thereof, *e.g.*, by breaking or forming a covalent or non-covalent bond, between a first and a

second atom of the analyte; or by changing the structure of a reagent, or a fragment or other derivative thereof, *e.g.*, by breaking or forming a covalent or non-covalent bond, between a first and a second atom of the reagent.

"Acquiring a sequence" as the term is used herein, refers to obtaining possession of a nucleotide sequence or amino acid sequence, by "directly acquiring" or "indirectly acquiring" the sequence. "Directly acquiring a sequence" means performing a process (*e.g.*, performing a synthetic or analytical method) to obtain the sequence, such as performing a sequencing method (*e.g.*, a Next Generation Sequencing (NGS) method). "Indirectly acquiring a sequence" refers to receiving information or knowledge of, or receiving, the sequence from another party or source (*e.g.*, a third party laboratory that directly acquired the sequence). The sequence acquired need not be a full sequence, *e.g.*, sequencing of at least one nucleotide, or obtaining information or knowledge, that identifies a LIX1L polymorphism or HLA-A*02:01 allele disclosed herein as being present in a subject constitutes acquiring a sequence.

Directly acquiring a sequence includes performing a process that includes a physical change in a physical substance, *e.g.*, a starting material, such as a tissue sample, *e.g.*, a biopsy, or an isolated nucleic acid (*e.g.*, DNA or RNA) sample. Exemplary changes include making a physical entity from two or more starting materials, shearing or fragmenting a substance, such as a genomic DNA fragment; separating or purifying a substance (*e.g.*, isolating a nucleic acid sample from a tissue); combining two or more separate entities into a mixture, performing a chemical reaction that includes breaking or forming a covalent or non-covalent bond. Directly acquiring a value includes performing a process that includes a physical change in a sample or another substance as described above.

"Complementary" refers to sequence complementarity between regions of two nucleic acid strands or between two regions of the same nucleic acid strand. It is known that an adenine residue of a first nucleic acid region is capable of forming specific hydrogen bonds ("base pairing") with a residue of a second nucleic acid region which is antiparallel to the first region if the residue is thymine or uracil. Similarly, it is known that a cytosine

residue of a first nucleic acid strand is capable of base pairing with a residue of a second nucleic acid strand which is antiparallel to the first strand if the residue is guanine. A first region of a nucleic acid is complementary to a second region of the same or a different nucleic acid if, when the two regions are arranged in an antiparallel fashion, at least one nucleotide residue of the first region is capable of base pairing with a residue of the second region. In certain embodiments, the first region comprises a first portion and the second region comprises a second portion, whereby, when the first and second portions are arranged in an antiparallel fashion, at least about 50%, at least about 75%, at least about 90%, or at least about 95% of the nucleotide residues of the first portion are capable of base pairing with nucleotide residues in the second portion. In other embodiments, all nucleotide residues of the first portion are capable of base pairing with nucleotide residues in the second portion.

The term "polymorphism" refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. Each divergent sequence is termed an allele, and can be part of a gene or located within an intergenic or non-genic sequence. A diallelic polymorphism has two alleles, and a triallelic polymorphism has three alleles. Diploid organisms can contain two alleles and may be homozygous or heterozygous for allelic forms.

The term "single polynucleotide polymorphism (SNP)" refers to a DNA sequence variation that involves a change in a single nucleotide. They present in humans with a frequency of about once in every 1000 bases and contribute to differences among individuals. The majority of SNPs have no effect. However, some affect the risk for certain diseases.

The term "genotype" refers to a description of the alleles of a gene or genes contained in an individual or a sample. Diploid individuals have a genotype that comprises two different sequences (heterozygous) or one sequence (homozygous) at a polymorphic site. In embodiments, no distinction is made between the genotype of an individual and the genotype of a sample originating from the individual. Although

typically a genotype is determined from samples of diploid cells, a genotype can be determined from a sample of haploid cells, such as a sperm cell.

The term "gene" is used broadly to refer to any segment of nucleic acid associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. For example, "gene" refers to a nucleic acid fragment that expresses mRNA, functional RNA, or specific protein, including regulatory sequences. "Genes" also include non- expressed DNA segments that, for example, form recognition sequences for other proteins. "Genes" can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

In the context of nucleotide sequence, the term "substantially identical" is used herein to refer to a first nucleic acid sequence that contains a sufficient or minimum number of nucleotides that are identical to aligned nucleotides in a second nucleic acid sequence such that the first and second nucleotide sequences encode a polypeptide having common functional activity, or encode a common structural polypeptide domain or a common functional polypeptide activity. For example, nucleotide sequences having at least about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to a reference sequence, *e.g.*, a sequence provided herein.

The term “functional variant” refers to polypeptides that have a substantially identical amino acid sequence to the naturally-occurring sequence, or are encoded by a substantially identical nucleotide sequence, and are capable of having one or more activities of the naturally-occurring sequence.

Calculations of homology or sequence identity between sequences (the terms are used interchangeably herein) are performed as follows.

To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for

comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology").

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

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) *J. Mol. Biol.* 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used unless otherwise specified) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller ((1989) *CABIOS*, 4:11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid (SEQ ID NO: 1) molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

It is understood that the molecules of the present invention may have additional conservative or non-essential amino acid substitutions, which do not have a substantial effect on their functions.

"Sequencing" a nucleic acid molecule requires determining the identity of at least 1 nucleotide in the molecule. In embodiments, the identity of less than all of the nucleotides in a molecule are determined. In other embodiments, the identity of a majority or all of the nucleotides in the molecule is determined.

"Next-generation sequencing or NGS or NG sequencing" as used herein, refers to any sequencing method that determines the nucleotide sequence of either individual nucleic acid molecules (*e.g.*, in single molecule sequencing) or clonally expanded proxies for individual nucleic acid molecules in a highly parallel fashion (*e.g.*, greater than 10^5

molecules are sequenced simultaneously). In one embodiment, the relative abundance of the nucleic acid species in the library can be estimated by counting the relative number of occurrences of their cognate sequences in the data generated by the sequencing experiment. Next generation sequencing methods are known in the art, and are described, *e.g.*, in Metzker, M. (2010) *Nature Biotechnology Reviews* 11:31-46, incorporated herein by reference. Next generation sequencing can detect a variant present in less than 5% of the nucleic acids in a sample.

As used herein, the term "LIX1L gene" refers to a gene encoding the LIX1-like protein. The human amino acid and nucleic acid sequences can be found in a public database, such as GenBank and UniProt. For example, the mRNA sequence can be found at Accession No. NM_153713.2 (provided herein as SEQ ID NO:2) and the protein sequence can be found at Accession No. NP_714924.1 (provided herein as SEQ ID NO:1).

As used herein "LIX1L variant" refers to any alteration in the nucleic acid sequence of LIX1L, including in the promoter sequence and termination sequence. In some embodiments, the LIX1L variant comprises one or more polymorphisms, e.g., one, two, three, four, or more, polymorphisms, in the LIX1 gene. In some embodiments, the one or more polymorphisms comprise single nucleotide polymorphisms. In some embodiments, the polymorphisms result in missense mutations, nonsense mutations, read-through mutations, splice donor and acceptor site mutations, which alter the protein sequence. In some embodiments, a LIX1 variant comprises one or more polymorphisms that encodes one or more amino acid substitutions relative to a wild type sequence selected from R114H, V173L, S286C, and L322F.

A "biomarker," "marker" or "genetic biomarker signature" is a genomic DNA sequence (e.g., an LIX1L genetic polymorphism described herein or the HLA-A*02:01 allele) that are positively or negatively associated with MS, e.g., PPMS. The genomic DNA sequence can be in a biological sample (*e.g.*, a whole blood or peripheral sample) obtained from a subject having MS. In embodiments, as compared to the genomic DNA

sequence of an average or median value for an MS patient population, a healthy control, or a healthy subject population (*e.g.,* a control); such genomic DNA sequences are positively or negatively associated with presence of a disease state, such as MS. For example, a marker of the invention which is associated with having PPMS can be present in the genomic DNA in a biological sample obtained from a subject having, or suspected of having, PPMS as compared to a biological sample obtained from a healthy subject.

Multiple Sclerosis (MS) is "treated," "inhibited" or "reduced," if at least one symptom of the disease is reduced, alleviated, terminated, slowed, or prevented. As used herein, MS is also "treated," "inhibited," or "reduced," if recurrence or relapse of the disease is reduced, slowed, delayed, or prevented. Exemplary clinical symptoms of MS that can be used to aid in determining the disease status in a subject can include optic neuritis, diplopia, nystagmus, ocular dysmetria, internuclear opthalmoplegia, movement and sound phosphenes, afferent pupillary defect, paresis, monoparesis, paraparesis, hemiparesis, quadraparesis, plegia, paraplegia, hemiplegia, tetraplegia, quadraplegia, spasticity, dysarthria, muscle atrophy, spasms, cramps, hypotonia, clonus, myoclonus, myokymia, restless leg syndrome, footdrop, dysfunctional reflexes, paraesthesia, anaesthesia, neuralgia, neuropathic and neurogenic pain, l'hermitte's, proprioceptive dysfunction, trigeminal neuralgia, ataxia, intention tremor, dysmetria, vestibular ataxia, vertigo, speech ataxia, dystonia, dysdiadochokinesia, frequent micturation, bladder spasticity, flaccid bladder, detrusor-sphincter dyssynergia, erectile dysfunction, anorgasmy, frigidity, constipation, fecal urgency, fecal incontinence, depression, cognitive dysfunction, dementia, mood swings, emotional lability, euphoria, bipolar syndrome, anxiety, aphasia, dysphasia, fatigue, Uhthoff's symptom, gastroesophageal reflux, and sleeping disorders.

The terms "therapy" or "treatment" (*e.g.*, MS therapy or MS treatment) are used interchangeably herein.

As used herein, "Primary Progressive Multiple Sclerosis" (PPMS) is characterized by a steady progression of increasing neurological deficits without distinct attacks or remissions.

Cerebral lesions, diffuse spinal cord damage and evidence of axonal loss are evident on the MRI of patients with PPMS.

As used herein, the "Expanded Disability Status Scale" or "EDSS" is intended to have its customary meaning in the medical practice. EDSS is a rating system that is frequently used for classifying and standardizing MS. The accepted scores range from 0 (normal) to 10 (death due to MS). Typically patients having an EDSS score of about 4-6 will have moderate disability (*e.g.,* limited ability to walk), whereas patients having an EDSS score of about 7 or 8 will have severe disability (*e.g.,* will require a wheelchair). More specifically, EDSS scores in the range of 1-3 refer to an MS patient who is fully ambulatory, but has some signs in one or more functional systems; EDSS scores in the range higher than 3 to 4.5 show moderate to relatively severe disability; an EDSS score of 5 to 5.5 refers to a disability impairing or precluding full daily activities; EDSS scores of 6 to 6.5 refer to an MS patient requiring intermittent to constant, or unilateral to bilateral constant assistance (cane, crutch or brace) to walk; EDSS scores of 7 to 7.5 means that the MS patient is unable to walk beyond five meters even with aid, and is essentially restricted to a wheelchair; EDSS scores of 8 to 8.5 refer to patients that are restricted to bed; and EDSS scores of 9 to 10 mean that the MS patient is confined to bed, and progressively is unable to communicate effectively or eat and swallow, until death due to MS.

"Responsiveness," to "respond" to treatment, and other forms of this verb, as used herein, refer to the reaction of a subject to treatment with an MS therapy. As an example, a subject responds to treatment with an MS therapy if at least one symptom of multiple sclerosis in the subject is reduced or retarded by about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more. In another example, a subject responds to treatment with an MS therapy, if at least one symptom of multiple sclerosis in the subject is reduced by about 5%, 10%, 20%, 30%, 40%, 50% or more as determined by any appropriate measure, *e.g.*, Expanded Disability Status Scale (EDSS) or determining the extent of other symptoms such as relapse rate, muscle weakness, tingling, and numbness. In another example, a subject responds to treatment with an MS therapy, if the subject

experiences a life expectancy extended by about 5%, 10%, 20%, 30%, 40%, 50% or more beyond the life expectancy predicted if no treatment is administered. In another example, a subject responds to treatment with an MS therapy, if the subject has an increased disease-free survival, overall survival or increased time to progression. Several methods can be used to determine if a patient responds to a treatment including the EDSS criteria, as set forth above.

The term "probe" or "detection reagent" refers to any molecule which is capable of selectively binding to a specifically intended target molecule, for example a biomarker as described herein. Probes or detection reagents can be either synthesized by one skilled in the art, or derived from appropriate biological preparations. For purposes of detection of the target molecule, probes can be specifically designed to be labeled, as described herein. Examples of molecules that can be utilized as probes or detection reagents include, but are not limited to, RNA, DNA, proteins, antibodies, and organic monomers.

"Sample," "tissue sample," "patient sample," "patient cell or tissue sample" or "specimen" each refers to a biological sample obtained from a tissue or bodily fluid of a subject or patient. The source of the tissue sample can be solid tissue as from a fresh, frozen and/or preserved organ, tissue sample, biopsy, or aspirate; blood (e.g., whole blood or peripheral blood) or any blood constituents (*e.g.,* serum, plasma); bodily fluids such as urine, cerebral spinal fluid, whole blood, plasma and serum. The sample can include a non-cellular fraction (*e.g.*, urine, plasma, serum, or other non-cellular body fluid). In some embodiments, the body fluid from which the sample is obtained from an individual comprises blood (*e.g.*, whole blood). In certain embodiments, the blood can be further processed to obtain plasma or serum. In another embodiment, the sample contains a tissue, cells (*e.g.*, peripheral blood mononuclear cells (PBMC)). For example, the sample can be a fine needle biopsy sample, an archival sample (*e.g.*, an archived sample with a known diagnosis and/or treatment history), a histological section (*e.g.*, a frozen or formalin-fixed section, *e.g.*, after long term storage), among others. The term sample includes any material obtained and/or derived from a biological sample, including

a polypeptide, and nucleic acid (*e.g.*, genomic DNA, cDNA, RNA) purified or processed from the sample. Purification and/or processing of the sample can involve one or more of extraction, concentration, antibody isolation, sorting, concentration, fixation, addition of reagents and the like. The sample can contain compounds that are not naturally intermixed with the tissue in nature such as preservatives, anticoagulants, buffers, fixatives, nutrients, antibiotics or the like.

Various aspects of the invention are described in further detail below. Additional definitions are set out throughout the specification.

LIX1L

Analysis of genomic sequence data in subjects having primary progressive multiple sclerosis (PPMS), relapsing remitting multiple sclerosis (RRMS) and unaffected subjects has led to the identification of variants of LIX1L which are associated with PPMS.

Limb expression 1-like (*LIX1L*) encodes a 36.5 KDa protein and is located on chromosome 1q21.1. LIX1-like protein (the protein encoded by *LIX1L)* is a putative RNA binding protein with an RNA binding motif. It is expressed on cancer cells and is thought to be involved with apoptosis inhibition. (Nakamura et al., 2015 Scientific Reports 5:13474) *LIX1L* expression increases cell proliferation (which may reflect a decrease in apoptosis). (Nakamura et al. supra)

There is some evidence for a role of *LIX1L* in multiple sclerosis. LIX1-like protein interacts with calplain small subunit 2. (Huttlin et al. 2015 Cell 162(2);425-40) Calpains are known to play a role in neuropathology in the rat experimental autoimmune encephalomyelitis (EAE) model, which is commonly accepted as an animal model for multiple sclerosis. (Park et al. 2012 J Neurosci Res. 90(11):2134-45).

The sequence of LIX1L polypeptide is shown below:

METMRAQRLQPGVGTSGRGTLRALRPGVTGAAAATATPPAGPPPAPPPPAPPPPPLLLSG
APGLPLPPGAAGSPAVLREAVEAVVRSFAKHTQGYGRVNVVEALQEFWQMKQSRGADLKN
GALVVYEMVPSNSPPYVCYVTLPGGSCFGSFQFCPTKAEARRSAAKIALMNSVFNEHPSR

RITDEFIEKSVSEALASFNGNREEADNPNTGIGAFRFMLESNKGKSMLEFQELMTVFQLL
HWNGSLKAMRERQCSRQEVLAHYSHRALDDDIRHQMALDWVSREQSVPGALSRELASTER
ELDEARLAGKELRFHKEKKDILVLAAGQLGNMHSSNC (SEQ ID NO: 1)

The sequence of LIX1L mRNA is shown below:

CTGAGTCCCCAGGCGCTAGCTGACTGGGGTGGTTGGGACCGTTAGCTCGGCAGGCTGGCAGGCTCCGGGG
CGCTACTCCATTGGCGGCCGGGATGGAGACTATGCGAGCGCAGCGGCTGCAGCCTGGTGTGGGCACCAGC
GGGAGGGGCACTCTCCGAGCGCTGCGGCCCGGAGTGACTGGGGCCGCGGCCGCCACCGCCACACCCCCTG
CGGGCCCCCCGCCTGCCCCGCCGCCTCCCGCACCGCCCCCGCCGCCGCTGCTCCTGTCTGGGGCCCCAGG
ACTACCCCTGCCCCCCGGCGCCGCCGGCAGCCCGGCAGTGCTGCGAGAGGCCGTGGAGGCCGTGGTGAGG
AGCTTCGCCAAGCACACGCAGGGCTATGGCCGAGTGAATGTGGTGGAGGCACTTCAGGAATTCTGGCAGA
TGAAGCAGTCCCGTGGTGCTGACTTAAAGAATGGGGCTCTAGTGGTTTATGAGATGGTTCCCTCCAACAG
CCCTCCTTATGTCTGCTATGTCACCCTGCCTGGGGGAAGCTGCTTTGGGAGTTTCCAGTTTTGCCCCACA
AAAGCTGAGGCCCGGAGGAGTGCTGCAAAGATTGCGCTAATGAATTCTGTGTTTAATGAACATCCTTCCC
GAAGAATCACTGATGAGTTCATCGAGAAGAGTGTCTCTGAGGCCCTGGCATCTTTTAATGGCAACAGGGA
GGAAGCTGACAACCCAAATACAGGGATTGGTGCCTTCCGATTCATGCTGGAATCCAACAAGGGCAAATCA
ATGTTGGAGTTCCAGGAGCTAATGACAGTTTTTCAACTGCTACACTGGAATGGCAGCCTTAAGGCCATGA
GGGAACGACAATGCTCTCGGCAGGAGGTGTTGGCTCATTATTCGCACCGGGCCCTGGATGATGATATTCG
CCACCAAATGGCCTTGGACTGGGTGAGCCGGGAGCAGAGTGTGCCGGGGGCACTGTCTAGAGAGCTGGCC
TCTACTGAGCGGGAGCTGGATGAAGCCCGACTGGCAGGCAAGGAGCTGCGCTTCCACAAGGAGAAGAAAG
ATATTCTTGTGCTGGCTGCTGGGCAGTTGGGCAATATGCATTCTTCCAACTGCTAGGCATCCACCCACAT
AACTCCCCAGGCTTTCCACAGCCTTTTTTTATGTCTCCTTTCTAAGATTTAGGATGATTTTTTGTACATA
CTTTCTCATTTTTATACTTTAAAAAATATATATGTGTATAAATTCTACACCTAGATTCCTATTTGCTAAG
AGATCCCTTTTCTTACTACCAGTTTTTGGATGTAGTTTTATTTGAAACATCTTCAGTCCACTTTACAACA
AAGAGCAGCTTGTCTTTGCAGCTTTGTTAGCTCTTAAACTTCCAGATTAACTGTGTAGCCATTTCAGTAG
CACTAAAAGATTAACTCTAGTGTTCATGTGTCCTTCTTTTCAAATATCAGGTAACTTGAATAAGGATTAT
GTGCCCCACCCTTACTCTCATTCCTGCTTCCTCTTGGACTCAAACAGGGTATGAGTATGAAGATTTTGCC
TTTAGTTCCTGAACTGAACCTGCTTGCTATCCCTTTCCTCCCCACCACTACCTTATTCCTTCTCTGCCTC
CAAATTGCCACTTTGTTTTGAGGCTTCCTTCCCTACCTTATTATTCTGAAGGAAGTAGAGATCTTGCTTC
TGAAACCCCTCCTAAGAAACTGCCCAGGGACAAGATAAATTACAAACAATTCATGGGAGTTTACTACCTA
AGTTGCTTCTAGGGCATATGTATACCATACTAGTAGTCTAGATTTCTGGATATACTCTACAGTAGATGGG
GGTTATGGTTGAAACTGATTCTCTTTCAGTATTCCCTCTAAACATCTCCCCTACTCCCCCAGCTTAGTTA
AACCCTGCGTTTGGACCTTCCTGCCTGCAGCTATTAGTAGAAAGTAAAACATATTTCCATATTTCCCTTC
ACCTAACATTTTATTTTTTGGAAGCGTTATCAGTCCTATTTGGTTAGTGAGAACCATGTTCCCCTTATTC
CCGCTAAGTTGCTGCTGTTTACTACCTTAGATTCTCATTTGTTTTCTCTTTCTTTCCTCTCCTTCCACAT
TAATTATTAGAACATAAGTTGATCAGGAAAATTAAATGAGACTTTAGTATTTTGGCACTTCCTAATTGAC
ACCTTGGGAGACTGCAGGAAGGGAAAGAGAATCAATGATCAGTTATTTGTGTGTGTGTGTGGTTTTTTTT
TTTTTTTGGAAGACAGAGTCTCACACTTTTGCCCAGGCTGGAGTGCAGTGGCGTGATCTCGGCTCACTGC
AACCTCTGCCTCCAGGGTTCAAGCGATTCTCTTGCCTCAGCTTCCCGAGTAGCTGGGATTACAGGCACAC
GCCACAACGCCTGGCTAATTTTTTATATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGTTCT
TGAACTCCTGACATAAGGTAATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGGCAC
TGAGCCTGGCCTATTTGTGGTTTTTTTTTGGTTTTGTTTTTGTTTTTGTTTTTGAGACGGAGTCTTGCTC
TGTCGCCCAGGCTAGAGTGCAGTGGCGCGATCTCGGCTCACTGCAACCTCCGCCTCCCAGATTCACACCA
TTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGTGCCCACCACCACGCCCAGCTAATTTTTTGT
ATTTTTAGTAGAGACGGGGTTTCACCATGTTAGCCAGGACGGTCTCCATCTCCTGACCTCGTGATCCACC
TGCCTCGGCCTCCCAAACTATTTGTGTGTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGG

CTGGAGTGCAGTGGCGCTATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCACTCTCCTGCCT
CAGCCTCCTTAGTAGCTGGGACTACAGGCGCCCGCCACCATGCCTGGCTAATTTTTTTGTATTTTTAGTA
GAGACGGGGTTTCACCGTGTTTGACAGGATGGTCTCGATTTCCTGACTTCGTGATCCGCCCGCCTCGGCC
TCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCACTATTTGTGTTTTTAACACCATTCTC
CCCCACTTCTCTCCTGGGTGACATAAGAGAGAAATAACCTGTAGTACAGCAGCTAAAGTATTCTCCTTTC
AGAGAATTTTTTTGGAGGTCTCTAATATATATTTCCCCCTTGTCTCTGTGATCTCTTATTTATACTATAT
TATTGTCCCATGTACTTTCTAAACTGAGCTTGGAACATTTAGTATTCCTGCAATTGGACTTCCCACTTAA
CAATTATACAGACTTTGCTTTTAGAAATAGATTAGGTTCCAAACAGAAAGTTCAAGTGTAACAACAACAA
TAAAAATAGATTATGAAACAGGCTATAATTGGCTCTTTTGGATTTGATAGGGGCAAGATGAAAGGCAACT
TTCTTGCTTTTGAAATCATGTTGGGTAAGAGGTAAGGAATCCAGCTACAATTTTATTAGTGCTTGAAACG
GGCTTCCTTGAATTCTCCAGGCCCTATCATTTTTTTTTTCTTACTAATCAGAAGAGAGCTGGGGTAGAAG
CCCCATGTTTGTATTCCATGAAACACGTCGGGTTGGAGTAAAGGCAAAAACAGCTAGACACACCAGGTGT
GTCTGTTTGACATTTATAAGCTGGCACTCATCAACACTCCTGTTTCTCCTTTCTCTGGGACGTGTGGATT
AAGGGGTGTGAGTTGTGGGAAGAATTGCCCTCGTACCTCCTGGATTTATTATTTTTCTCAAATACCAACC
AGTAAGATCCCAAATAACTTGAGAAAAATTGTTTCCTGATCTGTCCACTTCTGGTGTCAAAGATTTTACT
CATCTTCTTAGTACATTCTATGTATTTTATATGTATAATTTTATACAATTAAAAATAGATTTTTGTCTAG
TG (SEQ ID NO: 2)

LIX1L Variants

In some embodiments, the LIX1 variants include LIX1L gene polymorphisms, e.g., single nucleotide polymorphisms. Approximately 90% of all polymorphisms in the human genome are SNPs. SNPs are single base positions in DNA at which different alleles, or alternative nucleotides, exist in a population. The SNP position (interchangeably referred to herein as "SNP," "SNP site," "SNP locus," or "SNP marker") is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations). An individual may be homozygous or heterozygous for an allele at each SNP position.

SNPs can be in the gene coding region (i.e., exons) or in introns, in upstream and downstream sequence. Missense mutations, nonsense mutations, and read-through mutations are types of SNPs in the exons. In some embodiments, the LIX1L variant comprises a missense mutation. A missense mutation changes a codon coding for one amino acid to a codon coding for a different amino acid. In some embodiments, the LIX1L variant comprises a nonsense mutation. A nonsense mutation results in a type of codon change in which a stop codon is formed, thereby leading to premature termination of a polypeptide chain and a truncated protein. In some embodiments, the LIX1L variant

comprises a read-through mutation. A read-through mutation is another type of codon change that causes the destruction of a stop codon, thereby resulting in an extended polypeptide product. Other types of SNPs can include splice donor and acceptor site mutations, mutations in promoter sequences, mutations in termination signals and mutations in ribosome binding sites. In some embodiments, the LIX1L variant comprises a mutation in a splice donor or acceptor site. Splice donor or acceptor site mutations alter splicing resulting in a mis-spliced protein. In some embodiments, the LIX1L variant comprises a SNP in a promoter sequence. In some embodiments, the LIX1L variant comprises a SNP in a termination signal. In some embodiments, the LIX1L variant comprises a SNP in a ribosome binding site.

In one embodiment, the LIX1L variant comprises a single polymorphism. In another embodiment, the LIX1L variant comprises two polymorphisms. In another embodiment, the LIX1L variant comprises three or more polymorphisms. In another embodiment, the LIX1L variant comprises four or more polymorphisms.

In some embodiments, the LIX1L variant results in an altered level of LIX1L. In some embodiments, LIX1L transcription levels are downregulated 1, 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100% relative to a wild-type sequence. In some embodiments, LIX1L transcription levels are upregulated 1, 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100% relative to a wild-type sequence. In some embodiments, LIX1L translation levels are downregulated 1, 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100% relative to a wild-type sequence. In some embodiments, LIX1L translation levels are upregulated 1, 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100% relative to a wild-type sequence.

HLA locus

The human leukocyte antigen (HLA) system is a gene complex encoding the major histocompatibility complex (MHC) proteins in humans. It encodes cell surface molecules specialized to present antigenic peptides to the T-cell receptor (TCR) on T

cells. The HLA gene complex is located within a 3 Mbp region within chromosome 6p21. The HLA genes are highly polymorphic.

The HLA-A gene encodes class I MHC molecules. Class I MHC molecules are transmembrane glycoproteins on the surface of all nucleated cells that present peptides from inside the cell. The HLA-A*02:01 allele has been shown to be associated with a decreased risk of multiple sclerosis in cohorts of patients having multiple sclerosis compared to healthy controls (Brynedal et al., 2007 PLOS One 2(7): e664). These studies do not evaluate the role of different forms of MS.

The sequence of the HLA-A*02:01 allele is shown below:

MAVMAPRTLVLLLSGALALTQTWAGSHSMRYFFTSVSRPGRGEPRFIAVGYVD
DTQFVRFDSDAASQRMEPRAPWIEQEGPEYWDGETRKVKAHSQTHRVDLGTLR
GYYNQSEAGSHTVQRMYGCDVGSDWRFLRGYHQYAYDGKDYIALKEDLRSWT
AADMAAQTTKHKWEAAHVAEQLRAYLEGTCVEWLRRYLENGKETLQRTDAPK
THMTHHAVSDHEATLRCWALSFYPAEITLTWQRDGEDQTQDTELVETRPAGDG
TFQKWAAVVVPSGQEQRYTCHVQHEGLPKPLTLRWEPSSQPTIPIVGIIAGLVLF
GAVITGAVVAAVMWRRKSSDRKGGSYSQAASSDSAQGSDVSLTACKV
(SEQ ID NO: 3)

In some embodiments, the HLA-A*02:01 allele is more highly associated with RRMS than PPMS. In some embodiments, the odds ratio for HLA-A*02:01 in RRMS is 0.60, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, or 0.69. In some embodiments, the odds ratio for HLA-A*02:01 in PPMS is 0.75, 0.76, 0.77, 0.78, 0.79, 0.80, 0.81, 0.82, 0.83, 0.84, or 0.85.

Methods for Detection or Determining LIX1L variant or HLA-A*02:01 allele

Methods for evaluating polymorphisms and alleles at genomic loci are known to those of skill in the art. In one embodiment, the LIX1L variant or HLA-A*02:01 allele is detected in a nucleic acid molecule by a method chosen from one or more of: nucleic acid hybridization assay, amplification-based assays (*e.g.*, polymerase chain reaction (PCR)), PCR-RFLP assay, real-time PCR, sequencing, screening analysis (including metaphase

cytogenetic analysis by standard karyotype methods, FISH (*e.g.*, break away FISH), spectral karyotyping or MFISH, comparative genomic hybridization), *in situ* hybridization, SSP, HPLC or mass-spectrometric genotyping.

Additional exemplary methods include, traditional "direct probe" methods such as Southern blots or *in situ* hybridization (*e.g.,* fluorescence in situ hybridization (FISH) and FISH plus SKY), and "comparative probe" methods such as comparative genomic hybridization (CGH), *e.g.,* cDNA-based or oligonucleotide-based CGH, can be used. The methods can be used in a wide variety of formats including, but not limited to, substrate (e.g., membrane or glass) bound methods or array-based approaches.

In certain embodiments, the evaluation methods include the probes/primers described herein.

Nucleic Acid Molecules and Probes

One aspect of the invention pertains to isolated nucleic acid molecules that correspond to one or markers of the invention. The nucleic acid molecules of the invention include those nucleic acid molecules which reside in genomic regions identified herein. Isolated nucleic acid molecules of the invention also include nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules that correspond to a marker of the invention and fragments of such nucleic acid molecules, e.g., those suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded; in certain embodiments the nucleic acid molecule is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. In certain embodiments, an "isolated" nucleic acid molecule is free of sequences (such as

protein-encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, less than about 4 kB, less than about 3 kB, less than about 2 kB, less than about 1 kB, less than about 0.5 kB or less than about 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, the nucleic acids are isolated from a *e.g.,* blood sample or peripheral blood mononuclear cells (PBMCs).

The language "substantially free of other cellular material or culture medium" includes preparations of nucleic acid molecule in which the molecule is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, nucleic acid molecule that is substantially free of cellular material includes preparations of nucleic acid molecule having less than about 30%, less than about 20%, less than about 10%, or less than about 5% (by dry weight) of other cellular material or culture medium.

If so desired, a nucleic acid molecule of the present invention, can be isolated using standard molecular biology techniques and the sequence information in the database records described herein. Using all or a portion of such nucleic acid sequences, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook et al., ed., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA, or genomic DNA as a template and appropriate oligonucleotide primers according to

standard PCR amplification techniques. The nucleic acid molecules so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which has a nucleotide sequence complementary to the nucleotide sequence of a nucleic acid corresponding to a marker of the invention. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence, wherein the full length nucleic acid sequence comprises a marker of the invention. Such nucleic acid molecules can be used, for example, as a probe or primer. The probe/primer typically is used as one or more substantially purified oligonucleotides. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, at least about 15, at least about 25, at least about 50, at least about 75, at least about 100, at least about 125, at least about 150, at least about 175, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 500, at least about 600, at least about 700, at least about 800, at least about 900, at least about 1 kb, at least about 2 kb, at least about 3 kb, at least about 4 kb, at least about 5 kb, at least about 6 kb, at least about 7 kb, at least about 8 kb, at least about 9 kb, at least about 10 kb, at least about 15 kb, at least about 20 kb, at least about 25 kb, at least about 30 kb, at least about 35 kb, at least about 40 kb, at least about 45 kb, at least about 50 kb, at least about 60 kb, at least about 70 kb, at least about 80 kb, at least about 90 kb, at least about 100 kb, at least about 200 kb, at least about 300 kb, at least about 400 kb, at least about 500 kb, at least about 600 kb, at least about 700 kb, at least about 800 kb, at least about 900 kb, at least about 1 mb,

at least about 2 mb, at least about 3 mb, at least about 4 mb, at least about 5 mb, at least about 6 mb, at least about 7 mb, at least about 8 mb, at least about 9 mb, at least about 10 mb or more consecutive nucleotides of a nucleic acid of the invention.

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts (e.g., mRNA) or genomic sequences corresponding to one or more markers of the invention. The probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues have the nucleic acid sequence of the markers described herein, such as by measuring presence or levels of a nucleic acid molecule in a sample of cells from a subject, e.g., detecting genomic DNA or mRNA levels.

In certain embodiments, labels, dyes, or labeled probes and/or primers are used to detect genomic DNA, amplified or unamplified mRNAs. The skilled artisan will recognize which detection methods are appropriate based on the sensitivity of the detection method and the abundance of the target. Depending on the sensitivity of the detection method and the abundance of the target, amplification may or may not be required prior to detection. One skilled in the art will recognize the detection methods where mRNA amplification is preferred.

In one aspect, oligonucleotide probes or primers present in an amplification reaction are suitable for monitoring the amount of amplification product produced as a function of time. In certain aspects, probes having different single stranded versus double stranded character are used to detect the nucleic acid. In another aspect, oligonucleotide probes or primers present in an amplification reaction are suitable for detecting the presence or absence of a particular polymorphism, with polymorphism specific probes only allowing for amplification when the polymorphism is present. Probes include, but are not limited to, the 5′-exonuclease assay (e.g., TaqMan™) probes (see U.S. Pat. No. 5,538,848), stem-loop molecular beacons (see, e.g., U.S. Pat. Nos. 6,103,476 and 5,925,517), stemless or linear beacons (see, e.g., WO 9921881, U.S. Pat. Nos. 6,485,901

and 6,649,349), peptide nucleic acid (PNA) Molecular Beacons (see, e.g., U.S. Pat. Nos. 6,355,421 and 6,593,091), linear PNA beacons (see, e.g. U.S. Pat. No. 6,329,144), non-FRET probes (see, e.g., U.S. Pat. No. 6,150,097), Sunrise™/AmplifluorB™ probes (see, e.g., U.S. Pat. No. 6,548,250), stem-loop and duplex Scorpion™ probes (see, e.g., U.S. Pat. No. 6,589,743), bulge loop probes (see, e.g., U.S. Pat. No. 6,590,091), pseudo knot probes (see, e.g., U.S. Pat. No. 6,548,250), cyclicons (see, e.g., U.S. Pat. No. 6,383,752), MGB Eclipse™ probe (Epoch Biosciences), hairpin probes (see, e.g., U.S. Pat. No. 6,596,490), PNA light-up probes, antiprimer quench probes (Li et al., Clin. Chem. 53:624-633 (2006)), self-assembled nanoparticle probes, and ferrocene-modified probes described, for example, in U.S. Pat. No. 6,485,901.

In certain embodiments, one or more of the primers in an amplification reaction can include a label. In yet further embodiments, different probes or primers comprise detectable labels that are distinguishable from one another. In some embodiments a nucleic acid, such as the probe or primer, may be labeled with two or more distinguishable labels.

In some aspects, a label is attached to one or more probes and has one or more of the following properties: (i) provides a detectable signal; (ii) interacts with a second label to modify the detectable signal provided by the second label, e.g., FRET (Fluorescent Resonance Energy Transfer); (iii) stabilizes hybridization, e.g., duplex formation; and (iv) provides a member of a binding complex or affinity set, e.g., affinity, antibody-antigen, ionic complexes, hapten-ligand (e.g., biotin-avidin). In still other aspects, use of labels can be accomplished using any one of a large number of known techniques employing known labels, linkages, linking groups, reagents, reaction conditions, and analysis and purification methods.

Labels include, but are not limited to: light-emitting, light-scattering, and light-absorbing compounds which generate or quench a detectable fluorescent, chemiluminescent, or bioluminescent signal (see, e.g., Kricka, L., Nonisotopic DNA Probe Techniquies, Academic Press, San Diego (1992) and Garman A., Non-Radioactive

Labeling, Academic Press (1997).). Fluorescent reporter dyes useful as labels include, but are not limited to, fluoresceins (see, e.g. U.S. Pat. Nos. 5,188,934, 6,008,379, and 6,020,481), rhodamines (see, e.g., U.S. Pat. Nos. 5,366,860, 5,847,162, 5,936,087, 6,051,719, and 6,191,278), benzophenoxazines (see, e.g., U.S. Pat. No. 6,140,500), energy-transfer fluorescent dyes, comprising pairs of donors and acceptors (see, e.g., U.S. Pat. Nos. 5,863,727; 5,800,996; and 5,945,526), and cyanines (see, e.g., WO 9745539), lissamine, phycoerythrin, Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, FluorX (Amersham), Alexa 350, Alexa 430, AMCA, BODIPY 630/650, BODIPY 650/665, BODIPY-FL, BODIPY-R6G, BODIPY-TMR, BODIPY-TRX, Cascade Blue, Cy3, Cy5, 6-FAM, Fluorescein Isothiocyanate, HEX, 6-JOE, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, REG, Rhodamine Green, Rhodamine Red, Renographin, ROX, SYPRO, TAMRA, Tetramethylrhodamine, and/or Texas Red, as well as any other fluorescent moiety capable of generating a detectable signal. Examples of fluorescein dyes include, but are not limited to, 6-carboxyfluorescein; 2′,4′,1,4-tetrachlorofluorescein; and 2′,4′,5′,7′,1,4-hexachlorofluorescein. In certain aspects, the fluorescent label is selected from SYBR-Green, 6-carboxyfluorescein ("FAM"), TET, ROX, VICTM, and JOE. For example, in certain embodiments, labels are different fluorophores capable of emitting light at different, spectrally-resolvable wavelengths (e.g., 4-differently colored fluorophores); certain such labeled probes are known in the art and described above, and in U.S. Pat. No. 6,140,054. A dual labeled fluorescent probe that includes a reporter fluorophore and a quencher fluorophore is used in some embodiments. It will be appreciated that pairs of fluorophores are chosen that have distinct emission spectra so that they can be easily distinguished.

In still a further aspect, labels are hybridization-stabilizing moieties which serve to enhance, stabilize, or influence hybridization of duplexes, e.g., intercalators and intercalating dyes (including, but not limited to, ethidium bromide and SYBR-Green), minor-groove binders, and cross-linking functional groups (see, e.g., Blackburn et al., eds. "DNA and RNA Structure" in Nucleic Acids in Chemistry and Biology (1996)).

The invention further encompasses nucleic acid molecules that are substantially homologous to the genes or gene products described herein such that they are at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5% or greater. In other embodiments, the invention further encompasses nucleic acid molecules that are substantially homologous to the genes or gene products described herein such that they differ by only or at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1 kb, at least 2 kb, at least 3 kb, at least 4 kb, at least 5 kb, at least 6 kb, at least 7 kb, at least 8 kb, at least 9 kb, at least 10 kb, at least 15 kb, at least 20 kb, at least 25 kb, at least 30 kb, at least 35 kb, at least 40 kb, at least 45 kb, at least 50 kb nucleotides or any range in between.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 7, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 550, at least 650, at least 700, at least 800, at least 900, at least 1000, at least 1200, at least 1400, at least 1600, at least 1800, at least 2000, at least 2200, at least 2400, at least 2600, at least 2800, at least 3000, at least 3500, at least 4000, at least 4500, or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule corresponding to a marker of the invention or to a nucleic acid molecule encoding a protein corresponding to a marker of the invention. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which

nucleotide sequences at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, or at least 85% identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *e.g.*, sections 6.3.1-6.3.6 of Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989). Another, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50-65°C.

The methods described herein can also include molecular beacon nucleic acid molecules having at least one region which is complementary to a nucleic acid molecule of the invention, such that the molecular beacon is useful for quantitating the presence of the nucleic acid molecule of the invention in a sample. A "molecular beacon" nucleic acid is a nucleic acid molecule comprising a pair of complementary regions and having a fluorophore and a fluorescent quencher associated therewith. The fluorophore and quencher are associated with different portions of the nucleic acid in such an orientation that when the complementary regions are annealed with one another, fluorescence of the fluorophore is quenched by the quencher. When the complementary regions of the nucleic acid molecules are not annealed with one another, fluorescence of the fluorophore is quenched to a lesser degree. Molecular beacon nucleic acid molecules are described, for example, in U.S. Patent 5,876,930.

DNA Sequencing

The invention also includes methods of sequencing nucleic acids. In one embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence at least a portion of a LIX1L variant or HLA-A*02:01 allele. In one embodiment, the LIX1L variant or HLA-A*02:01 allele sequence is compared to a corresponding reference (control) sequence.

In one embodiment, the sequence of the LIX1L variant or HLA-A*02:01 allele nucleic acid molecule is determined by a method that includes one or more of: hybridizing

an oligonucleotide, *e.g.,* an allele specific oligonucleotide for one alteration described herein to said nucleic acid; hybridizing a primer, or a primer set (*e.g.*, a primer pair), that amplifies a region comprising the LIX1L variant polymorphism(s) or HLA-A*02:01 allele; amplifying, *e.g.*, specifically amplifying, a region comprising the LIX1L variant polymorphism(s) or HLA-A*02:01 allele; attaching an adapter oligonucleotide to one end of a nucleic acid that comprises the LIX1L variant polymorphism(s) or HLA-A*02:01 allele; generating an optical, *e.g.*, a colorimetric signal, specific to the presence of the one of the LIX1L variant polymorphism(s) or HLA-A*02:01 allele; hybridizing a nucleic acid comprising the LIX1L variant polymorphism(s) or HLA-A*02:01 allele to a second nucleic acid, *e.g.*, a second nucleic acid attached to a substrate; generating a signal, *e.g.*, an electrical or fluorescent signal, specific to the presence of the LIX1L variant polymorphism(s) or HLA-A*02:01 allele; and incorporating a nucleotide into an oligonucleotide that is hybridized to a nucleic acid that contains the LIX1L variant polymorphism(s) or HLA-A*02:01 allele.

In another embodiment, the sequence is determined by a method that comprises one or more of: determining the nucleotide sequence from an individual nucleic acid molecule, *e.g.*, where a signal corresponding to the sequence is derived from a single molecule as opposed, *e.g.*, from a sum of signals from a plurality of clonally expanded molecules; determining the nucleotide sequence of clonally expanded proxies for individual nucleic acid molecules; massively parallel short-read sequencing; template-based sequencing; pyrosequencing; real-time sequencing comprising imaging the continuous incorporation of dye-labeling nucleotides during DNA synthesis; nanopore sequencing; sequencing by hybridization; nano-transistor array based sequencing; polony sequencing; scanning tunneling microscopy (STM) based sequencing; or nanowire-molecule sensor based sequencing.

Any method of sequencing known in the art can be used. Exemplary sequencing reactions include those based on techniques developed by Maxam and Gilbert (*Proc. Natl Acad Sci USA* (1977) 74:560) or Sanger (Sanger *et al.* (1977) *Proc. Nat. Acad. Sci*

74:5463). Any of a variety of automated sequencing procedures can be utilized when performing the assays (*Biotechniques* (1995) 19:448), including sequencing by mass spectrometry (see, for example, U.S. Patent Number 5,547,835 and international patent application Publication Number WO 94/16101, entitled *DNA Sequencing by Mass Spectrometry* by H. Köster; U.S. Patent Number 5,547,835 and international patent application Publication Number WO 94/21822 entitled *DNA Sequencing by Mass Spectrometry Via Exonuclease Degradation* by H. Köster), and U.S. Patent Number 5,605,798 and International Patent Application No. PCT/US96/03651 entitled *DNA Diagnostics Based on Mass Spectrometry* by H. Köster; Cohen *et al.* (1996) *Adv Chromatogr* 36:127-162; and Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38:147-159).

Sequencing of nucleic acid molecules can also be carried out using next-generation sequencing (NGS). Next-generation sequencing includes any sequencing method that determines the nucleotide sequence of either individual nucleic acid molecules or clonally expanded proxies for individual nucleic acid molecules in a highly parallel fashion (*e.g.*, greater than 10^5 molecules are sequenced simultaneously). In one embodiment, the relative abundance of the nucleic acid species in the library can be estimated by counting the relative number of occurrences of their cognate sequences in the data generated by the sequencing experiment. Next generation sequencing methods are known in the art, and are described, *e.g.*, in Metzker, M. (2010) *Nature Biotechnology Reviews* 11:31-46, incorporated herein by reference.

In one embodiment, the next-generation sequencing allows for the determination of the nucleotide sequence of an individual nucleic acid molecule (*e.g.*, Helicos BioSciences' HeliScope Gene Sequencing system, and Pacific Biosciences' PacBio RS system). In other embodiments, the sequencing method determines the nucleotide sequence of clonally expanded proxies for individual nucleic acid molecules (*e.g.*, the Solexa sequencer, Illumina Inc., San Diego, Calif; 454 Life Sciences (Branford, Conn.), and Ion Torrent). *e.g.*, massively parallel short-read sequencing (*e.g.*, the Solexa sequencer, Illumina Inc., San Diego, Calif.), which generates more bases of sequence per sequencing unit than other

sequencing methods that generate fewer but longer reads. Other methods or machines for next-generation sequencing include, but are not limited to, the sequencers provided by 454 Life Sciences (Branford, Conn.), Applied Biosystems (Foster City, Calif.; SOLiD sequencer), and Helicos BioSciences Corporation (Cambridge, Mass.).

Platforms for next-generation sequencing include, but are not limited to, Roche/454's Genome Sequencer (GS) FLX System, Illumina/Solexa's Genome Analyzer (GA), Life/APG's Support Oligonucleotide Ligation Detection (SOLiD) system, Polonator's G.007 system, Helicos BioSciences' HeliScope Gene Sequencing system, and Pacific Biosciences' PacBio RS system.

NGS technologies can include one or more of steps, *e.g.*, template preparation, sequencing and imaging, and data analysis.

Next Gen Sequencing

Template preparation

Methods for template preparation can include steps such as randomly breaking nucleic acids (*e.g.*, genomic DNA or cDNA) into smaller sizes and generating sequencing templates (*e.g.*, fragment templates or mate-pair templates). The spatially separated templates can be attached or immobilized to a solid surface or support, allowing massive amounts of sequencing reactions to be performed simultaneously. Types of templates that can be used for NGS reactions include, *e.g.*, clonally amplified templates originating from single DNA molecules, and single DNA molecule templates.

Methods for preparing clonally amplified templates include, *e.g.*, emulsion PCR (emPCR) and solid-phase amplification.

EmPCR can be used to prepare templates for NGS. Typically, a library of nucleic acid fragments is generated, and adapters containing universal priming sites are ligated to the ends of the fragment. The fragments are then denatured into single strands and captured by beads. Each bead captures a single nucleic acid molecule. After amplification and enrichment of emPCR beads, a large amount of templates can be attached or immobilized in

a polyacrylamide gel on a standard microscope slide (*e.g.*, Polonator), chemically crosslinked to an amino-coated glass surface (*e.g.*, Life/APG; Polonator), or deposited into individual PicoTiterPlate (PTP) wells (*e.g.*, Roche/454), in which the NGS reaction can be performed.

Solid-phase amplification can also be used to produce templates for NGS. Typically, forward and reverse primers are covalently attached to a solid support. The surface density of the amplified fragments is defined by the ratio of the primers to the templates on the support. Solid-phase amplification can produce hundreds of millions spatially separated template clusters (*e.g.*, Illumina/Solexa). The ends of the template clusters can be hybridized to universal sequencing primers for NGS reactions.

Other methods for preparing clonally amplified templates also include, *e.g.*, Multiple Displacement Amplification (MDA) (Lasken R. S. *Curr Opin Microbiol.* 2007; 10(5):510-6). MDA is a non-PCR based DNA amplification technique. The reaction involves annealing random hexamer primers to the template and DNA synthesis by high fidelity enzyme, typically Φ29 at a constant temperature. MDA can generate large sized products with lower error frequency.

Template amplification methods such as PCR can be coupled with NGS platforms to target or enrich specific regions of the genome (*e.g.*, exons). Exemplary template enrichment methods include, *e.g.*, microdroplet PCR technology (Tewhey R. *et al., Nature Biotech.* 2009, 27:1025-1031), custom-designed oligonucleotide microarrays (*e.g.*, Roche/NimbleGen oligonucleotide microarrays), and solution-based hybridization methods (*e.g.*, molecular inversion probes (MIPs) (Porreca G. J. *et al., Nature Methods*, 2007, 4:931-936; Krishnakumar S. *et al., Proc. Natl. Acad. Sci. USA*, 2008, 105:9296-9310; Turner E. H. *et al., Nature Methods*, 2009, 6:315-316), and biotinylated RNA capture sequences (Gnirke A. *et al.*, *Nat. Biotechnol.* 2009; 27(2):182-9)

Single-molecule templates are another type of templates that can be used for NGS reaction. Spatially separated single molecule templates can be immobilized on solid supports by various methods. In one approach, individual primer molecules are covalently

attached to the solid support. Adapters are added to the templates and templates are then hybridized to the immobilized primers. In another approach, single-molecule templates are covalently attached to the solid support by priming and extending single-stranded, single-molecule templates from immobilized primers. Universal primers are then hybridized to the templates. In yet another approach, single polymerase molecules are attached to the solid support, to which primed templates are bound.

Sample Preparation

Isolated nucleic acid samples (*e.g.*, genomic DNA samples) can be fragmented or sheared by practicing routine techniques. For example, genomic DNA can be fragmented by physical shearing methods, enzymatic cleavage methods, chemical cleavage methods, and other methods well known to those skilled in the art. In some embodiments, a nucleic acid library is prepared for nucleic acid sequencing. The nucleic acid library can contain all or substantially all of the complexity of the genome. The term "substantially all" in this context refers to the possibility that there can in practice be some unwanted loss of genome complexity during the initial steps of the procedure. The methods described herein also are useful in cases where the nucleic acid library is a portion of the genome, *i.e.*, where the complexity of the genome is reduced by design. In some embodiments, any selected portion of the genome can be used with the methods described herein. In certain embodiments, the entire exome or a subset thereof is isolated.

Methods can further include isolating a nucleic acid sample to provide a library (*e.g.*, a nucleic acid library). In certain embodiments, the nucleic acid sample includes whole genomic, subgenomic fragments, or both. The isolated nucleic acid samples can be used to prepare nucleic acid libraries. Thus, in one embodiment, the methods featured in the invention further include isolating a nucleic acid sample to provide a library (*e.g.*, a nucleic acid library as described herein). Protocols for isolating and preparing libraries from whole genomic or subgenomic fragments are known in the art (*e.g.*, Illumina's genomic DNA sample preparation kit). In certain embodiments, the genomic or subgenomic DNA fragment

is isolated from a subject's sample. In certain embodiments, the nucleic acid sample used to generate the nucleic acid library is less than 5, less than 1 microgram, less than 500 ng, less than 200 ng, less than 100 ng, less than 50 ng or less than 20 ng (*e.g.*, 10 ng or less).

In still other embodiments, the nucleic acid sample used to generate the library includes RNA or cDNA derived from RNA. In some embodiments, the RNA includes total cellular RNA. In other embodiments, certain abundant RNA sequences (*e.g.*, ribosomal RNAs) have been depleted. In some embodiments, the poly(A)-tailed mRNA fraction in the total RNA preparation has been enriched. In some embodiments, the cDNA is produced by random-primed cDNA synthesis methods. In other embodiments, the cDNA synthesis is initiated at the poly(A) tail of mature mRNAs by priming by oligo(dT)-containing oligonucleotides. Methods for depletion, poly(A) enrichment, and cDNA synthesis are well known to those skilled in the art.

The method can further include amplifying the nucleic acid sample (*e.g.*, DNA or RNA sample) by specific or non-specific nucleic acid amplification methods that are well known to those skilled in the art. In certain embodiments, the nucleic acid sample is amplified, *e.g.*, by whole-genome amplification methods such as random-primed strand-displacement amplification.

In other embodiments, the nucleic acid sample is fragmented or sheared by physical or enzymatic methods and ligated to synthetic adapters, size-selected (*e.g.*, by preparative gel electrophoresis) and amplified (*e.g.*, by PCR). In other embodiments, the fragmented and adapter-ligated group of nucleic acids is used without explicit size selection or amplification prior to hybrid selection.

In other embodiments, the isolated DNA (*e.g.*, the genomic DNA) is fragmented or sheared. In some embodiments, the library includes less than 50% of genomic DNA, such as a subfraction of genomic DNA that is a reduced representation or a defined portion of a genome, *e.g.*, that has been subfractionated by other means. In other embodiments, the library includes all or substantially all genomic DNA.

In some embodiments, the library includes less than 50% of genomic DNA, such as a subfraction of genomic DNA that is a reduced representation or a defined portion of a genome, *e.g.*, that has been subfractionated by other means. In other embodiments, the library includes all or substantially all genomic DNA. Protocols for isolating and preparing libraries from whole genomic or subgenomic fragments are known in the art (*e.g.*, Illumina's genomic DNA sample preparation kit). Alternative DNA shearing methods can be more automatable and/or more efficient (*e.g.*, with degraded FFPE samples). Alternatives to DNA shearing methods can also be used to avoid a ligation step during library preparation.

The methods described herein can be performed using a small amount of nucleic acids, *e.g.*, when the amount of source DNA is limiting (*e.g.*, even after whole-genome amplification). In one embodiment, the nucleic acid comprises less than about 5 μg, 4 μg, 3 μg, 2 μg, 1 μg, 0.8 μg, 0.7 μg, 0.6 μg, 0.5 μg, or 400 ng, 300 ng, 200 ng, 100 ng, 50 ng, or 20 ng or less of nucleic acid sample. For example, to prepare 500 ng of hybridization-ready nucleic acids, one typically begins with 3 μg of genomic DNA. One can start with less, however, if one amplifies the genomic DNA (*e.g.*, using PCR) before the step of solution hybridization. Thus it is possible, but not essential, to amplify the genomic DNA before solution hybridization.

In some embodiments, a library is generated using DNA (*e.g.*, genomic DNA) from a sample tissue, and a corresponding library is generated with RNA (or cDNA) isolated from the same sample tissue.

Next-Gen Sequencing and imaging

Exemplary sequencing and imaging methods for NGS include, but are not limited to, cyclic reversible termination (CRT), sequencing by ligation (SBL), single-molecule addition (pyrosequencing), and real-time sequencing.

CRT uses reversible terminators in a cyclic method that minimally includes the steps of nucleotide incorporation, fluorescence imaging, and cleavage. Typically, a DNA polymerase incorporates a single fluorescently modified nucleotide corresponding to the

complementary nucleotide of the template base to the primer. DNA synthesis is terminated after the addition of a single nucleotide and the unincorporated nucleotides are washed away. Imaging is performed to determine the identity of the incorporated labeled nucleotide. Then in the cleavage step, the terminating/inhibiting group and the fluorescent dye are removed. Exemplary NGS platforms using the CRT method include, but are not limited to, Illumina/Solexa Genome Analyzer (GA), which uses the clonally amplified template method coupled with the four-color CRT method detected by total internal reflection fluorescence (TIRF); and Helicos BioSciences/HeliScope, which uses the single-molecule template method coupled with the one-color CRT method detected by TIRF.

SBL uses DNA ligase and either one-base-encoded probes or two-base-encoded probes for sequencing. Typically, a fluorescently labeled probe is hybridized to its complementary sequence adjacent to the primed template. DNA ligase is used to ligate the dye-labeled probe to the primer. Fluorescence imaging is performed to determine the identity of the ligated probe after non-ligated probes are washed away. The fluorescent dye can be removed by using cleavable probes to regenerate a 5'-PO_4 group for subsequent ligation cycles. Alternatively, a new primer can be hybridized to the template after the old primer is removed. Exemplary SBL platforms include, but are not limited to, Life/APG/SOLiD (support oligonucleotide ligation detection), which uses two-base-encoded probes.

Pyrosequencing method is based on detecting the activity of DNA polymerase with another chemiluminescent enzyme. Typically, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step. The template DNA is immobile, and solutions of A, C, G, and T nucleotides are sequentially added and removed from the reaction. Light is produced only when the nucleotide solution complements the first unpaired base of the template. The sequence of solutions which produce chemiluminescent signals allows the determination of the sequence of the template. Exemplary pyrosequencing

platforms include, but are not limited to, Roche/454, which uses DNA templates prepared by emPCR with 1-2 million beads deposited into PTP wells.

Real-time sequencing involves imaging the continuous incorporation of dye-labeled nucleotides during DNA synthesis. Exemplary real-time sequencing platforms include, but are not limited to, Pacific Biosciences platform, which uses DNA polymerase molecules attached to the surface of individual zero-mode waveguide (ZMW) detectors to obtain sequence information when phospholinked nucleotides are being incorporated into the growing primer strand; Life/VisiGen platform, which uses an engineered DNA polymerase with an attached fluorescent dye to generate an enhanced signal after nucleotide incorporation by fluorescence resonance energy transfer (FRET); and LI-COR Biosciences platform, which uses dye-quencher nucleotides in the sequencing reaction.

Other sequencing methods for NGS include, but are not limited to, nanopore sequencing, sequencing by hybridization, nano-transistor array based sequencing, polony sequencing, scanning tunneling microscopy (STM) based sequencing, and nanowire-molecule sensor based sequencing.

Nanopore sequencing involves electrophoresis of nucleic acid molecules in solution through a nano-scale pore which provides a highly confined space within which single-nucleic acid polymers can be analyzed. Exemplary methods of nanopore sequencing are described, *e.g.*, in Branton D. *et al.*, *Nat Biotechnol.* 2008; 26(10):1146-53.

Sequencing by hybridization is a non-enzymatic method that uses a DNA microarray. Typically, a single pool of DNA is fluorescently labeled and hybridized to an array containing known sequences. Hybridization signals from a given spot on the array can identify the DNA sequence. The binding of one strand of DNA to its complementary strand in the DNA double-helix is sensitive to even single-base mismatches when the hybrid region is short or is specialized mismatch detection proteins are present. Exemplary methods of sequencing by hybridization are described, *e.g.*, in Hanna G.J. *et al.*, *J. Clin. Microbiol.* 2000; 38 (7): 2715–21; and Edwards J.R. *et al.*, *Mut. Res.* 2005; 573 (1-2): 3–12.

Polony sequencing is based on polony amplification and sequencing-by-synthesis via multiple single-base-extensions (FISSEQ). Polony amplification is a method to amplify DNA *in situ* on a polyacrylamide film. Exemplary polony sequencing methods are described, *e.g.*, in US Patent Application Publication No. 2007/0087362.

Nano-transistor array based devices, such as Carbon NanoTube Field Effect Transistor (CNTFET), can also be used for NGS. For example, DNA molecules are stretched and driven over nanotubes by micro-fabricated electrodes. DNA molecules sequentially come into contact with the carbon nanotube surface, and the difference in current flow from each base is produced due to charge transfer between the DNA molecule and the nanotubes. DNA is sequenced by recording these differences. Exemplary Nano-transistor array based sequencing methods are described, *e.g.*, in U.S. Patent Application Publication No. 2006/0246497.

Scanning tunneling microscopy (STM) can also be used for NGS. STM uses a piezo-electric-controlled probe that performs a raster scan of a specimen to form images of its surface. STM can be used to image the physical properties of single DNA molecules, *e.g.*, generating coherent electron tunneling imaging and spectroscopy by integrating scanning tunneling microscope with an actuator-driven flexible gap. Exemplary sequencing methods using STM are described, *e.g.*, in U.S. Patent Application Publication No. 2007/0194225.

A molecular-analysis device which is comprised of a nanowire-molecule sensor can also be used for NGS. Such device can detect the interactions of the nitrogenous material disposed on the nanowires and nucleic acid molecules such as DNA. A molecule guide is configured for guiding a molecule near the molecule sensor, allowing an interaction and subsequent detection. Exemplary sequencing methods using nanowire-molecule sensor are described, *e.g.*, in U.S. Patent Application Publication No. 2006/0275779.

Double ended sequencing methods can be used for NGS. Double ended sequencing uses blocked and unblocked primers to sequence both the sense and antisense strands of DNA. Typically, these methods include the steps of annealing an unblocked primer to a first strand of nucleic acid; annealing a second blocked primer to a second strand of nucleic acid;

elongating the nucleic acid along the first strand with a polymerase; terminating the first sequencing primer; deblocking the second primer; and elongating the nucleic acid along the second strand. Exemplary double ended sequencing methods are described, *e.g.*, in U.S. Patent Serial No. 7,244,567.

Next-Gen Sequencing Data analysis

After NGS reads have been generated, they can be aligned to a known reference sequence or assembled *de novo*.

For example, identifying genetic variations such as single-nucleotide polymorphism and structural variants in a sample (*e.g.*, a tumor sample) can be accomplished by aligning NGS reads to a reference sequence (*e.g.*, a wild-type sequence). Methods of sequence alignment for NGS are described *e.g.*, in Trapnell C. and Salzberg S.L. *Nature Biotech.*, 2009, 27:455-457.

Examples of *de novo* assemblies are described, *e.g.*, in Warren R. *et al.*, *Bioinformatics*, 2007, 23:500-501; Butler J. *et al., Genome Res.*, 2008, 18:810-820; and Zerbino D.R. and Birney E., *Genome Res.*, 2008, 18:821-829.

Sequence alignment or assembly can be performed using read data from one or more NGS platforms, *e.g.*, mixing Roche/454 and Illumina/Solexa read data.

Algorithms and methods for data analysis are described in USSN 61/428,568, filed December 30, 2010, incorporated herein by reference.

Amplification based sequence analysis

Amplification may be performed according to various techniques known in the art, such as by polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA) and nucleic acid sequence based amplification (NASBA). These techniques can be performed using commercially available reagents and protocols. Preferred techniques use allele-specific PCR or PCR-SSCP. Amplification usually requires the use of specific nucleic acid primers, to initiate the reaction. Nucleic

acid primers useful for amplifying sequences from a gene locus are able to specifically hybridize with a portion of the gene locus that flank a target region of said locus, said target region being altered in subjects having PPMS. Examples of such target regions are polymorphisms in the LIX1L gene, e.g., nucleic acids encoding the R114H, V173L, S286C, and L322F mutations in the LIX1L protein sequence. Another particular object of this invention resides in a nucleic acid primer useful for amplifying sequences from the gene or locus including surrounding regions. Such primers are preferably complementary to, and hybridize specifically to nucleic acid sequences in the gene locus. Particular primers are able to specifically hybridize with a portion of the gene locus that flank a target region of said locus, said target region being altered in in subjects having PPMS.

The invention also relates to a nucleic acid primer, said primer being complementary to and hybridizing specifically to a portion of a gene locus coding sequence (e.g., gene or RNA) altered in subjects having PPMS. In this regard, particular primers of this invention are specific for altered sequences in the gene. By using such primers, the detection of an amplification product indicates the presence of an alteration in the gene locus. In contrast, the absence of amplification product indicates that the specific alteration is not present in the sample. The invention also concerns the use of a nucleic acid primer or a pair of nucleic acid primers as described above in a method of detecting the presence of or predisposition to Primary Progressive MS in a subject.

Selective hybridization based sequence analysis

Hybridization detection methods are based on the formation of specific hybrids between complementary nucleic acid sequences that serve to detect nucleic acid sequence alteration(s). A particular detection technique involves the use of a nucleic acid probe specific for wild-type or altered gene or RNA, followed by the detection of the presence of a hybrid. The probe may be in suspension or immobilized on a substrate or support (as in nucleic acid array or chips technologies). The probe is typically labeled to facilitate detection of hybrids. In this regard, a particular embodiment of this invention comprises

contacting the sample from the subject with a nucleic acid probe specific for an altered gene locus, and assessing the formation of a hybrid. In a particular preferred embodiment, the method comprises contacting simultaneously the sample with a set of probes that are specific, respectively, for the wild type gene locus and for various altered forms thereof. In this embodiment, it is possible to detect directly the presence of various forms of alterations in the gene locus in the sample. Also, various samples from various subjects may be treated in parallel. Within the context of this invention, a probe refers to a polynucleotide sequence which is complementary to and capable of specific hybridization with a (target portion of a) gene or RNA, and which is suitable for detecting polynucleotide polymorphisms associated with alleles which predispose to or are associated with PPMS.

The method of the invention employs a nucleic acid probe specific for an altered (e.g., a mutated) gene or RNA, i.e., a nucleic acid probe that specifically hybridizes to said altered gene or RNA and essentially does not hybridize to a gene or RNA lacking said alteration. Specificity indicates that hybridization to the target sequence generates a specific signal which can be distinguished from the signal generated through non-specific hybridization. Perfectly complementary sequences are preferred to design probes according to this invention. It should be understood, however, that certain mismatch may be tolerated, as long as the specific signal may be distinguished from non-specific hybridization.

Particular examples of such probes are nucleic acid sequences complementary to a target portion of the genomic region including the LIX1L gene locus carrying one or more of polymorphisms, e.g., nucleic acids encoding the R114H, V173L, S286C, and L322F mutations in the LIX1L protein sequence.

The sequence of the probes can be derived from the sequences of the gene as provided in the present application. Nucleotide substitutions may be performed, as well as chemical modifications of the probe, as described herein. The invention also concerns

the use of a nucleic acid probe as described above in a method of detecting the presence of or predisposition to PPMS or an associated disorder in a subject.

Expression Level Analysis

In certain embodiments, the LIX1 variant or HLA-A*02:01expression level can also be assayed. LIX1 variant or HLA-A*02:01expression can be assessed by any of a wide variety of methods for detecting expression of a transcribed molecule or protein. Non-limiting examples of such methods include immunological methods for detection of secreted, cell-surface, cytoplasmic, or nuclear proteins, protein purification methods, protein function or activity assays, nucleic acid hybridization methods, nucleic acid reverse transcription methods, and nucleic acid amplification methods.

In certain embodiments, activity of a particular gene is characterized by a measure of gene transcript (*e.g.,* mRNA), by a measure of the quantity of translated protein, or by a measure of gene product activity. LIX1 variant or HLA-A*02:01expression can be monitored in a variety of ways, including by detecting mRNA levels, protein levels, or protein activity, any of which can be measured using standard techniques. Detection can involve quantification of the level of gene expression (*e.g.,* genomic DNA, cDNA, mRNA, protein, or enzyme activity), or, alternatively, can be a qualitative assessment of the level of gene expression, in particular in comparison with a control level. The type of level being detected will be clear from the context.

Methods of detecting and/or quantifying the LIX1 variant or HLA-A*02:01 gene transcript (mRNA or cDNA made therefrom) using nucleic acid hybridization techniques are known to those of skill in the art (see Sambrook *et al. supra*). For example, one method for evaluating the presence, absence, or quantity of cDNA involves a Southern transfer as described above. Briefly, the mRNA is isolated (*e.g.,* using an acid guanidinium-phenol-chloroform extraction method, Sambrook *et al. supra.*) and reverse transcribed to produce cDNA. The cDNA is then optionally digested and run on a gel in buffer and transferred to membranes. Hybridization is then carried out using the nucleic

acid probes specific for the LIX1 variant or HLA-A*02:01cDNA, *e.g.*, using the probes and primers described herein.

In other embodiments, LIX1 variant or HLA-A*02:01expression is assessed by preparing genomic DNA or mRNA/cDNA (*i.e.,* a transcribed polynucleotide) from cells in a subject sample, and by hybridizing the genomic DNA or mRNA/cDNA with a reference polynucleotide which is a complement of a polynucleotide comprising the LIX1 variant or HLA-A*02:01, and fragments thereof. cDNA can, optionally, be amplified using any of a variety of polymerase chain reaction methods prior to hybridization with the reference polynucleotide. Expression of LIX1 variant or HLA-A*02:01can likewise be detected using quantitative PCR (QPCR) to assess the level of LIX1 variant or HLA-A*02:01expression.

Polypeptide Detection

The activity or level of a LIX1 variant or HLA-A*02:01polypeptide can also be detected and/or quantified by detecting or quantifying the expressed polypeptide. The LIX1 variant or HLA-A*02:01polypeptide can be detected and quantified by any of a number of means known to those of skill in the art. These can include analytic biochemical methods such as electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, and the like, or various immunological methods such as fluid or gel precipitin reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmunoassay (RIA), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, Western blotting, immunohistochemistry (IHC) and the like. A skilled artisan can adapt known protein/antibody detection methods.

Another agent for detecting a LIX1 variant or HLA-A*02:01polypeptide is an antibody molecule capable of binding to a polypeptide corresponding to a marker of the invention, *e.g.*, an antibody with a detectable label. Techniques for generating antibodies are described herein. The term "labeled", with regard to the probe or antibody, is

intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin.

In another embodiment, the antibody is labeled, *e.g.*, a radio-labeled, chromophore-labeled, fluorophore-labeled, or enzyme-labeled antibody. In another embodiment, an antibody derivative (*e.g.*, an antibody conjugated with a substrate or with the protein or ligand of a protein-ligand pair {*e.g.*, biotin-streptavidin}), or an antibody fragment (*e.g.*, a single-chain antibody, an isolated antibody hypervariable domain, *etc.*) which binds specifically with a LIX1 variant or HLA-A*02:01protein, is used.

LIX1 variant or HLA-A*02:01polypeptides from cells can be isolated using techniques that are known to those of skill in the art. The protein isolation methods employed can, for example, be such as those described in Harlow and Lane (Harlow and Lane, 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

Means of detecting proteins using electrophoretic techniques are well known to those of skill in the art (see generally, R. Scopes (1982) *Protein Purification*, Springer-Verlag, N.Y.; Deutscher, (1990) *Methods in Enzymology* Vol. 182: *Guide to Protein Purification*, Academic Press, Inc., N.Y.).

In another embodiment, Western blot (immunoblot) analysis is used to detect and quantify the presence of a polypeptide in the sample.

In another embodiment, the polypeptide is detected using an immunoassay. As used herein, an immunoassay is an assay that utilizes an antibody to specifically bind to the analyte. The immunoassay is thus characterized by detection of specific binding of a polypeptide to an anti-antibody as opposed to the use of other physical or chemical properties to isolate, target, and quantify the analyte.

The LIX1 variant or HLA-A*02:01polypeptide is detected and/or quantified using any of a number of immunological binding assays (see, *e.g.,* U.S. Patent Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also Asai (1993) *Methods in Cell Biology* Volume 37: *Antibodies in Cell Biology*, Academic Press, Inc. New York; Stites & Terr (1991) *Basic and Clinical Immunology* 7th Edition.

Kits

A kit is any manufacture (*e.g.,* a package or container) comprising at least one reagent, *e.g.,* a probe, primers, or an antibody, for specifically detecting a marker of the invention, the manufacture being promoted, distributed, or sold as a unit for performing the methods of the present invention. When the kits and methods of the invention are used for carrying out the methods of the invention, probes/primers/antibodies corresponding to a biomarker comprising one or more of the LIX1L polymorphisms described herein, e.g., R114H, V173L, S286C, and L322F, can be selected such that a positive result is obtained in at least about 20%, at least about 40%, at least about 60%, at least about 80%, at least about 90%, at least about 95%, at least about 99% or in 100% of subjects afflicted with multiple sclerosis, of the corresponding sub-type, of primary progressive nature.

The invention includes kits and methods for assaying serum in a sample (*e.g.,* a sample obtained from a subject). These kits and methods are substantially the same as those described above, except that, where necessary, the kits and methods are adapted for use with certain types of samples. For example, when the sample is a serum sample, it can be necessary to adjust the ratio of compounds in the kits of the invention, or the methods used. Such methods are well known in the art and within the skill of the ordinary artisan.

The invention thus includes a kit for evaluating MS, e.g., primary progressive MS in a subject (*e.g.,* in a sample such as a serum sample). The kit can comprise one or more

reagents capable of identifying one or more of the genetic polymorphisms in the LIX1L gene described herein, *e.g.,* binding specifically with a nucleic acid corresponding one or more of the genetic polymorphisms in the LIX1L gene described herein. Suitable reagents for binding with a nucleic acid include complementary nucleic acids. For example, the nucleic acid reagents can include oligonucleotides (labeled or non-labeled) fixed to a substrate, labeled oligonucleotides not bound with a substrate, pairs of PCR primers, molecular beacon probes, and the like.

The kit of the invention can optionally comprise additional components useful for performing the methods of the invention. By way of example, the kit can comprise fluids (*e.g.,* SSC buffer) suitable for annealing complementary nucleic acids or for binding an antibody with a protein with which it specifically binds, one or more sample compartments, an instructional material which describes performance of a method of the invention, a reference sample for comparison of amplification levels of the biomarkers described herein, and the like.

A kit of the invention can comprise a reagent useful for determining protein level or protein activity of a marker.

Multiple Sclerosis and Methods of Diagnosis

Multiple sclerosis (MS) is a central nervous system disease that is characterized by inflammation and loss of myelin sheaths.

Patients having MS can be identified by clinical criteria establishing a diagnosis of clinically definite MS as defined by Poser et al., *Ann. Neurol.* 13:227, 1983. Briefly, an individual with clinically definite MS has had two attacks and clinical evidence of either two lesions or clinical evidence of one lesion and paraclinical evidence of another, separate lesion. Definite MS may also be diagnosed by evidence of two attacks and oligoclonal bands of IgG in cerebrospinal fluid or by combination of an attack, clinical evidence of two lesions and oligoclonal band of IgG in cerebrospinal fluid. The McDonald criteria can also be used to diagnose MS. (McDonald et al., 2001,

Recommended diagnostic criteria for Multiple sclerosis: guidelines from the International Panel on the Diagnosis of Multiple Sclerosis, *Ann Neurol* 50:121-127). The McDonald criteria include the use of MRI evidence of CNS impairment over time to be used in diagnosis of MS, in the absence of multiple clinical attacks. Effective treatment of multiple sclerosis may be evaluated in several different ways. The following parameters can be used to gauge effectiveness of treatment. Two exemplary criteria include: EDSS (extended disability status scale), and appearance of exacerbations on MRI (magnetic resonance imaging).

The EDSS is a means to grade clinical impairment due to MS (Kurtzke, *Neurology* 33:1444, 1983). Eight functional systems are evaluated for the type and severity of neurologic impairment. Briefly, prior to treatment, patients are evaluated for impairment in the following systems: pyramidal, cerebella, brainstem, sensory, bowel and bladder, visual, cerebral, and other. Follow-ups are conducted at defined intervals. The scale ranges from 0 (normal) to 10 (death due to MS). A decrease of one full step indicates an effective treatment (Kurtzke, *Ann. Neurol.* 36:573-79, 1994), while an increase of one full step will indicate the progression or worsening of disease (*e.g.,* exacerbation). Typically patients having an EDSS score of about 6 will have moderate disability (*e.g.,* walk with a cane), whereas patients having an EDSS score of about 7 or 8 will have severe disability (*e.g.,* will require a wheelchair).

Exacerbations are defined as the appearance of a new symptom that is attributable to MS and accompanied by an appropriate new neurologic abnormality (IFNB MS Study Group, supra). In addition, the exacerbation must last at least 24 hours and be preceded by stability or improvement for at least 30 days. Briefly, patients are given a standard neurological examination by clinicians. Exacerbations are mild, moderate, or severe according to changes in a Neurological Rating Scale (Sipe et al., *Neurology* 34:1368, 1984). An annual exacerbation rate and proportion of exacerbation-free patients are determined.

Therapy can be deemed to be effective using a clinical measure if there is a statistically significant difference in the rate or proportion of exacerbation-free or relapse-free patients between the treated group and the placebo group for either of these measurements. In addition, time to first exacerbation and exacerbation duration and severity may also be measured. A measure of effectiveness as therapy in this regard is a statistically significant difference in the time to first exacerbation or duration and severity in the treated group compared to control group. An exacerbation-free or relapse-free period of greater than one year, 18 months, or 20 months is particularly noteworthy. Clinical measurements include the relapse rate in one and two-year intervals, and a change in EDSS, including time to progression from baseline of 1.0 unit on the EDSS that persists for six months. On a Kaplan-Meier curve, a delay in sustained progression of disability shows efficacy. Other criteria include a change in area and volume of T2 images on MRI, and the number and volume of lesions determined by gadolinium enhanced images.

MRI can be used to measure active lesions using gadolinium-DTPA-enhanced imaging (McDonald et al., *Ann. Neurol.* 36:14, 1994) or the location and extent of lesions using T2-weighted techniques. Briefly, baseline MRIs are obtained. The same imaging plane and patient position are used for each subsequent study. Positioning and imaging sequences can be chosen to maximize lesion detection and facilitate lesion tracing. The same positioning and imaging sequences can be used on subsequent studies. The presence, location and extent of MS lesions can be determined by radiologists. Areas of lesions can be outlined and summed slice by slice for total lesion area. Three analyses may be done: evidence of new lesions, rate of appearance of active lesions, percentage change in lesion area (Paty et al., *Neurology* 43:665, 1993). Improvement due to therapy can be established by a statistically significant improvement in an individual patient compared to baseline or in a treated group versus a placebo group.

Exemplary symptoms associated with multiple sclerosis, which can be treated with the methods described herein or managed using symptom management therapies,

include: optic neuritis, diplopia, nystagmus, ocular dysmetria, internuclear opthalmoplegia, movement and sound phosphenes, afferent pupillary defect, paresis, monoparesis, paraparesis, hemiparesis, quadraparesis, plegia, paraplegia, hemiplegia, tetraplegia, quadraplegia, spasticity, dysarthria, muscle atrophy, spasms, cramps, hypotonia, clonus, myoclonus, myokymia, restless leg syndrome, footdrop, dysfunctional reflexes, paraesthesia, anaesthesia, neuralgia, neuropathic and neurogenic pain, l'hermitte's, proprioceptive dysfunction, trigeminal neuralgia, ataxia, intention tremor, dysmetria, vestibular ataxia, vertigo, speech ataxia, dystonia, dysdiadochokinesia, frequent micturation, bladder spasticity, flaccid bladder, detrusor-sphincter dyssynergia, erectile dysfunction, anorgasmy, frigidity, constipation, fecal urgency, fecal incontinence, depression, cognitive dysfunction, dementia, mood swings, emotional lability, euphoria, bipolar syndrome, anxiety, aphasia, dysphasia, fatigue, Uhthoff's symptom, gastroesophageal reflux, and sleeping disorders.

In some embodiments, symptoms of MS comprise optic neuritis, diplopia, nystagmus, ocular dysmetria, internuclear opthalmoplegia, movement and sound phosphenes, afferent pupillary defect, paresis, monoparesis, paraparesis, hemiparesis, quadraparesis, plegia, paraplegia, hemiplegia, tetraplegia, quadraplegia, spasticity, dysarthria, muscle atrophy, spasms, cramps, hypotonia, clonus, myoclonus, myokymia, restless leg syndrome, footdrop, dysfunctional reflexes, paraesthesia, anaesthesia, neuralgia, neuropathic and neurogenic pain, l'hermitte's, proprioceptive dysfunction, trigeminal neuralgia, ataxia, intention tremor, dysmetria, vestibular ataxia, vertigo, speech ataxia, dystonia, dysdiadochokinesia, frequent micturation, bladder spasticity, flaccid bladder, detrusor-sphincter dyssynergia, erectile dysfunction, anorgasmy, frigidity, constipation, fecal urgency, fecal incontinence, depression, cognitive dysfunction, dementia, mood swings, emotional lability, euphoria, bipolar syndrome, anxiety, aphasia, dysphasia, fatigue, Uhthoff's symptom, gastroesophageal reflux, and sleeping disorders.

Each case of MS displays one of several patterns of presentation and subsequent course. Most commonly, MS first manifests itself as a series of attacks followed by complete or partial remissions as symptoms mysteriously lessen, only to return later after a period of stability. This is called relapsing-remitting MS (RRMS). Primary-progressive MS (PPMS) is characterized by a gradual clinical decline with no distinct remissions, although there may be temporary plateaus or minor relief from symptoms. Secondary-progressive MS (SPMS) begins with a relapsing-remitting course followed by a later primary-progressive course. Rarely, patients may have a progressive-relapsing (PRMS) course in which the disease takes a progressive path punctuated by acute attacks. PPMS, SPMS, and PRMS are sometimes lumped together and called chronic progressive MS.

A few patients experience malignant MS, defined as a swift and relentless decline resulting in significant disability or even death shortly after disease onset. This decline may be arrested or decelerated by determining the likelihood of the patient to respond to a therapy early in the therapeutic regime and switching the patient to an agent that they have the highest likelihood of responding to.

Therapeutic Methods

"Treat," "treatment," and other forms of this word refer to the administration of an MS therapeutic agent, alone or in combination with one or more symptom management agents, to a subject, *e.g.*, an MS patient, to impede progression of multiple sclerosis, to induce remission, to extend the expected survival time of the subject and or reduce the need for medical interventions (*e.g.*, hospitalizations). In those subjects, treatment can include, but is not limited to, inhibiting or reducing one or more symptoms such as numbness, tingling, muscle weakness; reducing relapse rate, reducing size or number of sclerotic lesions; inhibiting or retarding the development of new lesions; prolonging survival, or prolonging progression-free survival, and/or enhanced quality of life.

As used herein, unless otherwise specified, the terms "prevent," "preventing" and "prevention" contemplate an action that occurs before a subject begins to suffer from the a multiple sclerosis relapse and/or which inhibits or reduces the severity of the disease.

As used herein, and unless otherwise specified, the terms "manage," "managing" and "management" encompass preventing the progression of MS symptoms in a patient who has already suffered from the disease, and/or lengthening the time that a patient who has suffered from MS remains in remission. The terms encompass modulating the threshold, development and/or duration of MS, or changing the way that a patient responds to the disease.

As used herein, and unless otherwise specified, a "therapeutically effective amount" of a compound is an amount sufficient to provide a therapeutic benefit in the treatment or management of multiple sclerosis, or to delay or minimize one or more symptoms associated with MS. A therapeutically effective amount of a compound means an amount of therapeutic agent, alone or in combination with other therapeutic agents, which provides a therapeutic benefit in the treatment or management of MS. The term "therapeutically effective amount" can encompass an amount that improves overall therapy, reduces or avoids symptoms or causes of the disease, or enhances the therapeutic efficacy of another therapeutic agent.

As used herein, and unless otherwise specified, a "prophylactically effective amount" of a compound is an amount sufficient to prevent relapse of MS, or one or more symptoms associated with the disease, or prevent its recurrence. A prophylactically effective amount of a compound means an amount of the compound, alone or in combination with other therapeutic agents, which provides a prophylactic benefit in the prevention of MS relapse. The term "prophylactically effective amount" can encompass an amount that improves overall prophylaxis or enhances the prophylactic efficacy of another prophylactic agent.

As used herein, the term "patient" or "subject" refers to an animal, typically a human (*i.e.,* a male or female of any age group, *e.g.,* a pediatric patient (*e.g.,* infant, child, adolescent) or adult patient (*e.g.,* young adult, middle-aged adult or senior adult) or other mammal, such as a primate (*e.g.,* cynomolgus monkey, rhesus monkey), that will be or has been the object of treatment, observation, and/or experiment. When the term is used

in conjunction with administration of a compound or drug, then the patient has been the object of treatment, observation, and/or administration of the compound or drug.

In one embodiment, the subject has PPMS. In other embodiments, the MS subject has at least one newly developed lesion. In other embodiment, the MS subject has at least one pre-existing lesion. In other embodiments, the subject has a baseline EDSS score of 1.5 to 7.

The methods described herein permit one of skill in the art to identify a monotherapy that an MS patient is most likely to respond to, thus eliminating the need for administration of multiple therapies to the patient to ensure that a therapeutic effect is observed. However, in one embodiment, combination treatment of an individual with MS is contemplated.

It will be appreciated that the MS therapy, as described above and herein, can be administered in combination with one or more additional therapies to treat and/or reduce the symptoms of MS described herein, particularly to treat patients with moderate to severe disability (*e.g.,* EDSS score of 5.5 or higher). The pharmaceutical compositions can be administered concurrently with, prior to, or subsequent to, one or more other additional therapies or therapeutic agents. In general, each agent will be administered at a dose and/or on a time schedule determined for that agent. In will further be appreciated that the additional therapeutic agent utilized in this combination can be administered together in a single composition or administered separately in different compositions. The particular combination to employ in a regimen will take into account compatibility of the pharmaceutical composition with the additional therapeutically active agent and/or the desired therapeutic effect to be achieved. In general, it is expected that additional therapeutic agents utilized in combination be utilized at levels that do not exceed the levels at which they are utilized individually. In some embodiments, the levels utilized in combination will be lower than those utilized individually.

MS Therapeutic Agents, Compositions and Administration

There are several medications presently used to modify the course of multiple sclerosis. Such agents include, but are not limited to, dialkyl fumarates (*e.g.*, DMF or others of Formula A herein), Beta interferons (*e.g.,* Avonex®, Rebif®, Betaseron®, Betaferon®, among others)), glatiramer (Copaxone®), natalizumab (Tysabri®), and mitoxantrone (Novantrone®).

It will be appreciated that the MS therapies, as described above and herein, can be administered in combination with one or more additional therapies to treat and/or reduce the symptoms of MS described herein, particularly to treat patients with moderate to severe disability (*e.g.,* EDSS score of 5.5 or higher). The pharmaceutical compositions can be administered concurrently with, prior to, or subsequent to, one or more other additional therapies or therapeutic agents. In general, each agent will be administered at a dose and/or on a time schedule determined for that agent. It will further be appreciated that the additional therapeutic agent utilized in this combination can be administered together in a single composition or administered separately in different compositions. The particular combination to employ in a regimen will take into account compatibility of the pharmaceutical composition with the additional therapeutically active agent and/or the desired therapeutic effect to be achieved. In general, it is expected that additional therapeutic agents utilized in combination be utilized at levels that do not exceed the levels at which they are utilized individually. In some embodiments, the levels utilized in combination will be lower than those utilized individually.

Immunomodulatory Agents for MS Therapy

Several immunomodulatory agents are presently used to modify the course of multiple sclerosis in patients. Such agents include, but are not limited to, an IFN-β 1 molecule; a polymer of glutamic acid, lysine, alanine and tyrosine, *e.g.*, glatiramer; an antibody or fragment thereof against alpha-4 integrin, *e.g.*, natalizumab; an anthracenedione molecule, *e.g.*, mitoxantrone; a fingolimod, *e.g.*, FTY720; a dimethyl fumarate, *e.g.*, an oral dimethyl fumarate; an antibody to the alpha subunit of the IL-2

receptor of T cells (CD25), *e.g.*, daclizumab; an antibody against CD52, *e.g.*, alemtuzumab; an inhibitor of a dihydroorotate dehydrogenase, *e.g.*, teriflunomide; an antibody to CD20, *e.g.*, ocrelizumab; a corticosteroid; and anti-Lingo-1 antibodies. The reparative agents disclosed herein can be used in combination with any of these agents.

Exemplary immunomodulatory agents are described in more detail as follows.

IFNβ agents (Beta interferons)

One known therapy for MS includes treatment with interferon beta. Interferons (IFNs) are natural proteins produced by the cells of the immune systems of most animals in response to challenges by foreign agents such as viruses, bacteria, parasites and tumor cells. Interferons belong to the large class of glycoproteins known as cytokines. Interferon beta has 165 amino acids. Interferons alpha and beta are produced by many cell types, including T-cells and B-cells, macrophages, fibroblasts, endothelial cells, osteoblasts and others, and stimulate both macrophages and NK cells. Interferon gamma is involved in the regulation of immune and inflammatory responses. It is produced by activated T-cells and Th1 cells.

Several different types of interferon are now approved for use in humans. Interferon alpha (including forms interferon alpha-2a, interferon alpha-2b, and interferon alfacon-1) was approved by the United States Food and Drug Administration (FDA) as a treatment for Hepatitis C. There are two currently FDA-approved types of interferon beta. Interferon beta 1a (Avonex®) is identical to interferon beta found naturally in humans, and interferon beta 1b (Betaseron®) differs in certain ways from interferon beta 1a found naturally in humans, including that it contains a serine residue in place of a cysteine residue at position 17. Other uses of interferon beta have included treatment of AIDS, cutaneous T-cell lymphoma, Acute Hepatitis C (non-A, non-B), Kaposi's sarcoma, malignant melanoma, hairy cell leukemia, and metastatic renal cell carcinoma.

IFNβ agents can be administered to the subject by any method known in the art, including systemically (*e.g.*, orally, parenterally, subcutaneously, intravenously, rectally, intramuscularly, intravitreally, intraperitoneally, intranasally, transdermally, or by inhalation

or intracavitary installation). Typically, the IFNβ agents are administered subcutaneously, or intramuscularly.

IFNβ agents can be used as a monotherapy (*i.e.,* as a single "disease modifying therapy") although the treatment regimen can further comprise the use of "symptom management therapies" such as antidepressants, analgesics, anti-tremor agents, etc. In one embodiment, the IFNβ agent is an IFNβ-1A agent (*e.g.,* Avonex®, Rebif®). In another embodiment, the INFβ agent is an INFβ-1B agent (*e.g.,* Betaseron®, Betaferon®, Extavia®).

Avonex®, an Interferon β-1a, is indicated for the treatment of patients with relapsing forms of MS to slow the accumulation of physical disability and decrease the frequency of clinical exacerbations. Avonex® (Interferon beta-1a) is a 166 amino acid glycoprotein with a predicted molecular weight of approximately 22,500 daltons. It is produced by recombinant DNA technology using genetically engineered Chinese Hamster Ovary cells into which the human interferon beta gene has been introduced. The amino acid sequence of Avonex® is identical to that of natural human interferon beta. The recommended dosage of Avonex® (Interferon beta-1a) is 30 mcg injected intramuscularly once a week. Avonex® is commercially available as a 30 mcg lyophilized powder vial or as a 30 mcg prefilled syringe.

Interferon beta 1a (Avonex®) is identical to interferon beta found naturally in humans (AVONEX®, *i.e.*, Interferon beta Ia (SwissProt Accession No. P01574 and gi:50593016).

Methods for making Avonex® are known in the art.

Other interferon molecules include, e.g., other interferons and fragments, analogues, homologues, derivatives, and natural variants thereof with substantially similar biological activity. The INFβ agent can be modified to increase one or more pharmacokinetic properties. For example, the INFβ agent can be a modified form of interferon 1a to include a pegylated moiety. PEGylated forms of interferon beta 1a are described in, *e.g.*, Baker, D.P. *et al.* (2006) *Bioconjug Chem* 17(1):179-88; Arduini, RM *et al.* (2004) *Protein Expr Purif* 34(2):229-42; Pepinsky, RB *et al.* (2001) *J. Pharmacol. Exp. Ther.* 297(3):1059-66; Baker, D.P. et al. (2010) *J Interferon Cytokine Res* 30(10):777-85 (all of which are incorporated herein by reference in their entirety, and describe a human interferon beta 1a modified at its

N-terminal alpha amino acid to include a PEG moiety, *e.g.*, a 20 kDa mPEG-O-2-methylpropionaldehyde moiety). Pegylated forms of IFN beta 1a can be administered by, *e.g.*, injectable routes of administration (*e.g.*, subcutaneously).

Rebif® is also an Interferon β-1a agent, while Betaseron®, Betaferon®, and Extavia® are Interferon β-1b agents. Both Rebif® and Betaseron® are formulated for administration by subcutaneous injection.

Dosages of IFNβ agents to administer can be determined by one of skill in the art, and include clinically acceptable amounts to administer based on the specific interferon-beta agent used. For example, AVONEX® is typically administered at 30 microgram once a week via intramuscular injection. Other forms of interferon beta 1a, specifically REBIF®, is administered, for example, at 22 microgram three times a week or 44 micrograms once a week, via subcutaneous injection. Interferon beta- 1A can be administered, *e.g.*, intramuscularly, in an amount of between 10 and 50 μg. For example, AVONEX® can be administered every five to ten days, *e.g.*, once a week, while Rebif® can be administered three times a week.

Anti-VLA4 antibody (e.g., Natalizumab (Tysabri®))

Anti-VLA4 antibodies (*e.g.*, Natalizumab) inhibit the migration of leukocytes from the blood to the central nervous system. These antibodies bind to VLA-4 (also called α4βl) on the surface of activated T-cells and other mononuclear leukocytes. They can disrupt adhesion between the T-cell and endothelial cells, and thus prevent migration of mononuclear leukocytes across the endothelium and into the parenchyma. As a result, the levels of pro-inflammatory cytokines can also be reduced. Natalizumab can decrease the number of brain lesions and clinical relapses and accumulation of disability in patients with relapse remitting multiple sclerosis and relapsing secondary-progressive multiple sclerosis.

Natalizumab and related VLA-4 binding antibodies are described, *e.g.*, in U.S. Pat. No. 5,840,299. Monoclonal antibodies 21.6 and HP1/2 are exemplary murine monoclonal antibodies that bind VLA-4. Natalizumab is a humanized version of murine monoclonal

antibody 21.6 (see, *e.g.*, U.S. Pat. No. 5,840,299). A humanized version of HP 1/2 has also been described (see, *e.g.*, U.S. Pat. No. 6,602,503). Several additional VLA-4 binding monoclonal antibodies, such as HP2/1, HP2/4, L25 and P4C2, are described, *e.g.*, in U.S. Pat. No. 6,602,503; Sanchez-Madrid *et al*, (1986) *Eur. J. Immunol* 16:1343-1349; Hemler *et al*, (1987) *J Biol. Chem.* 2:11478-11485; Issekutz *et al.* (1991) *J Immunol* 147: 109 (TA-2 mab); Pulido *et al.* (1991) *J Biol. Chem.* 266: 10241-10245; and U.S. Pat. No. 5,888,507). The contents of the aforesaid publications (including the antibody compositions, dosages, methods of administration and production) are incorporated herein by reference in their entirety.

Dimethyl Fumarate (Tecfidera®)

Dimethyl fumarate, DMF, (Tecfidera®) is a fumaric acid ester. DMF is thought to decrease leukocyte passage through the blood brain barrier and exert neuroprotective effects by the activation of antioxidative pathways, specifically through activation of the Nrf-2 pathway (Lee *et al.* (2008) *Int MS Journal* 15: 12-18). Research also suggests that BG-12® has the potential to reduce the activity and impact of inflammatory cells on the CNS and induce direct cytoprotective responses in CNS cells. These effects may enhance the CNS cells' ability to mitigate the toxic inflammatory and oxidative stress that plays a role in MS pathophysiology.

Glatiramer acetate (Copaxone®)

Copaxone® (glatiramer acetate) consists of the acetate salts of synthetic polypeptides, specifically the four naturally occurring amino acids: L-glutamic acid, L-alanine, L-tyrosine, and L-lysine (Bornstein *et al.* (1987) *N Engl J Med.* 317: 408-414). Copaxone® exhibits structural similarity to myelin basic protein and is thought to function as an immune modulator by shifting the T helper cell type 1 response towards a T helper cell type 2 response (Duda *et al.* (2000) *J Clin Invest* 105: 967-976; Nicholas *et al.* (2011) *Drug Design, Development, and Therapy* 5: 255-274).

Mitoxantrone (Novantrone®, an anthracenedione molecule)

Mitoxantrone is an anthracenedione molecule (1,4-dihydroxy-5,8-bis[2-(2-hydroxyethylamino) ethylamino]-anthracene-9,10-dione) and a type II topoisomerase inhibitor that disrupts DNA synthesis and repair of cells. It is used to treat cancers and MS. Mitoxantrone is used to treat several forms of advancing MS, including secondary progressive MS, progressive relapsing MS, and advanced relapsing-remitting MS. For example, mitoxantrone is effective in slowing the progression of secondary progressive MS and extending the time between relapses in relapsing-remitting MS and progressive relapsing MS (Fox E (2006) *Clin Ther* **28** (4): 461–74).

Fingolimod (Gilenya®; sphingosine 1-phosphate receptor modulator)

Fingolimod is an immunomodulating drug, approved for treating MS. It has reduced the rate of relapses in relapsing-remitting multiple sclerosis by over half, but may have serious adverse effects. Fingolimod is a sphingosine 1-phosphate receptor modulator, which sequesters lymphocytes in lymph nodes, preventing them from moving to the central nervous system for autoimmune responses in MS.

Antibodies to the alpha subunit of the IL-2 receptor of T cells (CD25) (e.g., daclizumab HYP; ZINBRYTA®)

An antibody to the alpha subunit of the IL-2 receptor of T cells (CD25), *e.g.*, daclizumab HYP, can be used in the methods disclosed herein. Daclizumab HYP is a therapeutic humanized monoclonal antibody to the alpha subunit of the IL-2 receptor of T cells (CD25). Daclizumab HYP showed efficacy in reducing lesions and annualized relapse rate in patients with relapsing-remitting multiple sclerosis (Kappos *et al.* (2015). *N. Engl. J. Med.* 373 (15): 1418-28).

Antibody against CD52, e.g., alemtuzumab

Antibodies against CD52, *e.g.*, alemtuzumab (currently under further development as Lemtrada®), bind to CD52, which is a protein present on the surface of mature lymphocytes, but not on stem cells. Phase III studies reported positive results comparing alemtuzumab with Rebif® (high-dose subcutaneous interferon beta-1a) in the treatment of patients with relapsing-remitting MS (RRMS). Alemtuzumab has been approved in Europe.

Antibody to CD20, e.g., ocrelizumab

Antibodies against CD20, *e.g.*, ocrelizumab, rituximab, ofatumumab, target mature B lymphocytes. Phase 2 clinical studies of rituximab and ocrelizumab in relapse remitting MS have demonstrated a statistically significant reduction in disease activity measured by brain lesions (*e.g.*, measured by MRI scans) and relapse rate compared to placebo. Phase 3 studies of ocrelizumab showed both reduction in relapse rate and disability compared to interferon beta-1a (*e.g.*, Rebif®).

Inhibitors of dihydroorotate dehydrogenase, e.g., teriflunomide

Inhibitors of dihydroorotate dehydrogenase, *e.g.*, teriflunomide, inhibit pyrimidine synthesis. Teriflunomide (also known as A77 1726 or) is an active metabolite of leflunomide. Teriflunomide inhibits rapidly dividing cells, including activated T cells, which are thought to drive the disease process in MS. Teriflunomide was investigated in clinical trials as a medication for treating MS. (Vollmer *EMS News* (May 28, 2009)).

Steroids

Steroids, *e.g.*, corticosteroid, and ACTH agents can be used to treat acute relapses in relapsing-remitting MS or secondary progressive MS. Such agents include, but are not limited to, Depo-Medrol®, Solu-Medrol®, Deltasone®, Delta-Cortef®, Medrol®, Decadron®, and Acthar®.

LINGO-1

Anti-Lingo-1 antibodies function by restoring myelin. Anti-Lingo-1 antibodies can be used to treat relapsing MS.

Symptom management

In certain embodiments, the method further includes the use of one or more symptom management therapies, such as antidepressants, analgesics, anti-tremor agents, among others. Treatment of a subject with a disease modifying MS therapy can be combined with one or more of the following therapies often used in symptom management of subjects having MS: IMURAN® (azathioprine), CYTOXAN® (cyclophosphamide), NEOSAR® (cyclophosphamide), SANDIMMUNE® (cyclosporine), methotrexate, LEUSTATIN® (cladribine), TEGRETOL® (carbamazepine), EPITOL® (carbamazepine), ATRETOL® (carbamazepine), CARBATROL® (carbamazepine), NEURONTIN® (gabapentin), TOPAMAX® (topiramate), ZONEGRAN® (zonisamide), DILANTIN® (phenytoin), NORPRAMIN® (desipramine), ELAVIL® (amitriptyline), TOFRANIL® (imipramine), IMAVATE® (imipramine), JANIMINE® (imipramine), SINEQUAN® (doxepine), ADAPIN® (doxepine), TRIADAPIN® (doxepine), ZONALON® (doxepine), VIVACTIL® (protriptyline), MARINOL® (synthetic cannabinoids), TRENTAL® (pentoxifylline), NEUROFEN® (ibuprofen), aspirin, acetaminophen, ATARAX® (hydroxyzine), PROZAC® (fluoxetine), ZOLOFT® (sertraline), LUSTRAL® (sertraline), EFFEXOR XR® (venlafaxine), CELEXA® (citalopram), PAXIL®, SEROXAT®, DESYREL® (trazodone), TRIALODINE® (trazodone), PAMELOR® (nortriptyline), AVENTYL® (imipramine), PROTHIADEN® (dothiepin), GAMANIL® (lofepramine), PARNATE® (tranylcypromine), MANERIX® (moclobemide), AURORIX® (moclobemide), WELLBUTRIN SR® (bupropion), AMFEBUTAMONE® (bupropion), SERZONE® (nefazodone), REMERON® (mirtazapine), AMBIEN® (zolpidem), XANAX® (alprazolam), RESTORIL® (temazepam), VALIUM® (diazepam), BUSPAR® (buspirone), SYMMETREL® (amantadine), CYLERT® (pemoline), PROVIGIL®

(modafinil), DITROPAN XL® (oxybutynin), DDAVP® (desmopressin, vasopressin), DETROL® (tolterodine), URECHOLINE® (bethane), DIBENZYLINE® (phenoxybenzamine), HYTRIN® (terazosin), PRO-BANTHINE® (propantheline), URISPAS® (hyoscyamine), CYSTOPAS® (hyoscyamine), LIORESAL® (baclofen), HIPREX® (methenamine), MANDELAMINE® (metheneamine), MACRODANTIN® (nitrofurantoin), PYRIDIUM® (phenazopyridine), CIPRO® (ciprofloxacin), DULCOLAX® (bisacodyl), BISACOLAX® (bisacodyl), SANI-SUPP® (glycerin), METAMUCIL® (psyllium hydrophilic mucilloid), FLEET ENEMA® (sodium phosphate), COLACE® (docusate), THEREVAC PLUS®, KLONOPIN® (clonazepam), RIVOTRIL® (clonazepam), DANTRIUM® (dantrolen sodium), CATAPRES® (clonidine), BOTOX® (botulinum toxin), NEUROBLOC® (botulinum toxin), ZANAFLEX® (tizanidine), SIRDALUD® (tizanidine), MYSOLINE® (primidone), DIAMOX® (acetozolamide), SINEMET® (levodopa, carbidopa), LANIAZID® (isoniazid), NYDRAZID® (isoniazid), ANTIVERT® (meclizine), BONAMINE® (meclizine), DRAMAMINE® (dimenhydrinate), COMPAZINE® (prochlorperazine), TRANSDERM® (scopolamine), BENADRYL® (diphenhydramine), ANTEGREN® (natalizumab), CAMPATH-1H® (alemtuzumab), FAMPRIDINE® (4-aminopyridine), GAMMAGARD® (IV immunoglobulin), GAMMAR-IV® (IV immunoglobulin), GAMIMUNE N® (IV immunoglobulin), IVEEGAM® (IV immunoglobulin), PANGLOBULIN® (IV immunoglobulin), SANDOGLOBULIN® (IV immunoglobulin), VENOBLOGULIN® (IV immunoglobulin), pregabalin, ziconotide, and ANERGIX-MS®.

EXAMPLES

The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of

illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

Example 1: Identification of Genetic Markers for Multiple Sclerosis (MS)

Introduction

The strongest risk factor for MS is thought to be the human leukocyte antigen (HLA) region in the major histocompatibility complex (MHC) on chromosome 6. The HLA Class II allele *HLA-DRB1*15:01* (OR = 3.92, $P = 2x10^{-686}$) is the strongest HLA risk factor for MS. The HLA Class I allele *HLA-A*02:01* (OR = 0.67, $P = 10x10^{-70}$) is an independent protective allele also associated with MS.

Over 100 common non-MHC variants are also associated with MS (**Table 1**), and they are enriched for T helper cell differentiation. However, these only account for 25% of the heritability (portion of disease explained by genetics) of MS (Moutsianas et al., Nat. Genet. 2015). Furthermore, 90% of the associated common variants from genome-wide association studies (GWAS) are non-coding and the causal gene is often not obvious in these GWAS loci. Therefore, exome sequencing rare variant association studies are advantageous for genetic studies of MS because they identify both coding variants associated with disease, and the causal gene in the locus.

Table 1. Genome-wide association studies and Immunochip study of MS

Year	First Author	Journal	Number of Genome-Wide Significant Regions
2016	Andlauer	Sci Adv	15
2013	Int. MS Genetics Consortium (IMSGC)	Nat Genet	97 (Immunochip)
		Brain	5
2011		Nat	51
2011	Patsopoulos	Ann Neurol	9
2010	Sanna	Nat Genet	2

Primary progressive MS (PPMS) is the next frontier in the treatment of MS. PPMS comprises 10% of MS cases, and no PPMS therapies have yet been approved. The genetics of PPMS have not yet been thoroughly interrogated. Coding variants associated with PPMS, but not RRMS, could help identify PPMS specific targets.

Methods

In this rare variant association study of MS subtypes, 1,589 Swedish samples (509 PPMS cases, 552 relapsing remitting MS (RRMS) cases, and 528 healthy controls) were whole exome sequenced. To increase statistical power, 2,996 Swedish whole exome sequences from a dbGaP study that included healthy controls and schizophrenia cases (53%) (phs000473.v1.p1) (total N = 4,585) were included. Stringent Quality Control (QC) was applied. Only samples without excessive contamination, relatedness and excess of singletons in addition to gender discrepancy and ancestry population outliers were considered. Only variants with high genotype quality, sufficient coverage, within population Hardy-Weinberg equilibrium, with non-significant differential call rate and read depth were considered.

Variants were annotated with the Ensembl Variant Effect Predictor (VEP). EPACTS was used for single variant tests that compared 509 PPMS cases to 3,524 controls (total N = 4,033), 552 RRMS cases to 3,524 controls (total N = 4,076), and 509 PPMS cases to 552 RRMS cases (total N = 1,061). The significance threshold for the single variant tests was $P < 5x10^{-8}$.

Because each rare variant has too few observations to test individually, rare variants on the gene level as a unit were analyzed. Gene-based association tests in R SKAT-O with resampling was conducted. Gene-based tests were restricted to genes with >1 rare (MAF < 0.005 in the Broad Institute's Exome Aggregation Consortium (ExAC) non-Finnish Europeans (NFE) population) stop gain, splice donor, splice acceptor, or missense variant, where missense variants were required to be predicted damaging from computational algorithms (Combined Annotation Dependent Depletion (CADD) C score

> 20 or Polymorphism Phenotyping v2 (PolyPhen-2) "probably damaging"). The significance threshold (Bonferroni correction for 12,750 gene-based tests) was $P < 3.92 \times 10^{-6}$ (**Table 2**).

Table 2. Number of genes and variants tested in gene-based rare variant tests

Study	Number of Genes	Number of Variants	Significance Threshold
PPMS vs. Controls	12,715	89,762	3.9e-6
RRMS vs. Controls	12,930	91,378	
MS vs. Controls	12,715	89,762	

Results

In the single variant analysis, the *HLA-DRB1*15:01* signal was similar in PPMS and RRMS but *HLA-A*02:01* (tagged by rs1143146, $r^2 = 0.95$, controls MAF = 0.55) reached genome-wide significance in RRMS (OR = 0.61, $P = 2.8 \times 10^{-12}$) (**Figure 1**) and was only suggestive of association in PPMS (OR = 0.79, $P = 9.6 \times 10^{-4}$) (**Figure 2**). Rs1143146 was under-represented in RRMS (MAF = 0.43) vs. PPMS (MAF = 0.50) ($P = 5.8 \times 10^{-3}$). **Figure 3** shows a comparison of the *HLA-DRB1*15:01* and *HLA-A*02:01* allele signals in RRMS and PPMS subjects.

A putative gene-based signal was observed in *LIX1L* ($P = 3.8 \times 10^{-6}$) in the PPMS vs. controls analysis. Four rare, damaging *LIX1L* variants were analyzed, and were carried by four different PPMS cases and no RRMS patients or controls. The rare R11H missense mutation (rs142637090) has been observed in only 5 out of 60,703 exomes in ExAC, and the other three variants have not been observed in the ExAC database. The variants are R114H, V173L, S286C, and L322F. Only one of the four variants (R114H) is located in the *LIX1L* putative RNA binding domain. All four variants are in the limb expression 1 pfam domain (**Figure 4**).

The *LIX1* (Limb & CNS Expressed 1) gene was initially identified in a screen for genes transiently expressed during early chicken limb development (Swindell et al., 2001, Mechanisms of Development). The limb expression 1 Pfam domain is present in only two human proteins: *LIX1* and *LIX1L*. *LIX1* is specific to brain and spinal cord; *LIX1L* is expressed in brain and spinal cord but is not specific. The BioPlex protein-protein interaction network indicates that *LIX1L* interacts with the calpain *CAPNS2* and the endogenous calpain inhibitor CAST.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

The invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, and descriptive terms from one or more of the listed claims are introduced into another claim. Where elements are presented as lists, *e.g.*, in Markush group format or as an alternative, each subgroup of the elements is also disclosed, and any element(s) can be removed from the group.

It should it be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements and/or features, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements and/or features. It is also noted that the terms "comprising" and "containing" are intended to be open and permit the inclusion of additional elements or steps. Where ranges are given, endpoints are included. Furthermore, unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or sub–range within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

What is claimed is:

1. A method of evaluating, monitoring, stratifying, or treating, a subject, comprising:

(optionally) a) obtaining a biological sample from a subject;

b) acquiring a genotype (e.g., of a nucleic acid sequence in the biological sample) to identify the presence of a genetic polymorphism in the LIX1L gene;

c) determining from the genotype an increased likelihood of developing or having Primary Progressive Multiple Sclerosis (PPMS) when the genetic polymorphism in the LIX1L gene is/are present; and

d) responsive to said determining,

i) classifying said subject,

ii) selecting said subject for treatment with an multiple sclerosis (MS) therapy, or

iii) administering an MS therapy to said subject.

2. A method of treating Primary Progressive Multiple Sclerosis (PPMS), comprising:

(optionally) a) obtaining a biological sample from a subject;

b) acquiring a genotype from a subject (e.g., of a nucleic acid sequence in the biological sample) to identify the presence or absence of a genetic polymorphism in the LIX1L gene,

c) if the genetic polymorphism in the LIX1L gene is/are present in the subject, the subject has an increased likelihood of developing or having Primary Progressive Multiple Sclerosis (PPMS); and

d) responsive to the presence of the genetic polymorphism in the LIX1L gene,

i) administering a MS therapy;

ii) selecting or altering a dosing of a MS therapy; or

iii) selecting or altering the schedule or time course of a MS therapy.

3. A method of evaluating a subject's likelihood of developing Primary Progressive Multiple Sclerosis (PPMS), comprising:

(optionally) a) identifying a subject who is a blood relative of an individual with MS;

(optionally) b) obtaining a biological sample from the subject;

c) acquiring a genotype from a subject (e.g., of a nucleic acid sequence in the biological sample) to identify the presence of a genetic polymorphism in the LIX1L gene,

d) determining from the genotype an increased likelihood of developing or having Multiple Sclerosis (MS) when the genetic polymorphism in the LIX1L gene is/are present.

4. The method of claim 3, said individual with MS had Primary Progressive Multiple Sclerosis (PPMS).

5. The method of any of claims 1-4, wherein acquiring a genotype comprises acquiring a sequence for a position in an LIX1L gene by sequencing, thereby determining if the polymorphism in the LIX1L gene is present in the nucleic acid molecule.

6. The method of claim 5, wherein the sequence acquired is compared to a reference sequence, or a wild type reference sequence.

7. The method of any of claims 1-4, wherein acquiring a genotype comprises hybridizing one or more probes to the nucleic acid sequence in in an LIX1L gene, wherein the probes are complimentary to the genetic polymorphism in the LIX1L gene, and detecting the polymorphism in the LIX1L gene.

8. The method of any of claims 1-7, wherein the genetic polymorphism comprises single nucleotide polymorphisms.

9. The method of any of claims 1-7, wherein the genetic polymorphism comprises one or more of a stop gain, a splice donor, a splice acceptor, or a missense variant in the LIX1L sequence, or a combination thereof.

10. The method of claim 9, wherein the genetic polymorphism comprise one or more of R114H, V173L, S286C, and L322F, or a combination of two, three or all thereof.

11. The method of claim 10, wherein the genetic polymorphism comprises R114H.

12. The method of claim 10, wherein the genetic polymorphism comprises V173L.

13. The method of claim 10, wherein the genetic polymorphism comprises S286C.

14. The method of claim 10, wherein the genetic polymorphisms comprises L322F.

15. The method of claim 10, wherein the genetic polymorphism comprises one or both of R114H and V173L.

16. The method of claim 10, wherein the genetic polymorphism comprises one or both of R114H and S286C.

17. The method of claim 10, wherein the genetic polymorphism comprises one or both of R114H and L322F.

18. The method of claim 10, wherein the genetic polymorphism comprises one or both of V173L and S286C.

19. The method of claim 10, wherein the genetic polymorphisms comprises one or both of V173L and L322F.

20. The method of claim 10, wherein the genetic polymorphism comprises one or both of S286C and L322F.

21. The method of claim 10, wherein the genetic polymorphism comprises one or more of R114H, V173L, and S286C, or a combination of at least two or three thereof.

22. The method of claim 10, wherein the genetic polymorphism comprises one or more of R114H, V173L, and L322F, or a combination of at least two or three thereof.

23. The method of claim 10, wherein the genetic polymorphism comprises one or more of R114H, S286C, and L322F, or a combination of at least two or three thereof.

24. The method of claim 10, wherein the genetic polymorphism comprises one or more of V173L, S286C, and L322F, or a combination of at least two or three thereof.

25. The method of claim 10, wherein the genetic polymorphism comprises one or more of R114H, V173L, S286C, and L322F, or a combination of at least two, three or all thereof.

26. A method of evaluating, monitoring, stratifying, or treating, a subject, comprising:

(optionally) a) obtaining a biological sample from a subject;

b) acquiring a genotype of a nucleic acid sequence in the biological sample to identify the presence of a HLA-A*02:01 allele,

c) determining from the genotype a decreased likelihood of developing or having Primary Progressive Multiple Sclerosis (PPMS) when the HLA-A*02:01 allele is present;

d) responsive to said determining,

i) classifying said subject,

ii) selecting said subject for treatment with an multiple sclerosis (MS) therapy, or

iii) administering an MS therapy to said subject.

27. The method of any of the preceding claims wherein the MS therapy comprises one or more of an IFN-β 1 molecule; a polymer of glutamic acid, lysine, alanine and tyrosine, *e.g.*, glatiramer (*e.g.*, Copaxone®); an antibody or fragment thereof against alpha-4 integrin, *e.g.*, natalizumab (*e.g.*, Tysabri®); an anthracenedione molecule, *e.g.*, mitoxantrone (*e.g.*, Novantrone®); a fingolimod, *e.g.*, FTY720 (*e.g.*, Gilenya®); a dimethyl fumarate, *e.g.*, an oral dimethyl fumarate (*e.g.*, Tecfidera®); an antibody to the alpha subunit of the IL-2 receptor of T cells (CD25), *e.g.*, daclizumab; an antibody against CD52, *e.g.*, alemtuzumab (*e.g.*, CAMPATH); an inhibitor of a dihydroorotate dehydrogenase, *e.g.*, leflunomide or an active metabolite thereof, *e.g.*, teriflunomide (*e.g.*, AUBAGIO); an antibody to CD20, *e.g.*, rituximab, or ocrelizumab; an anti-LINGO-1 antibody; or a corticosteroid.

28. The method of any of the preceding claims, wherein the subject, *e.g.*, a human subject, has multiple sclerosis (MS).

29. The method of claim 28, wherein the subject with MS has primary progressive MS.

30. The method of any of the preceding claims, wherein the sample comprises blood, e.g., whole blood or peripheral blood.

31. The method of claims 1, 3, or 26 wherein said method further comprises treating, or preventing in, the subject having MS one or more symptoms associated with MS by administering to a subject an MS therapy, in an amount sufficient to reduce one or more symptoms associated with MS.

32. The method of claim 31, wherein symptoms of MS comprise optic neuritis, diplopia, nystagmus, ocular dysmetria, internuclear opthalmoplegia, movement and sound phosphenes, afferent pupillary defect, paresis, monoparesis, paraparesis, hemiparesis, quadraparesis, plegia, paraplegia, hemiplegia, tetraplegia, quadraplegia, spasticity, dysarthria, muscle atrophy, spasms, cramps, hypotonia, clonus, myoclonus, myokymia, restless leg syndrome, footdrop, dysfunctional reflexes, paraesthesia, anaesthesia, neuralgia, neuropathic and neurogenic pain, l'hermitte's, proprioceptive dysfunction, trigeminal neuralgia, ataxia, intention tremor, dysmetria, vestibular ataxia, vertigo, speech ataxia, dystonia, dysdiadochokinesia, frequent micturation, bladder spasticity, flaccid bladder, detrusor-sphincter dyssynergia, erectile dysfunction, anorgasmy, frigidity, constipation, fecal urgency, fecal incontinence, depression, cognitive dysfunction, dementia, mood swings, emotional lability, euphoria, bipolar syndrome, anxiety, aphasia, dysphasia, fatigue, Uhthoff's symptom, gastroesophageal reflux, and sleeping disorders.

33. The method of claim 32, wherein said treating or preventing comprises reducing, retarding or preventing, a relapse, or the worsening of the disease, in the MS subject.

34. The method of any of the preceding claims, further comprising one or more of: performing a neurological examination, evaluating the subject's status on the Expanded Disability Status Scale (EDSS), or detecting the subject's lesion status as assessed using an MRI.

35. A method of determining the presence of a polymorphism in the LIX1L gene in a subject with MS, e.g., PPMS, comprising:

directly acquiring knowledge that a nucleic acid molecule or polypeptide having the genetic polymorphism in the LIX1L gene is present in the subject, e.g., in a sample from the subject.

36. The method of claim 35, wherein the sample comprises blood, e.g., whole blood or peripheral blood.

37. The method of claim 35, wherein the sample is a nucleic acid sample.

38. The method of claim 35, wherein the sample is a protein sample.

39. The method of claim 35, wherein the sample is acquired from the subject.

40. The method of claim 35, wherein the genetic polymorphism comprises one or more of R114H, V173L, S286C, and L322F, or a combination of two, three, or all thereof.

41. The method of claim 35, wherein the genetic polymorphism in the LIX1L gene is detected in a nucleic acid molecule.

42. The method of claim 41, comprising sequencing by a method chosen from one or more of: nucleic acid hybridization assay, amplification-based assays, PCR-RFLP assay, real-time PCR, sequencing, screening analysis, FISH, spectral karyotyping or MFISH, comparative genomic hybridization), *in situ* hybridization, SSP, HPLC or mass-spectrometric genotyping.

43. The method of claim 35, comprising acquiring a sequence for a position in a nucleic acid molecule by sequencing at least one nucleotide of the nucleic acid molecule, thereby determining that the genetic polymorphism in the LIX1L gene is present in the nucleic acid molecule.

44. The method of claim 43, wherein the sequence acquired is compared to a reference sequence, or a wild type reference sequence.

45. The method of claim 35, wherein a polypeptide having the genetic polymorphism in the LIX1L gene is detected.

46. The method of claim 35, comprising: contacting a protein sample with a reagent which specifically binds to a polypeptide having the genetic polymorphism in the LIX1L gene; and detecting the formation of a complex of the polypeptide and the reagent.

47. The method of claim 46, wherein the reagent is labeled with a detectable group to facilitate detection of the bound and unbound reagent.

48. The method of claim 46, wherein the reagent is an antibody molecule.

49. A kit for evaluating a subject's likelihood of developing multiple sclerosis (MS), comprising:

one or more nucleic acid reagents for detecting one or more Primary Progressive Multiple Sclerosis (PPMS)-associated genetic polymorphisms in the LIX1L gene; and

optionally, instructions for use.

50. The kit of claim 49, wherein the nucleic acid reagents comprise: a) probes complimentary to one or more genetic polymorphisms in the LIX1L gene; or b) allele-specific primers or primer pairs that provides for detection of the one or more genetic polymorphisms in the LIX1L gene.

51. The kit of claim 49, wherein the one or more genetic polymorphisms comprise one or more of R114H, V173L, S286C, and L322F.

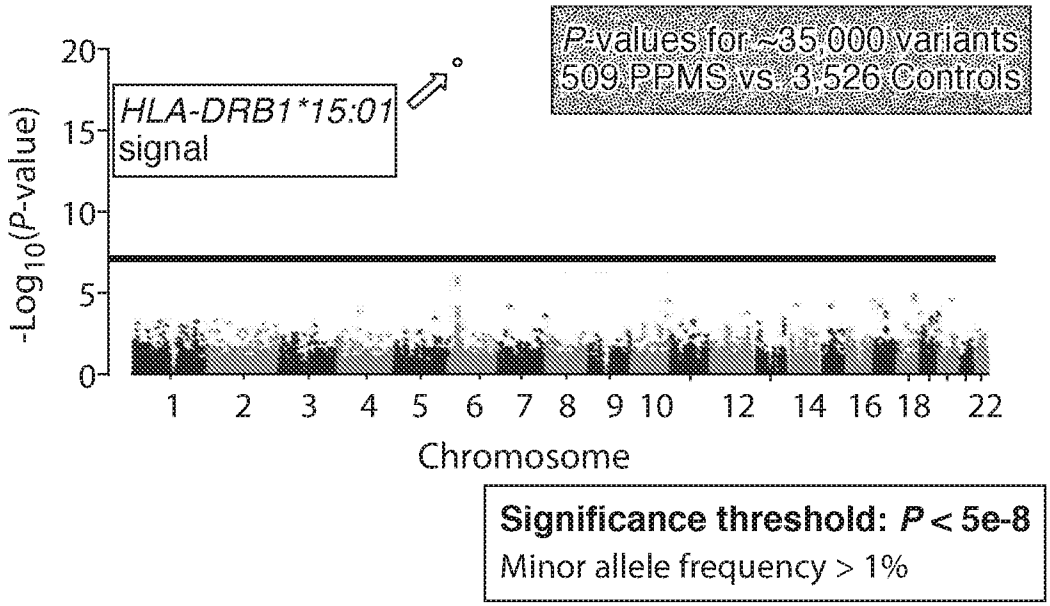
P-values for ~35,000 variants
509 PPMS vs. 3,526 Controls
HLA-DRB1*15:01
signal
-Log10(P-value)
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 12 14 16 18 22
Chromosome
Significance threshold: P < 5e-8
Minor allele frequency > 1%

FIG. 1

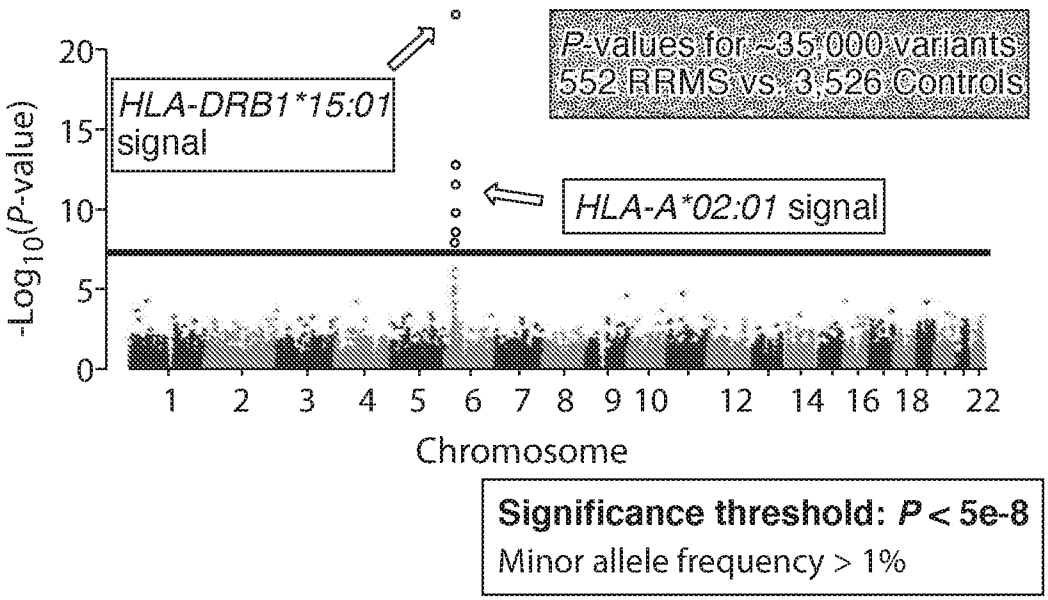
P-values for ~35,000 variants
552 RRMS vs. 3,526 Controls
HLA-DRB1*15:01 signal
HLA-A*02:01 signal
-Log10(P-value)
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 12 14 16 18 22
Chromosome
Significance threshold: P < 5e-8
Minor allele frequency > 1%

FIG. 2

Signal	PPMS vs. Controls		RRMS vs. Controls		PPMS vs. RRMS		Allele Frequency		
	P	OR	P	OR	P	OR	Controls 3,526	PPMS 509	RRMS 552
DRB1*15:01	**4.5e-20**	**1.92**	**4.5e-23**	**1.92**	**0.73**	**1.03**	**0.29**	**0.45**	**0.44**
*A*02:01*	9.6e-04	0.79	2.8e-12	0.61	5.8e-3	1.29	0.55	0.5	0.44

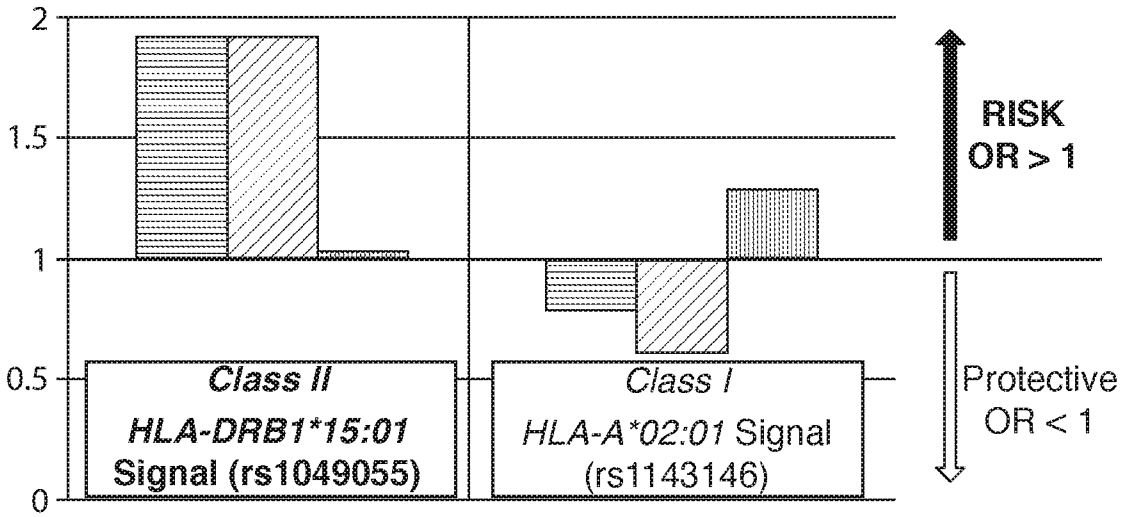

FIG. 3

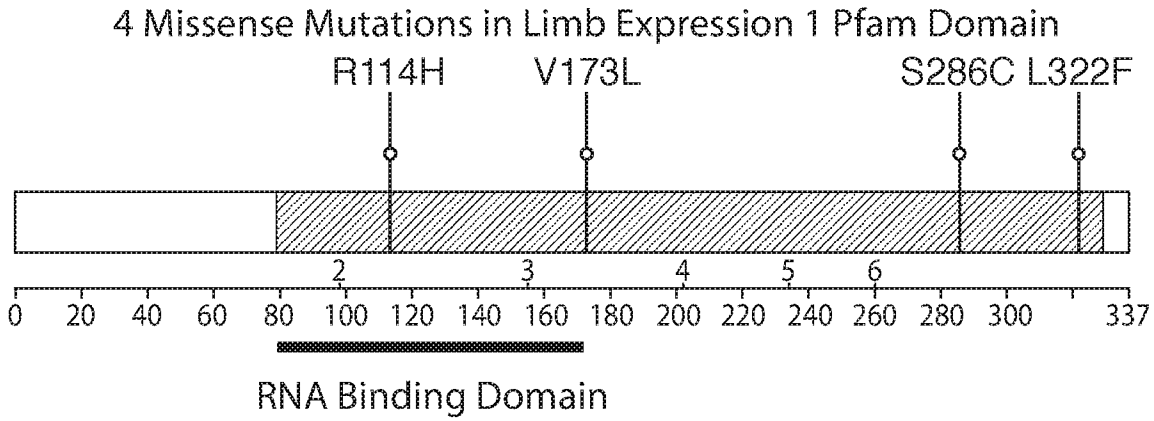
4 Missense Mutations in Limb Expression 1 Pfam Domain
R114H
V173L
S286C
L322F
2
3
4
5
6
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
337
RNA Binding Domain

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/049599

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

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)
C12Q

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)
EPO-Internal, BIOSIS, CHEM ABS Data, EMBASE, FSTA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LISA F. BARCELLOS ET AL: "Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis", HUMAN MOLECULAR GENETICS, vol. 15, no. 18, 15 September 2006 (2006-09-15), pages 2813-2824, XP055418718, gb ISSN: 0964-6906, DOI: 10.1093/hmg/ddl223	1-25, 27-34
A	abstract; page 2816, right col., last section; Fig. 3	35-51

-/--

[X] Further documents are listed in the continuation of Box C. [ ] 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
3 November 2017	09/01/2018

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Leber, Thomas

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/049599

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [ ] forming part of the international application as filed:

      [ ] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [X] furnished subsequent to the international filing date for the purposes of international search only:

      [X] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/049599

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-25, 35-51(completely); 27-34(partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/049599

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LOSONCZI E ET AL: "Tumour necrosis factor alpha gene (TNF-alpha) -376 polymorphism in Hungarian patients with primary progressive multiple sclerosis", JOURNAL OF NEUROIMMUNOLOGY, ELSEVIER SCIENCE PUBLISHERS BV, NL, vol. 208, no. 1-2, 31 March 2009 (2009-03-31), pages 115-118, XP026026877, ISSN: 0165-5728, DOI: 10.1016/J.JNEUROIM.2009.01.004 [retrieved on 2009-02-06]	1-25, 27-34
A	abstract; Table 2	35-51

Y	WANG ZHE ET AL: "Nuclear Receptor NR1H3 in Familial Multiple Sclerosis", NEURON, CELL PRESS, US, vol. 90, no. 5, 1 June 2016 (2016-06-01), pages 948-954, XP029565204, ISSN: 0896-6273, DOI: 10.1016/J.NEURON.2016.04.039	1-25, 27-34
A	abstract	35-51

A	Nn Nn ET AL: "USER GUIDE TaqMan SNP Genotyping Assays", , 30 January 2014 (2014-01-30), pages 1-102, XP055418920, Retrieved from the Internet: URL:https://tools.thermofisher.com/content/sfs/manuals/TaqMan_SNP_Genotyping_Assays_man.pdf [retrieved on 2017-10-25] title sheet	1-25, 27-51

A	Nn Nn: "Reagent for SNP Genotyping Assays for Your search for rs145680047", , 30 January 2014 (2014-01-30), XP055418922, Retrieved from the Internet: URL:https://www.thermofisher.com/order/genome-database/browse/genotyping/keyword/rs145680047?SID=srch-uc-snp-rs145680047&mode=and [retrieved on 2017-10-25] the whole document	1-25, 27-51

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-25, 35-51(completely); 27-34(partially)

Methods of evaluating, monitoring, stratifying, treating a subject, etc. comprising acquiring a genotype, determing on the basis of this genotype an increased likelihood of developing or having Primary Progressive Multiple Sclerosis (PPMS) and, responsive to said determining classifying said subject, selecting said subject for treatment with an MS therapy or administering an MS therapy to said patient whereby the genotype is a polymorphism in the LIX1L gene.

---

2. claims: 26(completely); 27-34(partially)

Methods of evaluating, monitoring, stratifying, treating a subject, etc. comprising acquiring a genotype, determing on the basis of this genotype an increased likelihood of developing or having Primary Progressive Multiple Sclerosis (PPMS) and, responsive to said determining classifying said subject, selecting said subject for treatment with an MS therapy or administering an MS therapy to said patient whereby presence of the HLA-A*02.01 allele is determined.
