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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 June 2010 (10.06.2010)

PCT

(10) International Publication Number
WO 2010/065953 A1

(51) **International Patent Classification**:
C12Q 1/68 (2006.01) *C07H 21/04* (2006.01)

(21) **International Application Number**:
PCT/US2009/066980

(22) **International Filing Date**:
7 December 2009 (07.12.2009)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/200,948 5 December 2008 (05.12.2008) US

(71) **Applicant** *(for all designated States except US)*: **THE CHILDREN'S HOSPITAL OF PHILADELPHIA** [US/US]; Research Institute, Abramson Research Center, Suite 140, Office Of Technology Transfer, 3615 Civic Center Boulevard, Philadelphia, PA 19104 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **BISGAARD, Hans** [DK/DK]; Mosehøjvej 15, DK-2920 Cophenagen (DK). **SLEIMAN, Patrick, M.a.** [IE/US]; 225 S. 18th Street, Philadelphia, PA 19103 (US). **HAKONARSON, Hakon** [IS/US]; 1877 Covered Bridge Road, Malvern, PA 19355 (US).

(74) **Agents**: **RIGAUT, Kathleen, D.** et al.; Dann, Dorfman, Herrell & Skillman, 1601 Market Street, Suite 2400, Philadelphia, PA 19103-2307 (US).

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

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

Published:

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

WO 2010/065953 A1

(54) **Title**: T CELL RECEPTOR ALPHA GENES WHICH PREDISPOSE THE AIRWAY TO BACTERIAL COLONIZATION

(57) **Abstract**: A method for detecting a propensity for increased bacterial colonization of the airway and an increased risk for developing asthma is described The method comprises detecting the presence of at least one single nucleotide polymorphism on chromosome 14 in a target TCR alpha encoding polynucleotide wherein if said single nucleotide polymorphism is present, said patient has an increased risk for increased bacterial airway colonization and an increased risk for developing asthma

T CELL RECEPTOR ALPHA GENES WHICH PREDISPOSE THE AIRWAY TO BACTERIAL COLONIZATION

By

Hans Bisgaard

Patrick M.A. Sleiman

Hakon Hakonarson

This application claims priority to US Provisional Application 60/200,948, filed December 5, 2008, the entire contents being incorporated by reference herein as though set forth in full.

FIELD OF THE INVENTION

This invention relates to the fields of airway disease, immunity, bacterial colonization, and genetic testing. More specifically, the invention provides compositions and methods for the diagnosis and treatment of asthma and other allergic conditions.

BACKGROUND OF THE INVENTION

Several publications and patent documents are cited throughout the specification in order to describe the state of the art to which this invention pertains. Each of these citations is incorporated by reference herein as though set forth in full.

Asthma is a heterogeneous multifactorial disease that manifests as episodes of wheezing, coughing and breathlessness. Airway inflammation together with bronchial hyperresponsiveness, atopy, and elevated IgE and impaired adrenoceptor-mediated airway relaxation are characteristic features of bronchial asthma. Multiple genetic and environmental factors are known to modulate clinical expression of the disease and its associated phenotypes (*2*). However, the genetic susceptibility factors underpinning what is the most common chronic childhood disease remain largely unknown.

Asthma is a major health problem in children and an immense economic burden on the health care systems both in the US and the rest of the world. Since only a small proportion of the heritability of asthma has been unveiled, additional high frequency variants that confer increased asthma risk remain uncovered. GWA is the most robust approach to gene variant discovery (*4*). However, unlike the discovery of multiple variants underlying type-2 diabetes or height (*5*)(*6*), only one locus has been uncovered (*3*), suggesting that either substantially larger numbers of subjects need to be studied or individuals who might possess a higher proportion of 'genetic' disease either through lower age of onset or increased severity need to be ascertained.

Clearly a need exists to identify other genes involved in the manifestation of the asthmatic phenotype. Such knowledge will facilitate diagnosis of this condition as well as provide new targets for the development of potent therapeutics for the treatment of asthma.

SUMMARY OF THE INVENTION

In accordance with the present invention, compositions and methods are provided for identifying children exhibiting increased levels of bacterial colonization in their airways and therefore at an increased risk for developing asthma. An exemplary method entails detecting the presence of a single nucleotide polymorphism on chromosome 14 in a TCRa encoding polynucleotide wherein if the single nucleotide polymorphism is present, the patient has an a propensity for enhanced bacterial colonization of the airway and a subsequent increased risk for developing asthma. Exemplary single nucleotide polymorphisms associated with the development of asthma include, without limitation, those disclosed in Table 1 hereinbelow.

The methods of the invention can include alternative means for detecting the disclosed polymorphisms. For example, such methods of detection can further comprises processes such as specific hybridization, measurement of allele size, restriction fragment length polymorphism analysis, allele-specific hybridization analysis, single base primer extension reaction, and sequencing of an amplified polynucleotide.

In a preferred embodiment, the polymorphism is on chromosome 14 and is provided in Table 1.

In yet another aspect, nucleic acid molecules useful for amplifying the nucleic acids encoding the single nucleotide polymorphisms disclosed herein are provided. Also provided are solid supports comprising suitable nucleic acid targets to facilitate detection of such SNPS in patient samples. A suitable solid support for this process includes a microarray.

Finally, the invention also encompasses screening methods to identify agents which differentially modulate TCRa action or function in the SNP containing cells described herein. An exemplary method entails providing cells comprising at least one of the SNPs disclosed in Table 1; providing cells which express these gene(s) which lack the cognate polymorphisms (step b); contacting each cell type with a test agent and analyzing whether said agent alters T cell function and/or signaling , agents so identified are also within the scope of the invention.

Also provided are transgenic mice comprising the SNP containing nucleic acid molecules described herein. Such mice provide a superior in vivo screening tool to identify agents which modulate the progression and development of asthma.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a linkage disequilibrium heatmap.

Figure 2 shows multiple SNPs and their association with $FEV_{0.5}$.

Figure 3 shows multiple SNPs and their association with PD_{15} TcO_2.

Figure 4 shows multiple SNPs and their association with PD_{20} metacholine.

Figure 5 shows multiple SNPs and their association with FEV_1.

Figure 6 shows multiple SNPs and their association with FEV_1/FVC.

Figure 7 shows multiple SNPs and their association with MMEF.

Figure 8 shows multiple SNPs and their association with sRaw profiles.

Figure 9 shows multiple SNPs and their association with eosinophilocytes at age 6 years.

Figure 10 shows multiple SNPs and their association with asthma status at age 6 years.

Figure 11 shows multiple SNPs and their association with wheezy categorization.

Figure 12 shows multiple SNPs and their association with bacterial colonization.

Figure 13 shows multiple SNPs and their association with JJJ.

Figure 14 shows multiple SNPs and their association with J45.

Figure 15 shows multiple SNPs and their association with exacerbation.

Figure 16 shows multiple SNPs and their association with exacerbation incidence.

Figure 17 shows multiple SNPs and their association with eczema.

Figure 18 shows multiple SNPs and their association with asthma related events.

Figure 19 shows haplotype association analysis performed against colonization phenotype. Significant association detected with haplotypes formed by block 1 which includes the most highly associated SNP rs227870. A single haplotype (the second most common in Europeans) is highly associated with the colonization phenotype.

DETAILED DESCRIPTION OF THE INVENTION

Asthma is the most common chronic disease in children across all developed countries. The incidence of the disease is currently increasing world-wide. Family and twin studies strongly implicate genetic factors in an individual's risk of developing of asthma. Despite this, it has proven difficult to isolate disease genes that confer susceptibility to this disease using classical candidate gene and linkage approaches, with the notable exception of linkages to the HLA loci on chr 5q31-q33 and 6p21. Over the last two years, genome-wide association (GWA) studies have become feasible, where modern high-throughput SNP genotyping technologies can be applied to large and comprehensively phenotyped patient cohorts. Such approaches have enabled scientists to robustly associate specific variants with many complex diseases, including age-related macular degeneration, type 2 diabetes, breast cancer and IBD. In the asthma field, one study has been reported to date showing positive association with the ORMDL3 gene on chromosome 17q12, and our group has replicated this

finding and identified another common variant on chr 1q31 that resides in the DENND1B gene and predisposes to asthma.

This invention pertains to the discovery of another novel variant that confers a predisposition to bacterial colonization of the airways of newborns and infants and associates robustly with the development of asthma in later life. Association results are presented that for the first time make the link between bacterial colonization of newborns and infants and the development of asthma in later life. This discovery will facilitate identifying those patients most likely to develop asthma and also provides new targets for the design agents and compositions useful for the treatment of asthma.

Definitions:

For purposes of the present invention, "a" or "an" entity refers to one or more of that entity; for example, "a cDNA" refers to one or more cDNA or at least one cDNA. As such, the terms "a" or "an," "one or more" and "at least one" can be used interchangeably herein. It is also noted that the terms "comprising," "including," and "having" can be used interchangeably. Furthermore, a compound "selected from the group consisting of" refers to one or more of the compounds in the list that follows, including mixtures (i.e. combinations) of two or more of the compounds. According to the present invention, an isolated, or biologically pure molecule is a compound that has been removed from its natural milieu. As such, Aisolated@ and Abiologically pure@ do not necessarily reflect the extent to which the compound has been purified. An isolated compound of the present invention can be obtained from its natural source, can be produced using laboratory synthetic techniques or can be produced by any such chemical synthetic route.

"Enhanced bacterial colonization/asthma-associated SNP or specific marker" is a SNP or marker which is associated with an increased or decreased predisposition to bacterial colonization, such increased colonization conferring an risk of developing asthma not found normal patients who do not have this disease. Such markers may include but are not limited to nucleic acids, proteins encoded thereby, or other small molecules.

A "single nucleotide polymorphism (SNP)" refers to a change in which a single base in the DNA differs from the usual base at that position. These single base changes are called SNPs or "snips." Millions of SNP's have been cataloged in the human genome. Some SNPs

such as that which causes sickle cell are responsible for disease. Other SNPs are normal variations in the genome.

The term "genetic alteration" as used herein refers to a change from the wild-type or reference sequence of one or more nucleic acid molecules. Genetic alterations include without limitation, base pair substitutions, additions and deletions of at least one nucleotide from a nucleic acid molecule of known sequence.

The term "solid matrix" as used herein refers to any format, such as beads, microparticles, a microarray, the surface of a microtitration well or a test tube, a dipstick or a filter. The material of the matrix may be polystyrene, cellulose, latex, nitrocellulose, nylon, polyacrylamide, dextran or agarose.

The phrase "consisting essentially of" when referring to a particular nucleotide or amino acid means a sequence having the properties of a given SEQ ID NO:. For example, when used in reference to an amino acid sequence, the phrase includes the sequence per se and molecular modifications that would not affect the functional and novel characteristics of the sequence.

"Target nucleic acid" as used herein refers to a previously defined region of a nucleic acid present in a complex nucleic acid mixture wherein the defined wild-type region contains at least one known nucleotide variation which may or may not be associated with asthma. The nucleic acid molecule may be isolated from a natural source by cDNA cloning or subtractive hybridization or synthesized manually. The nucleic acid molecule may be synthesized manually by the triester synthetic method or by using an automated DNA synthesizer.

With regard to nucleic acids used in the invention, the term "isolated nucleic acid" is sometimes employed. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a prokaryote or eukaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule. An isolated

nucleic acid molecule inserted into a vector is also sometimes referred to herein as a recombinant nucleic acid molecule.

With respect to RNA molecules, the term "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form.

By the use of the term "enriched" in reference to nucleic acid it is meant that the specific DNA or RNA sequence constitutes a significantly higher fraction (2-5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two. However, it should be noted that "enriched" does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment (compared to the natural level, this level should be at least 2-5 fold greater, e.g., in terms of mg/ml). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones can be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10^{6}-fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. Thus the term "substantially pure" refers to a

preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid, oligonucleotide, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest.

The term "complementary" describes two nucleotides that can form multiple favorable interactions with one another. For example, adenine is complementary to thymine as they can form two hydrogen bonds. Similarly, guanine and cytosine are complementary since they can form three hydrogen bonds. Thus if a nucleic acid sequence contains the following sequence of bases, thymine, adenine, guanine and cytosine, a "complement" of this nucleic acid molecule would be a molecule containing adenine in the place of thymine, thymine in the place of adenine, cytosine in the place of guanine, and guanine in the place of cytosine. Because the complement can contain a nucleic acid sequence that forms optimal interactions with the parent nucleic acid molecule, such a complement can bind with high affinity to its parent molecule.

With respect to single stranded nucleic acids, particularly oligonucleotides, the term "specifically hybridizing" refers to the association between two single-stranded nucleotide molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule of the invention, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence. For example, specific hybridization can refer to a sequence which hybridizes to any asthma specific marker nucleic acid, but does not hybridize to other nucleotides. Also polynucleotide which "specifically hybridizes" may hybridize only to a TCRa specific marker, such as the TCRa-specific markers shown in the Tables contained herein. Appropriate conditions enabling specific hybridization of single stranded nucleic acid molecules of varying complementarity are well known in the art.

For instance, one common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology is set forth below (Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989):

$T_m = 81.5^{\circ}C + 16.6\text{Log}\,[Na+] + 0.41(\%\ G+C) - 0.63\ (\%\ \text{formamide}) - 600/\#bp\ \text{in duplex}$

As an illustration of the above formula, using [Na+] = [0.368] and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C.

The stringency of the hybridization and wash depend primarily on the salt concentration and temperature of the solutions. In general, to maximize the rate of annealing of the probe with its target, the hybridization is usually carried out at salt and temperature conditions that are 20-25°C below the calculated T_m of the hybrid. Wash conditions should be as stringent as possible for the degree of identity of the probe for the target. In general, wash conditions are selected to be approximately 12-20°C below the T_m of the hybrid. In regards to the nucleic acids of the current invention, a moderate stringency hybridization is defined as hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 µg/ml denatured salmon sperm DNA at 42°C, and washed in 2X SSC and 0.5% SDS at 55°C for 15 minutes. A high stringency hybridization is defined as hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 µg/ml denatured salmon sperm DNA at 42°C, and washed in 1X SSC and 0.5% SDS at 65°C for 15 minutes. A very high stringency hybridization is defined as hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 µg/ml denatured salmon sperm DNA at 42°C, and washed in 0.1X SSC and 0.5% SDS at 65°C for 15 minutes.

The term "oligonucleotide," as used herein is defined as a nucleic acid molecule comprised of two or more ribo- or deoxyribonucleotides, preferably more than three. The exact size of the oligonucleotide will depend on various factors and on the particular application and use of the oligonucleotide. Oligonucleotides, which include probes and primers, can be any length from 3 nucleotides to the full length of the nucleic acid molecule, and explicitly include every possible number of contiguous nucleic acids from 3 through the full length of the polynucleotide. Preferably, oligonucleotides are at least about 10

nucleotides in length, more preferably at least 15 nucleotides in length, more preferably at least about 20 nucleotides in length.

The term "probe" as used herein refers to an oligonucleotide, polynucleotide or nucleic acid, either RNA or DNA, whether occurring naturally as in a purified restriction enzyme digest or produced synthetically, which is capable of annealing with or specifically hybridizing to a nucleic acid with sequences complementary to the probe. A probe may be either single-stranded or double-stranded. The exact length of the probe will depend upon many factors, including temperature, source of probe and use of the method. For example, for diagnostic applications, depending on the complexity of the target sequence, the oligonucleotide probe typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. The probes herein are selected to be complementary to different strands of a particular target nucleic acid sequence. This means that the probes must be sufficiently complementary so as to be able to "specifically hybridize" or anneal with their respective target strands under a set of pre-determined conditions. Therefore, the probe sequence need not reflect the exact complementary sequence of the target. For example, a non-complementary nucleotide fragment may be attached to the 5' or 3' end of the probe, with the remainder of the probe sequence being complementary to the target strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the sequence of the target nucleic acid to anneal therewith specifically.

The term "primer" as used herein refers to an oligonucleotide, either RNA or DNA, either single-stranded or double-stranded, either derived from a biological system, generated by restriction enzyme digestion, or produced synthetically which, when placed in the proper environment, is able to functionally act as an initiator of template-dependent nucleic acid synthesis. When presented with an appropriate nucleic acid template, suitable nucleoside triphosphate precursors of nucleic acids, a polymerase enzyme, suitable cofactors and conditions such as a suitable temperature and pH, the primer may be extended at its 3' terminus by the addition of nucleotides by the action of a polymerase or similar activity to yield a primer extension product. The primer may vary in length depending on the particular conditions and requirement of the application. For example, in diagnostic applications, the oligonucleotide primer is typically 15-25 or more nucleotides in length. The primer must be of sufficient complementarity to the desired template to prime the synthesis of the desired

extension product, that is, to be able anneal with the desired template strand in a manner sufficient to provide the 3' hydroxyl moiety of the primer in appropriate juxtaposition for use in the initiation of synthesis by a polymerase or similar enzyme. It is not required that the primer sequence represent an exact complement of the desired template. For example, a non-complementary nucleotide sequence may be attached to the 5' end of an otherwise complementary primer. Alternatively, non-complementary bases may be interspersed within the oligonucleotide primer sequence, provided that the primer sequence has sufficient complementarity with the sequence of the desired template strand to functionally provide a template-primer complex for the synthesis of the extension product.

Polymerase chain reaction (PCR) has been described in US Patents 4,683,195, 4,800,195, and 4,965,188, the entire disclosures of which are incorporated by reference herein.

An "siRNA" refers to a molecule involved in the RNA interference process for a sequence-specific post-transcriptional gene silencing or gene knockdown by providing small interfering RNAs (siRNAs) that has homology with the sequence of the targeted gene. Small interfering RNAs (siRNAs) can be synthesized in vitro or generated by ribonuclease III cleavage from longer dsRNA and are the mediators of sequence-specific mRNA degradation. Preferably, the siRNA of the invention are chemically synthesized using appropriately protected ribonucleoside phosphoramidites and a conventional DNA/RNA synthesizer. The siRNA can be synthesized as two separate, complementary RNA molecules, or as a single RNA molecule with two complementary regions. Commercial suppliers of synthetic RNA molecules or synthesis reagents include Applied Biosystems (Foster City, CA, USA), Proligo (Hamburg, Germany), Dharmacon Research (Lafayette, Colo., USA), Pierce Chemical (part of Perbio Science, Rockford, Ill., USA), Glen Research (Sterling, Va., USA), ChemGenes (Ashland, Mass., USA) and Cruachem (Glasgow, UK). Specific siRNA constructs for inhibiting DENN/D1B mRNA, for example, may be between 15-35 nucleotides in length, and more typically about 21 nucleotides in length.

The term "vector" relates to a single or double stranded circular nucleic acid molecule that can be infected, transfected or transformed into cells and replicate independently or within the host cell genome. A circular double stranded nucleic acid molecule can be cut and thereby linearized upon treatment with restriction enzymes. An assortment of vectors, restriction enzymes, and the knowledge of the nucleotide sequences that are targeted by

restriction enzymes are readily available to those skilled in the art, and include any replicon, such as a plasmid, cosmid, bacmid, phage or virus, to which another genetic sequence or element (either DNA or RNA) may be attached so as to bring about the replication of the attached sequence or element. A nucleic acid molecule of the invention can be inserted into a vector by cutting the vector with restriction enzymes and ligating the two pieces together.

Many techniques are available to those skilled in the art to facilitate transformation, transfection, or transduction of the expression construct into a prokaryotic or eukaryotic organism. The terms "transformation", "transfection", and Atransduction@ refer to methods of inserting a nucleic acid and/or expression construct into a cell or host organism. These methods involve a variety of techniques, such as treating the cells with high concentrations of salt, an electric field, or detergent, to render the host cell outer membrane or wall permeable to nucleic acid molecules of interest, microinjection, PEG-fusion, and the like.

The term "promoter element" describes a nucleotide sequence that is incorporated into a vector that, once inside an appropriate cell, can facilitate transcription factor and/or polymerase binding and subsequent transcription of portions of the vector DNA into mRNA. In one embodiment, the promoter element of the present invention precedes the 5' end of the marker nucleic acid molecule such that the latter is transcribed into mRNA. Host cell machinery then translates mRNA into a polypeptide.

Those skilled in the art will recognize that a nucleic acid vector can contain nucleic acid elements other than the promoter element and the asthma specific marker encoding nucleic acid. These other nucleic acid elements include, but are not limited to, origins of replication, ribosomal binding sites, nucleic acid sequences encoding drug resistance enzymes or amino acid metabolic enzymes, and nucleic acid sequences encoding secretion signals, localization signals, or signals useful for polypeptide purification.

A "replicon" is any genetic element, for example, a plasmid, cosmid, bacmid, plastid, phage or virus, that is capable of replication largely under its own control. A replicon may be either RNA or DNA and may be single or double stranded.

An "expression operon" refers to a nucleic acid segment that may possess transcriptional and translational control sequences, such as promoters, enhancers, translational start signals (e.g., ATG or AUG codons), polyadenylation signals, terminators,

and the like, and which facilitate the expression of a polypeptide coding sequence in a host cell or organism.

As used herein, the terms "reporter," "reporter system", "reporter gene," or "reporter gene product" shall mean an operative genetic system in which a nucleic acid comprises a gene that encodes a product that when expressed produces a reporter signal that is a readily measurable, e.g., by biological assay, immunoassay, radio immunoassay, or by colorimetric, fluorogenic, chemiluminescent or other methods. The nucleic acid may be either RNA or DNA, linear or circular, single or double stranded, antisense or sense polarity, and is operatively linked to the necessary control elements for the expression of the reporter gene product. The required control elements will vary according to the nature of the reporter system and whether the reporter gene is in the form of DNA or RNA, but may include, but not be limited to, such elements as promoters, enhancers, translational control sequences, poly A addition signals, transcriptional termination signals and the like.

The introduced nucleic acid may or may not be integrated (covalently linked) into nucleic acid of the recipient cell or organism. In bacterial, yeast, plant and mammalian cells, for example, the introduced nucleic acid may be maintained as an episomal element or independent replicon such as a plasmid. Alternatively, the introduced nucleic acid may become integrated into the nucleic acid of the recipient cell or organism and be stably maintained in that cell or organism and further passed on or inherited to progeny cells or organisms of the recipient cell or organism. Finally, the introduced nucleic acid may exist in the recipient cell or host organism only transiently.

The term "selectable marker gene" refers to a gene that when expressed confers a selectable phenotype, such as antibiotic resistance, on a transformed cell.

The term "operably linked" means that the regulatory sequences necessary for expression of the coding sequence are placed in the DNA molecule in the appropriate positions relative to the coding sequence so as to effect expression of the coding sequence. This same definition is sometimes applied to the arrangement of transcription units and other transcription control elements (e.g. enhancers) in an expression vector.

The terms “recombinant organism,” or “transgenic organism” refer to organisms which have a new combination of genes or nucleic acid molecules. A new combination of

genes or nucleic acid molecules can be introduced into an organism using a wide array of nucleic acid manipulation techniques available to those skilled in the art. The term "organism" relates to any living being comprised of a least one cell. An organism can be as simple as one eukaryotic cell or as complex as a mammal. Therefore, the phrase "a recombinant organism" encompasses a recombinant cell, as well as eukaryotic and prokaryotic organism.

The term "isolated protein" or "isolated and purified protein" is sometimes used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein that has been sufficiently separated from other proteins with which it would naturally be associated, so as to exist in "substantially pure" form. "Isolated" is not meant to exclude artificial or synthetic mixtures with other compounds or materials, or the presence of impurities that do not interfere with the fundamental activity, and that may be present, for example, due to incomplete purification, addition of stabilizers, or compounding into, for example, immunogenic preparations or pharmaceutically acceptable preparations.

A "specific binding pair" comprises a specific binding member (sbm) and a binding partner (bp) which have a particular specificity for each other and which in normal conditions bind to each other in preference to other molecules. Examples of specific binding pairs are antigens and antibodies, ligands and receptors and complementary nucleotide sequences. The skilled person is aware of many other examples. Further, the term "specific binding pair" is also applicable where either or both of the specific binding member and the binding partner comprise a part of a large molecule. In embodiments in which the specific binding pair comprises nucleic acid sequences, they will be of a length to hybridize to each other under conditions of the assay, preferably greater than 10 nucleotides long, more preferably greater than 15 or 20 nucleotides long.

"Sample" or "patient sample" or "biological sample" generally refers to a sample which may be tested for a particular molecule, preferably an bacterial colonization/asthma specific marker molecule, such as a marker shown in the tables provided below. Samples may include but are not limited to cells, body fluids, including blood, serum, plasma, urine, saliva, tears, pleural fluid and the like.

The terms "agent" and "test compound" are used interchangeably herein and denote a chemical compound, a mixture of chemical compounds, a biological macromolecule, or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian) cells or tissues. Biological macromolecules include siRNA, shRNA, antisense oligonucleotides, peptides, peptide/DNA complexes, and any nucleic acid based molecule which exhibits the capacity to modulate the activity of the SNP containing nucleic acids described herein or their encoded proteins. Agents are evaluated for potential biological activity by inclusion in screening assays described hereinbelow.

METHODS OF USING ENHANCED BACTERIAL COLONIZATION/ASTHMA-ASSOCIATED SNPS FOR DIAGNOSING A PROPENSITY FOR THE DEVELOPMENT OF ASTHMA

Bacterial colonization/asthma-related-SNP containing nucleic acids, including but not limited to those listed in the Tables provided below may be used for a variety of purposes in accordance with the present invention. Enhanced bacterial colonization/asthma-associated SNP containing DNA, RNA, or fragments thereof may be used as probes to detect the presence of and/or expression of the same in biological samples. Methods in which the specific marker nucleic acids may be utilized as probes for such assays include, but are not limited to: (1) *in situ* hybridization; (2) Southern hybridization (3) northern hybridization; and (4) assorted amplification reactions such as polymerase chain reactions (PCR).

Further, assays for detecting enhanced bacterial colonization/asthma-associated SNPs or the proteins encoded thereby may be conducted on any type of biological sample, including but not limited to body fluids (including blood, urine, serum, gastric lavage), any type of cell (such as brain cells, white blood cells, mononuclear cells) or body tissue.

From the foregoing discussion, it can be seen that enhanced bacterial colonization/asthma-associated SNP containing nucleic acids, vectors expressing the same, SNP containing marker proteins and anti-SNP specific marker antibodies of the invention can be used to detect such SNPs in body tissue, cells, or fluid, and alter SNP containing marker protein expression for purposes of assessing the genetic and protein interactions involved in the predisposition to bacterial colonization and the subsequent development of asthma.

In most embodiments for screening for enhanced bacterial colonization/asthma-associated SNPs, the SNP containing nucleic acid in the sample will initially be amplified, e.g. using PCR, to increase the amount of the templates as compared to other sequences present in the sample. This allows the target sequences to be detected with a high degree of sensitivity if they are present in the sample. This initial step may be avoided by using highly sensitive array techniques that are becoming increasingly important in the art.

Alternatively, new detection technologies can overcome this limitation and enable analysis of small samples containing as little as 1μg of total RNA. Using Resonance Light Scattering (RLS) technology, as opposed to traditional fluorescence techniques, multiple reads can detect low quantities of mRNAs using biotin labeled hybridized targets and anti-biotin antibodies. Another alternative to PCR amplification involves planar wave guide technology (PWG) to increase signal-to-noise ratios and reduce background interference. Both techniques are commercially available from Qiagen Inc. (USA).

Thus any of the aforementioned techniques may be used to detect or quantify enhanced bacterial colonization/asthma-associated SNP marker expression and accordingly, diagnose an increased risk of developing asthma.

KITS AND ARTICLES OF MANUFACTURE

Any of the aforementioned products can be incorporated into a kit which may contain a enhanced bacterial colonization/asthma-associated SNP specific marker polynucleotide or one or more such markers immobilized on a Gene Chip, an oligonucleotide, a polypeptide, a peptide, an antibody, a label, marker, or reporter, a pharmaceutically acceptable carrier, a physiologically acceptable carrier, instructions for use, a container, a vessel for administration, an assay substrate, or any combination thereof.

METHODS OF USING ENHANCED BACTERIAL COLONIZATION/ASTHMA-ASSOCIATED SNPS FOR DEVELOPMENT OF THERAPEUTIC AGENTS

Since the SNPs identified herein have been associated with the etiology of asthma, methods for identifying agents that modulate the activity of the genes and their encoded products containing such SNPs should result in the generation of efficacious therapeutic agents for the treatment of this condition.

Chromosome 14 contains TCRa protein coding regions which provide suitable targets for the rational design of therapeutic agents which modulate their activity. Small peptide molecules corresponding to these regions may be used to advantage in the design of therapeutic agents which effectively modulate the activity of the encoded proteins.

Molecular modeling should facilitate the identification of specific organic molecules with capacity to bind to the active site of the proteins encoded by the SNP containing nucleic acids based on conformation or key amino acid residues required for function. A combinatorial chemistry approach will be used to identify molecules with greatest activity and then iterations of these molecules will be developed for further cycles of screening.

The polypeptides or fragments employed in drug screening assays may either be free in solution, affixed to a solid support or within a cell. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may determine, for example, formation of complexes between the polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between the polypeptide or fragment and a known substrate is interfered with by the agent being tested.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity for the encoded polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on Sep. 13, 1984. Briefly stated, large numbers of different, small peptide test compounds, such as those described above, are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with the target polypeptide and washed. Bound polypeptide is then detected by methods well known in the art.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as T cells) which have a nonfunctional or altered bacterial colonization/asthma associated TCRa gene. These host cell lines or cells therefore encode TCRa altered at the polypeptide level. The host cell lines or cells are exposed to known asthma triggers, including bacterial toxins, allergens, viral particles, or other activators and then grown in the presence of a test compound to determine if the compound is capable of regulating T cell

signialling or function in altered cells. The enhanced bacterial colonization/asthma-associated SNP containing TCRa encoding DNA molecules may be introduced singly into such host cells or in combination to assess the phenotype of cells conferred by such expression. Methods for introducing DNA molecules are also well known to those of ordinary skill in the art. Such methods are set forth in Ausubel et al. eds., Current Protocols in Molecular Biology, John Wiley & Sons, NY, N.Y. 1995, the disclosure of which is incorporated by reference herein.

A wide variety of expression vectors are available that can be modified to express the novel DNA sequences of this invention. The specific vectors exemplified herein are merely illustrative, and are not intended to limit the scope of the invention. Expression methods are described by Sambrook et al. Molecular Cloning: A Laboratory Manual or Current Protocols in Molecular Biology 16.3-17.44 (1989). Expression methods in Saccharomyces are also described in Current Protocols in Molecular Biology (1989).

Suitable vectors for use in practicing the invention include prokaryotic vectors such as the pNH vectors (Stratagene Inc., 11099 N. Torrey Pines Rd., La Jolla, Calif. 92037), pET vectors (Novogen Inc., 565 Science Dr., Madison, Wis. 53711) and the pGEX vectors (Pharmacia LKB Biotechnology Inc., Piscataway, N.J. 08854). Examples of eukaryotic vectors useful in practicing the present invention include the vectors pRc/CMV, pRc/RSV, and pREP (Invitrogen, 11588 Sorrento Valley Rd., San Diego, Calif. 92121); pcDNA3.1/V5&His (Invitrogen); baculovirus vectors such as pVL1392, pVL1393, or pAC360 (Invitrogen); and yeast vectors such as YRP17, YIP5, and YEP24 (New England Biolabs, Beverly, Mass.), as well as pRS403 and pRS413 Stratagene Inc.); Picchia vectors such as pHIL-D1 (Phillips Petroleum Co., Bartlesville, Okla. 74004); retroviral vectors such as PLNCX and pLPCX (Clontech); and adenoviral and adeno-associated viral vectors.

Promoters for use in expression vectors of this invention include promoters that are operable in prokaryotic or eukaryotic cells. Promoters that are operable in prokaryotic cells include lactose (lac) control elements, bacteriophage lambda (pL) control elements, arabinose control elements, tryptophan (trp) control elements, bacteriophage T7 control elements, and hybrids thereof. Promoters that are operable in eukaryotic cells include Epstein Barr virus promoters, adenovirus promoters, SV40 promoters, Rous Sarcoma Virus promoters, cytomegalovirus (CMV) promoters, baculovirus promoters such as AcMNPV polyhedrin promoter, Picchia promoters such as the alcohol oxidase promoter, and Saccharomyces

promoters such as the gal4 inducible promoter and the PGK constitutive promoter. In addition, a vector of this invention may contain any one of a number of various markers facilitating the selection of a transformed host cell. Such markers include genes associated with temperature sensitivity, drug resistance, or enzymes associated with phenotypic characteristics of the host organisms.

Host cells expressing the enhanced bacterial colonization/asthma-associated SNPs of the present invention or functional fragments thereof provide a system in which to screen potential compounds or agents for the ability to modulate bacterial colonization and/or the development of asthma. Thus, in one embodiment, the nucleic acid molecules of the invention may be used to create recombinant cell lines for use in assays to identify agents which modulate aspects of airway bacterial colonization and the subsequent cellular signaling associated with the development of asthma and aberrant bronchoconstriction. Also provided herein are methods to screen for compounds capable of modulating the function of proteins encoded by SNP containing nucleic acids.

Another approach entails the use of phage display libraries engineered to express fragment of the polypeptides encoded by the SNP containing nucleic acids on the phage surface. Such libraries are then contacted with a combinatorial chemical library under conditions wherein binding affinity between the expressed peptide and the components of the chemical library may be detected. US Patents 6,057,098 and 5,965,456 provide methods and apparatus for performing such assays.

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide in vivo. See, e.g., Hodgson, (1991) Bio/Technology 9:19-21. In one approach, discussed above, the three-dimensional structure of a protein of interest or, for example, of the protein-substrate complex, is solved by x-ray crystallography, by nuclear magnetic resonance, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., (1990) Science 249:527-533). In addition, peptides may be analyzed by an alanine scan (Wells, (1991) Meth. Enzym. 202:390-

411). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based.

One can bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original molecule. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of polypeptide activity. By virtue of the availability of SNP containing nucleic acid sequences described herein, sufficient amounts of the encoded polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

In another embodiment, the availability of enhanced bacterial colonization/asthma-associated SNP containing nucleic acids enables the production of strains of laboratory mice carrying the enhanced bacterial colonization/asthma-associated SNPs of the invention. Transgenic mice expressing the enhanced bacterial colonization/asthma-associated SNP of the invention provide a model system in which to examine the role of the altered TCRa protein encoded by the SNP containing nucleic acid in bacterial colonization mediated development and progression towards asthma. Methods of introducing transgenes in laboratory mice are known to those of skill in the art. Three common methods include: 1. integration of retroviral vectors encoding the foreign gene of interest into an early embryo; 2. injection of DNA into the pronucleus of a newly fertilized egg; and 3. the incorporation of genetically manipulated embryonic stem cells into an early embryo. Production of the transgenic mice described above will facilitate the molecular elucidation of the role that

altered TCRa protein plays in various processes associated with the increased bacterial colonization and progression towards the asthmatic phenotype. Such mice provide an in vivo screening tool to study putative therapeutic drugs in a whole animal model and are encompassed by the present invention.

The term "animal" is used herein to include all vertebrate animals, except humans. It also includes an individual animal in all stages of development, including embryonic and fetal stages. A "transgenic animal" is any animal containing one or more cells bearing genetic information altered or received, directly or indirectly, by deliberate genetic manipulation at the subcellular level, such as by targeted recombination or microinjection or infection with recombinant virus. The term "transgenic animal" is not meant to encompass classical cross-breeding or in vitro fertilization, but rather is meant to encompass animals in which one or more cells are altered by or receive a recombinant DNA molecule. This molecule may be specifically targeted to a defined genetic locus, be randomly integrated within a chromosome, or it may be extrachromosomally replicating DNA. The term "germ cell line transgenic animal" refers to a transgenic animal in which the genetic alteration or genetic information was introduced into a germ line cell, thereby conferring the ability to transfer the genetic information to offspring. If such offspring, in fact, possess some or all of that alteration or genetic information, then they, too, are transgenic animals.

The alteration of genetic information may be foreign to the species of animal to which the recipient belongs, or foreign only to the particular individual recipient, or may be genetic information already possessed by the recipient. In the last case, the altered or introduced gene may be expressed differently than the native gene. Such altered or foreign genetic information would encompass the introduction of enhanced bacterial colonization/asthma-associated SNP containing nucleotide sequences.

The DNA used for altering a target gene may be obtained by a wide variety of techniques that include, but are not limited to, isolation from genomic sources, preparation of cDNAs from isolated mRNA templates, direct synthesis, or a combination thereof.

A preferred type of target cell for transgene introduction is the embryonal stem cell (ES). ES cells may be obtained from pre-implantation embryos cultured in vitro (Evans et al., (1981) Nature 292:154-156; Bradley et al., (1984) Nature 309:255-258; Gossler et al., (1986) Proc. Natl. Acad. Sci. 83:9065-9069). Transgenes can be efficiently introduced into the ES

cells by standard techniques such as DNA transfection or by retrovirus-mediated transduction. The resultant transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The introduced ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal.

One approach to the problem of determining the contributions of altered TCRa genes and their expression products is to use isolated enhanced bacterial colonization/asthma-associated SNP containing TCRa genes as insertional cassettes to selectively inactivate a wild-type gene in totipotent ES cells (such as those described above) and then generate transgenic mice. The use of gene-targeted ES cells in the generation of gene-targeted transgenic mice was described, and is reviewed elsewhere (Frohman et al., (1989) Cell 56:145-147; Bradley et al., (1992) Bio/Technology 10:534-539).

Techniques are available to inactivate or alter any genetic region to a mutation desired by using targeted homologous recombination to insert specific changes into chromosomal alleles. However, in comparison with homologous extrachromosomal recombination, which occurs at a frequency approaching 100%, homologous plasmid-chromosome recombination was originally reported to only be detected at frequencies between 10^{-6} and 10^{-3}. Nonhomologous plasmid-chromosome interactions are more frequent occurring at levels 10^{5}-fold to 10^{2} fold greater than comparable homologous insertion.

To overcome this low proportion of targeted recombination in murine ES cells, various strategies have been developed to detect or select rare homologous recombinants. One approach for detecting homologous alteration events uses the polymerase chain reaction (PCR) to screen pools of transformant cells for homologous insertion, followed by screening of individual clones. Alternatively, a positive genetic selection approach has been developed in which a marker gene is constructed which will only be active if homologous insertion occurs, allowing these recombinants to be selected directly. One of the most powerful approaches developed for selecting homologous recombinants is the positive-negative selection (PNS) method developed for genes for which no direct selection of the alteration exists. The PNS method is more efficient for targeting genes which are not expressed at high levels because the marker gene has its own promoter. Non-homologous recombinants are selected against by using the Herpes Simplex virus thymidine kinase (HSV-TK) gene and selecting against its nonhomologous insertion with effective herpes drugs such as gancyclovir (GANC) or (1-(2-deoxy-2-fluoro-B-D arabinofluranosyl)-5-iodou- racil, (FIAU). By this

counter selection, the number of homologous recombinants in the surviving transformants can be increased. Utilizing enhanced bacterial colonization/asthma-associated SNP containing TCRa nucleic acid as a targeted insertional cassette provides means to detect a successful insertion as visualized, for example, by acquisition of immunoreactivity to an antibody immunologically specific for the polypeptide encoded by the enhanced bacterial colonization/asthma-associated SNP containing TCRa nucleic acid and, therefore, facilitates screening/selection of ES cells with the desired genotype.

As used herein, a knock-in animal is one in which the endogenous murine gene, for example, has been replaced with human enhanced bacterial colonization/asthma-associated SNP containing TCRa gene of the invention. Such knock-in animals provide an ideal model system for studying enhanced bacterial colonization and the subsequent increased risk for the development of asthma.

As used herein, the expression of a enhanced bacterial colonization/asthma-associated SNP containing TCRa nucleic acid, fragment thereof, or an enhanced bacterial colonization/asthma-associated SNP containing TCRa fusion protein can be targeted in a "tissue specific manner" or "cell type specific manner" using a vector in which nucleic acid sequences encoding all or a portion the TCRa SNP(s) are operably linked to regulatory sequences (e.g., promoters and/or enhancers) that direct expression of the encoded protein in a particular tissue or cell type. Such regulatory elements may be used to advantage for both in vitro and in vivo applications. Promoters for directing tissue specific proteins are well known in the art and described herein.

The nucleic acid sequence encoding the enhanced bacterial colonization/asthma-associated SNP containing TCRa of the invention may be operably linked to a variety of different promoter sequences for expression in transgenic animals. Such promoters include, but are not limited to a prion gene promoter such as hamster and mouse Prion promoter (MoPrP), described in U.S. Pat. No. 5,877,399 and in Borchelt et al., Genet. Anal. 13(6) (1996) pages 159-163; a rat neuronal specific enolase promoter, described in U.S. Pat. Nos. 5,612,486, and 5,387,742; a platelet-derived growth factor B gene promoter, described in U.S. Pat. No. 5,811,633; a brain specific dystrophin promoter, described in U.S. Pat. No. 5,849,999; a Thy-1 promoter; a PGK promoter; a CMV promoter, the TCRa promoter or other promoters active in T cells from different species, for the expression of transgenes in T cells and other target cells.

Methods of use for the transgenic mice of the invention are also provided herein. Transgenic mice into which a nucleic acid containing the enhanced bacterial colonization/asthma-associated SNP or its encoded TCRa protein have been introduced are useful, for example, to develop screening methods to screen therapeutic agents to identify those capable of modulating bacterial colonization and the increased risk for the development of asthma.

PHARMACEUTICALS AND METHODS OF TREATMENT

The elucidation of the role played by the SNPs described herein in modulating the increased risk for asthma facilitates the development of pharmaceutical compositions useful for treatment and diagnosis of asthma. These compositions may comprise, in addition to one of the above substances, a pharmaceutically acceptable excipient, carrier, buffer, stabilizer or other materials well known to those skilled in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material may depend on the route of administration, e.g. oral, intravenous, cutaneous or subcutaneous, nasal, aerosolized, intramuscular, and intraperitoneal routes.

The invention includes a method of treating asthma in a mammal. An exemplary method entails administering to the mammal a pharmaceutically effective amount of siRNA which down regulates expression of the altered TCRa genes of the invention. Preferably, the mammal is a human. The term "patient" as used herein refers to a human.

Specific siRNA preparations directed at inhibiting the expression of TCRa, as well as delivery methods are provided as a novel therapy to treat asthma. SiRNA oligonucleotides directed to TCRa specifically hybridize with nucleic acids encoding TCRa and interfere with TCRa gene expression. The siRNA can be delivered to a patient in vivo either systemically or locally with carriers, as discussed below. The compositions of the invention may be used alone or in combination with other agents or genes encoding proteins to augment the efficacy of the compositions.

A "membrane permeant peptide sequence" refers to a peptide sequence which is able to facilitate penetration and entry of the TCRa inhibitor across the cell membrane. Exemplary peptides include with out limitation, the signal sequence from Karposi fibroblast growth factor exemplified herein, the HIV tat peptide (Vives et al., J Biol. Chem., 272:16010-

16017, 1997), Nontoxic membrane translocation peptide from protamine (Park et al., FASEB J. 19(11):1555-7, 2005), CHARIOT® delivery reagent (Active Motif; US Patent 6,841,535) and the antimicrobial peptide Buforin 2.

In one embodiment of the invention siRNAs are delivered for therapeutic benefit. There are several ways to administer the siRNA of the invention to in vivo to treat or ameliorate the symptoms of asthma including, but not limited to, naked siRNA delivery, siRNA conjugation and delivery, liposome carrier-mediated delivery, polymer carrier delivery, nanoparticle compositions, plasmid-based methods, and the use of viruses.

siRNA composition of the invention can comprise a delivery vehicle, including liposomes, for administration to a subject, carriers and diluents and their salts, and/or can be present in pharmaceutically acceptable formulations. This can be necessary to allow the siRNA to cross the cell membrane and escape degradation. Methods for the delivery of nucleic acid molecules are described in Akhtar et al., 1992, Trends Cell Bio., 2, 139; Delivery Strategies for Antisense Oligonucleotide Therapeutics, ed. Akhtar, 1995, Maurer et al., 1999, Mol. Membr. Biol., 16, 129-140; Hofland and Huang, 1999, Handb. Exp. Pharmacol., 137, 165-192; and Lee et al., 2000, ACS Symp. Ser., 752, 184-192; Beigelman et al., U.S. Pat. No. 6,395,713 and Sullivan et al., PCT WO 94/02595 further describe the general methods for delivery of nucleic acid molecules. These protocols can be utilized for the delivery of virtually any nucleic acid molecule.

The frequency of administration of the siRNA to a patient will also vary depending on several factors including, but not limited to, the type and severity of the asthma to be treated, the route of administration, the age and overall health of the individual, the nature of the siRNA, and the like. It is contemplated that the frequency of administration of the siRNA to the patient may vary from about once every few months to about once a month, to about once a week, to about once per day, to about several times daily.

Pharmaceutical compositions that are useful in the methods of the invention may be administered systemically in parenteral, oral solid and liquid formulations, ophthalmic, suppository, aerosol, topical or other similar formulations. In addition to the appropriate siRNA, these pharmaceutical compositions may contain pharmaceutically-acceptable carriers and other ingredients known to enhance and facilitate drug administration. Thus such

compositions may optionally contain other components, such as adjuvants, e.g., aqueous suspensions of aluminum and magnesium hydroxides, and/or other pharmaceutically acceptable carriers, such as saline. Other possible formulations, such as nanoparticles, liposomes, resealed erythrocytes, and immunologically based systems may also be used to administer the appropriate siRNA to a patient according to the methods of the invention. The use of nanoparticles to deliver siRNAs, as well as cell membrane permeable peptide carriers that can be used are described in Crombez et al., Biochemical Society Transactions v35:p44 (2007).

Methods of the invention directed to treating asthma involve the administration of siRNA specific for desired targets in a pharmaceutical composition. siRNA is administered to an individual as a pharmaceutical composition comprising siRNA and a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are well known in the art and include aqueous solutions such as physiologically buffered saline, other solvents or vehicles such as glycols, glycerol, oils such as olive oil or injectable organic esters.

A pharmaceutically acceptable carrier can contain physiologically acceptable compounds that act, for example, to stabilize the siRNA or increase the absorption of the agent. Such physiologically acceptable compounds include, for example, carbohydrates, such as glucose, sucrose or dextrans, antioxidants, such as ascorbic acid or glutathione, chelating agents, low molecular weight proteins or other stabilizers or excipients. One skilled in the art would know that the choice of a pharmaceutically acceptable carrier, including a physiologically acceptable compound, depends, for example, on the route of administration of the siRNA.

One skilled in the art appreciates that a pharmaceutical composition comprising siRNA can be administered to a subject by various routes including, for example, orally or parenterally, such as intravenously (i.v.), intramuscularly, subcutaneously, intraorbitally, intranasally, intracapsularly, intraperitoneally (i.p.), intracisternally, intra-tracheally (i.t.), or intra-articularly or by passive or facilitated absorption. The same routes of administration can be used other pharmaceutically useful compounds, for example, small molecules, nucleic acid molecules, peptides, antibodies and polypeptides as discussed hereinabove.

A pharmaceutical composition comprising siRNA inhibitor also can be incorporated, if desired, into liposomes, microspheres, microbubbles, or other polymer matrices (Gregoriadis, Liposome Technology, Vols. I to III, 2nd ed., CRC Press, Boca Raton Fla. (1993)). Liposomes, for example, which consist of phospholipids or other lipids, are nontoxic, physiologically acceptable and metabolizable carriers that are relatively simple to make and administer.

The pharmaceutical preparation comprises a siRNA targeting a nucleic acid target of interest or an expression vector encoding the same. Such pharmaceutical preparations can be administered to a patient for treating asthma, for example..

Expression vectors for the expression of siRNA molecules preferably employ a strong promoter which may be constitutive or regulated. Such promoters are well known in the art and include, but are not limited to, RNA polymerase II promoters, the T7 RNA polymerase promoter, and the RNA polymerase III promoters U6 and H1 (see, e.g., Myslinski et al. (2001) Nucl. Acids Res., 29:2502 09).

A formulated siRNA composition can be a composition comprising one or more siRNA molecules or a vector encoding one or more siRNA molecules independently or in combination with a cationic lipid, a neutral lipid, and/or a polyethyleneglycol-diacylglycerol (PEG-DAG) or PEG-cholesterol (PEG-Chol) conjugate. Non-limiting examples of expression vectors are described in Paul et al., 2002, Nature Biotechnology, 19, 505; Miyagishi and Taira, 2002, Nature Biotechnology, 19, 497; Lee et al., 2002, Nature Biotechnology, 19, 500-505.

A lipid nanoparticle composition is a composition comprising one or more biologically active molecules independently or in combination with a cationic lipid, a neutral lipid, and/or a polyethyleneglycol-diacylglycerol (i.e., polyethyleneglycol diacylglycerol (PEG-DAG), PEG-cholesterol, or PEG-DMB) conjugate. In one embodiment, the biologically active molecule is encapsulated in the lipid nanoparticle as a result of the process of providing and aqueous solution comprising a biologically active molecule of the invention (i.e., siRNA), providing an organic solution comprising lipid nanoparticle, mixing the two solutions, incubating the solutions, dilution, ultrafiltration, resulting in concentrations suitable to produce nanoparticle compositions.

Nucleic acid molecules can be administered to cells by incorporation into other vehicles, such as biodegradable polymers, hydrogels, cyclodextrins. (see for example Gonzalez et al., 1999, Bioconjugate Chem., 10, 1068-1074; Wang et al., International PCT publication Nos. WO 03/47518 and WO 03/46185), poly(lactic-co-glycolic)acid (PLGA) and PLCA microspheres (see for example U.S. Pat. No. 6,447,796 and US Patent Application Publication No. US 2002130430), biodegradable nanocapsules, and bioadhesive microspheres, or by proteinaceous vectors (O'Hare and Normand, International PCT Publication No. WO 00/53722)

Cationic lipids and polymers are two classes of non-viral siRNA delivery which can form complexes with negatively charged siRNA. The self-assembly PEG-ylated polycation polyethylenimine (PEI) has also been used to condense and protect siRNAs (Schiffelers et al., 2004, Nuc. Acids Res. 32: 141-110). The siRNA complex can be condensed into a nanoparticle to allow efficient uptake of the siRNA through endocytosis. Also, the nucleic acid-condensing property of protamine has been combined with specific antibodies to deliver siRNAs and can be used in the invention (Song et al., 2005, Nat Biotech. 23:709-717).

In order to treat an individual having athma, to alleviate a sign or symptom of the disease, siRNA should be administered in an effective dose. The total treatment dose can be administered to a subject as a single dose or can be administered using a fractionated treatment protocol, in which multiple doses are administered over a more prolonged period of time, for example, over the period of a day to allow administration of a daily dosage or over a longer period of time to administer a dose over a desired period of time. One skilled in the art would know that the amount of siRNA required to obtain an effective dose in a subject depends on many factors, including the age, weight and general health of the subject, as well as the route of administration and the number of treatments to be administered. In view of these factors, the skilled artisan would adjust the particular dose so as to obtain an effective dose for treating an individual having asthma.

The effective dose of siRNA will depend on the mode of administration, and the weight of the individual being treated. The dosages described herein are generally those for an average adult but can be adjusted for the treatment of children. The dose will generally range from about 0.001 mg to about 1000 mg.

The concentration of siRNA in a particular formulation will depend on the mode and frequency of administration. A given daily dosage can be administered in a single dose or in multiple doses so long as the siRNA concentration in the formulation results in the desired daily dosage. One skilled in the art can adjust the amount of siRNA in the formulation to allow administration of a single dose or in multiple doses that provide the desired concentration of siRNA over a given period of time.

In an individual suffering from asthma, in particular a more severe form of the disease, administration of siRNA can be particularly useful when administered in combination, for example, with a conventional agent for treating such a disease. The skilled artisan would administer siRNA, alone or in combination and would monitor the effectiveness of such treatment using routine methods such as pulmonary function determination, radiologic, immunologic or, where indicated, histopathologic methods. Other conventional agents for the treatment of asthma include steroid or administration of other agents that alleviate the symptoms underlying the disease.

Administration of the pharmaceutical preparation is preferably in an "effective amount" this being sufficient to show benefit to the individual. This amount prevents, alleviates, abates, or otherwise reduces the severity of asthma symptoms in a patient.

The pharmaceutical preparation is formulated in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form, as used herein, refers to a physically discrete unit of the pharmaceutical preparation appropriate for the patient undergoing treatment. Each dosage should contain a quantity of active ingredient calculated to produce the desired effect in association with the selected pharmaceutical carrier. Procedures for determining the appropriate dosage unit are well known to those skilled in the art.

Dosage units may be proportionately increased or decreased based on the weight of the patient. Appropriate concentrations for alleviation of a particular pathological condition may be determined by dosage concentration curve calculations, as known in the art.

The following example is provided to illustrate certain embodiments of the invention. It is not intended to limit the invention in any way.

EXAMPLE I

Genetic variation in the T cell receptor alpha gene (TCRa) predisposes to bacterial colonization and associates with asthma.

Novel sequences variants of TCRa have been identified which correlate with increased bacterial colonization of airways in infants resulting in an increased risk for development of asthma later in life.

The materials and methods set forth below are provided to facilitate the practice of the present invention.

SNPs

32 SNPs located in the chromosome 14 spanning a 57.9 Kb region (22024524 — 22082384) were included in this study.

Table 1A: Description of the SNPs included in the analysis

Snp	Chr	Pos
rs996165	Chr14	22024524
rs3811215	Chr14	22024617
rs227869	Chr14	22026550
rs10083416	Chr14	22026757
rs227870	Chr14	22027011
rs1483971	Chr14	22027328
rs3751467	Chr14	22029432
rs2128997	Chr14	22030942
rs226985	Chr14	22033486
rs2170810	Chr14	22042582
rs6572493	Chr14	22042795
rs7147589	Chr14	22042826
rs17794137	Chr14	22043207
rs3811207	Chr14	22044096
rs1010116	Chr14	22047509

rs226998	Chr14	22048356
rs226999	Chr14	22048417
rs1483973	Chr14	22048600
rs1483974	Chr14	22048615
rs227000	Chr14	22049188
rs12887424	Chr14	22050248
rs170544	Chr14	22053063
rs3764159	Chr14	22054426
rs227028	Chr14	22054846
rs227011	Chr14	22060602
rs227019	Chr14	22064646
rs12587781	Chr14	22069457
rs227003	Chr14	22069902
rs1154153	Chr14	22071387
rs227008	Chr14	22074550
rs1263638	Chr14	22079839
rs7147412	Chr14	22082384

Statistical methods

The association between each SNP and the continuous endpoints ($FEV_{0.5}$, PD_{15} TcO_2, PD_{20} metacholine at age 6.5, FEV_1 at age 6, eosinophilocytes at age 6, FVC at age 6 and MMEF at age 6) was analyzed by linear regression. The association between each SNP and the binary endpoints (asthma status at age 6, bacterial colonization) was analyzed by logistic regression.

The association between each SNP and the censored endpoints (JJJ, J45, exacerbation, eczema, asthma composite score) was estimated using univariate Cox-proportional hazards regression. The significance of the association was given by the score test *p*-values.

The association between each SNP and the endpoints with longitudinal measurements (sRaw profiles, exacerbation incidences) was estimated using respectively linear mixed effect models and generalized estimating equations (GEE). The significance of the association was respectively given by the ANOVA test and Wald-test *p*-values.

The association between each SNP and categorical variables (wheezy categorization) was tested using chi-square tests of independence on contingency tables.

All analyses were performed using the R statistical software (version 2.7.0) — including the following packages: survival, SNPassoc, genetics, LDheatmap, nlme, geepack.

Genetics models

Five genetic models of inheritance were investigated (Table 2).

Table 2: Coding of the genetics data according to five models of inheritance (A stands for the most frequent allele and B for the rare allele).

	A/A	A/B	B/B
codominant	A/A	A/B	B/B
dominant	A/A	A/B-B/B	A/B-B/B
recessive	A/A-A/B	A/A-A/B	B/B
overdominant	A/A-B/B	A/B	A/A-B/B
log-additive	0	1	2

Graphical representation and result interpretation

The multi-SNPs association figures present the -log(*p*-values) of the association between the different endpoints and the children genotypes. Univariate tests are performed SNP by SNP. The horizontal lines represent the nominal (simple dashed-line) and Bonferroni corrected (bold dashed-line) *p*-value threshold.

Tables present the estimates (continuous variables), hazard-ratios (binary and censored data) and chi-square contributions (categorical data) for each genotype together with the levels of significance (*p*-values) according to the codominant, dominant and recessive genetic models.

Results

Descriptive statistics

Table 3 describes for each SNP the alleles, the major allele frequency, *p*-values of the Hardy-Weinberg equilibrium testing and the percentage of missing values.

The tests of the Hardy-Weinberg equilibrium were non-significant for all SNPs but *rs17794137* for which no estimate was available since only 2 out of 3 genotypes were represented in our cohort.

Table 3: SNPs overview

5

	Alleles	major.allele.freq	HWE	missing
rs996165	T/C	50.50	0.41	0.00
rs3811215	C/A	77.00	0.47	0.00
rs227869	A/G	56.90	0.75	2.70
rs10083416	G/A	70.70	1.00	0.00
rs227870	G/A	71.50	0.25	0.00
rs1483971	G/T	61.20	0.19	0.00
rs3751467	A/G	97.00	0.32	0.30
rs2128997	A/G	72.40	0.70	0.00
rs226985	T/C	65.70	0.02	0.00
rs2170810	A/G	65.40	0.21	0.30
rs6572493	G/A	58.50	0.45	0.00
rs7147589	G/A	76.20	1.00	0.00
rs17794137	G/A	95.00		0.00
rs3811207	G/T	70.10	0.38	4.80
rs1010116	G/A	91.40	0.01	0.00
rs226998	G/A	61.20	0.91	0.00
rs226999	G/A	87.40	0.16	0.00
rs1483973	C/T	51.70	0.68	0.00
rs1483974	C/A	84.80	0.03	0.00
rs227000	G/A	72.80	1.00	0.00
rs12887424	A/G	54.00	0.21	0.00
rs170544	C/T	90.20	0.56	0.00
rs3764159	A/C	85.30	0.68	0.00
rs227028	T/G	58.30	1.00	0.00
rs227011	G/A	91.00	0.75	0.00
rs227019	C/T	67.90	0.48	0.30
rs12587781	T/G	86.30	1.00	0.30
rs227003	G/A	66.00	0.25	0.30
rs1154153	A/G	86.20	1.00	0.00
rs227008	C/A	57.30	0.92	1.10
rs1263638	C/T	61.40	0.27	0.00
rs7147412	T/C	64.20	0.43	0.00

Figure 1 displays a pairwise linkage desequilibrium heatmap based on measures of allelic correlation (R^2). This correlation coefficient is defined as follows:

$$r=-D/\sqrt{p(A)p(a)p(B)p(b)}$$

with $D=p(AB)-p(A)p(B)$ and $p(A)$ defined as the observed probability of allele 'A' for marker 1; $p(a)=1-p(A)$ is defined as the observed probability of allele 'a' for marker 1; $p(B)$ is defined as the observed probability of allele 'B' for marker 2; $p(b)=1-p(B)$ is defined as the observed probability of allele 'b' for marker 2; p(AB) is defined as the probability of the marker allele pair 'AB'.

One can detect 2 blocks of linkage desequilibrium at each extremity of the region screened in this analysis.

Multiple SNPs association study with various clinical endpoints

$FEV_{0.5}$

Associations:

$FEV_{0.5}$ measures at birth were calibrated for lifespan, weight and BMI. The association between log-transformed $FEV_{0.5}$ values and genotypes was tested. Figure 2 shows the association of multiple SNPs with $FEV_{0.5}$.

Estimates: -

Table 4: $FEV_{0.5}$ estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	65.68	67.68	67.27	0.51	0.25	0.86
rs3811215	A/C	66.35	67.09	74.16	0.05	0.23	0.02
rs227869	A/G	66.76	65.82	70.90	0.04	0.76	0.01
rs10083416	A/G	67.71	66.71	64.96	0.53	0.36	0.36
rs227870	A/G	67.20	67.39	64.80	0.60	0.83	0.32
rs1483971	G/T	67.05	65.42	71.48	0.01	1.00	0.01
rs3751467	A/G	67.05	67.30		0.94		
rs2128997	A/G	67.55	66.25	67.63	0.67	0.46	0.81

rs226985	C/T	66.63	68.20	65.24	0.35	0.59	0.30
rs2170810	A/G	67.77	66.52	66.12	0.64	0.35	0.63
rs6572493	A/G	67.10	67.52	65.75	0.67	0.96	0.39
rs7147589	A/G	67.10	66.90	67.44	0.98	0.93	0.89
rs17794137	A/G	66.99	67.58	66.99	0.81		
rs3811207	G/T	67.82	66.72	63.88	0.36	0.30	0.21
rs1010116	A/G	67.03	66.91	68.75	0.95	0.96	0.74
rs226998	A/G	67.28	66.75	67.43	0.92	0.80	0.83
rs226999	A/G	66.62	68.86	66.10	0.46	0.27	0.83
rs1483973	C/T	65.57	67.59	67.51	0.47	0.22	0.73
rs1483974	A/C	67.19	66.62	69.13	0.91	0.75	0.80
rs227000	A/G	66.69	68.56	61.62	0.04	0.60	0.03
rs12887424	A/G	64.22	68.33	67.63	0.04	0.01	0.69
rs170544	C/T	66.30	70.55	64.83	0.08	0.03	0.82
rs3764159	A/C	67.50	66.80	56.84	0.04	0.30	0.01
rs227028	G/T	65.61	68.40	66.03	0.18	0.15	0.52
rs227011	A/G	66.44	70.25	64.83	0.15	0.06	0.82
rs227019	C/T	66.55	67.86	65.56	0.55	0.53	0.50
rs12587781	G/T	67.43	66.63	56.77	0.09	0.33	0.03
rs227003	A/G	67.06	67.39	65.62	0.73	0.94	0.45
rs1154153	A/G	67.43	66.76	56.77	0.09	0.37	0.03
rs227008	A/C	65.82	68.05	66.18	0.34	0.26	0.60
rs1263638	C/T	65.52	68.19	66.88	0.23	0.11	0.93
rs7147412	C/T	67.52	67.07	65.58	0.68	0.58	0.41

PD_{15} TcO_2

Associations:

PD_{15} TcO_2 measures at birth were calibrated for lifespan and weight. The association between log-transformed PD_{15} TcO_2 values and genotypes was tested. Figure 3 shows the associations of multiple SNPs with PD_{15} TcO_2.

Estimates: -

Table 5: PD_{15} TcO_2 estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	0.36	0.26	0.39	0.17	0.38	0.22

rs3811215	A/C	0.30	0.36	0.23	0.48	0.56	0.40
rs227869	A/G	0.35	0.30	0.29	0.72	0.42	0.70
rs10083416	A/G	0.29	0.35	0.34	0.56	0.29	0.84
rs227870	A/G	0.33	0.30	0.34	0.88	0.77	0.79
rs1483971	G/T	0.36	0.28	0.32	0.54	0.31	0.97
rs3751467	A/G	0.32	0.28		0.73		
rs2128997	A/G	0.30	0.34	0.32	0.87	0.61	0.98
rs226985	C/T	0.30	0.33	0.35	0.84	0.58	0.68
rs2170810	A/G	0.28	0.35	0.36	0.53	0.26	0.64
rs6572493	A/G	0.27	0.34	0.36	0.39	0.17	0.51
rs7147589	A/G	0.31	0.35	0.22	0.57	0.78	0.37
rs17794137	A/G	0.31	0.37	0.31	0.55		
rs3811207	G/T	0.29	0.33	0.41	0.58	0.36	0.45
rs1010116	A/G	0.32	0.32	0.28	0.98	0.95	0.85
rs226998	A/G	0.29	0.33	0.35	0.80	0.54	0.64
rs226999	A/G	0.30	0.36	0.45	0.65	0.39	0.56
rs1483973	C/T	0.29	0.33	0.32	0.85	0.57	0.91
rs1483974	A/C	0.31	0.33	0.18	0.88	0.90	0.64
rs227000	A/G	0.32	0.34	0.21	0.48	0.88	0.24
rs12887424	A/G	0.27	0.31	0.39	0.44	0.35	0.26
rs170544	C/T	0.29	0.46	0.21	0.15	0.07	0.74
rs3764159	A/C	0.35	0.23	0.49	0.15	0.13	0.44
rs227028	G/T	0.28	0.33	0.37	0.54	0.30	0.47
rs227011	A/G	0.29	0.53	0.21	0.05	0.02	0.74
rs227019	C/T	0.31	0.32	0.32	0.99	0.88	0.95
rs12587781	G/T	0.35	0.22	0.51	0.08	0.06	0.47
rs227003	A/G	0.35	0.31	0.22	0.29	0.27	0.15
rs1154153	A/G	0.35	0.22	0.51	0.08	0.06	0.47
rs227008	A/C	0.27	0.34	0.38	0.43	0.22	0.44
rs1263638	C/T	0.28	0.33	0.37	0.65	0.39	0.54
rs7147412	C/T	0.35	0.32	0.23	0.34	0.43	0.15

PD_{20} metacholine at age 6.5

Associations:

The association between the log-transformed PD_{20} measures after metacholine challenge tests at age 6.5 years and genotypes was tested. Figure 4 shows multiple SNPs and their association with PD_{20} metacholine.

Estimates: -

Table 6: PD_{20} metacholine estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	1.34	1.46	1.35	0.88	0.76	0.79
rs3811215	A/C	1.43	1.25	1.90	0.42	0.74	0.28
rs227869	A/G	1.54	1.29	1.47	0.60	0.42	0.77
rs10083416	A/G	1.28	1.59	1.30	0.42	0.26	0.79
rs227870	A/G	1.41	1.46	1.11	0.66	0.91	0.37
rs1483971	G/T	1.53	1.21	1.59	0.31	0.38	0.42
rs3751467	A/G	1.40	1.27		0.77		
rs2128997	A/G	1.30	1.65	1.07	0.24	0.34	0.32
rs226985	C/T	1.52	1.43	0.99	0.20	0.33	0.08
rs2170810	A/G	1.59	1.39	0.96	0.14	0.18	0.06
rs6572493	A/G	1.67	1.29	1.19	0.22	0.09	0.32
rs7147589	A/G	1.24	1.74	1.11	0.11	0.09	0.50
rs17794137	A/G	1.35	1.94	1.35	0.18		
rs3811207	G/T	1.58	1.37	0.98	0.29	0.24	0.18
rs1010116	A/G	1.45	1.22	0.64	0.36	0.29	0.23
rs226998	A/G	1.67	1.33	0.98	0.09	0.07	0.08
rs226999	A/G	1.46	1.28	0.61	0.21	0.27	0.10
rs1483973	C/T	1.62	1.34	1.27	0.52	0.26	0.56
rs1483974	A/C	1.36	1.50	1.41	0.86	0.58	0.99
rs227000	A/G	1.41	1.31	1.72	0.67	0.89	0.43
rs12887424	A/G	1.33	1.62	1.00	0.07	0.69	0.04
rs170544	C/T	1.44	1.24	1.00	0.75	0.46	0.77
rs3764159	A/C	1.33	1.63	1.30	0.53	0.30	0.88
rs227028	G/T	1.56	1.22	1.59	0.28	0.31	0.43
rs227011	A/G	1.41	1.33	1.00	0.92	0.75	0.77
rs227019	C/T	1.46	1.42	1.08	0.55	0.63	0.27
rs12587781	G/T	1.31	1.72	1.29	0.35	0.17	0.90
rs227003	A/G	1.41	1.44	1.26	0.87	0.93	0.60
rs1154153	A/G	1.31	1.72	1.29	0.35	0.17	0.90
rs227008	A/C	1.56	1.26	1.50	0.47	0.36	0.67
rs1263638	C/T	1.53	1.25	1.68	0.33	0.40	0.36
rs7147412	C/T	1.45	1.40	1.25	0.83	0.68	0.56

FEV_1 at age 6

Associations:

The association between log-transformed values of FEV_1 at age 6 and genotypes was tested after adjustment for height and sex. Figure 5 shows multiple SNPs and their association with FEV_1.Estimates: -

Table 7: FEV_1 at age 6 years estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	0.21	0.22	0.22	0.88	0.67	0.68
rs3811215	A/C	0.21	0.22	0.21	0.66	0.57	0.62
rs227869	A/G	0.22	0.22	0.22	0.77	0.54	0.57
rs10083416	A/G	0.21	0.21	0.22	0.78	0.97	0.49
rs227870	A/G	0.22	0.22	0.21	0.52	0.56	0.26
rs1483971	G/T	0.22	0.21	0.22	0.68	0.72	0.53
rs3751467	A/G	0.21	0.22	1.01	0.20		
rs2128997	A/G	0.21	0.22	0.22	0.78	0.50	0.72
rs226985	C/T	0.22	0.21	0.21	0.20	0.17	0.11
rs2170810	A/G	0.22	0.21	0.21	0.03	0.02	0.07
rs6572493	A/G	0.22	0.22	0.21	0.65	0.42	0.49
rs7147589	A/G	0.21	0.22	0.22	0.31	0.13	0.87
rs17794137	A/G	0.21	0.21	1.01	0.79		
rs3811207	G/T	0.23	0.23	0.22	0.02	0.02	0.04
rs1010116	A/G	0.21	0.22	0.24	0.07	0.17	0.03
rs226998	A/G	0.22	0.22	0.22	0.80	0.51	0.90
rs226999	A/G	0.22	0.22	0.23	0.21	0.53	0.08
rs1483973	C/T	0.21	0.21	0.21	0.70	0.91	0.45
rs1483974	A/C	0.21	0.21	0.22	0.88	0.78	0.65
rs227000	A/G	0.21	0.21	0.21	0.85	0.60	0.74
rs12887424	A/G	0.23	0.24	0.24	0.03	0.01	0.38
rs170544	C/T	0.21	0.22	0.23	0.44	0.37	0.29
rs3764159	A/C	0.22	0.22	0.19	0.03	0.39	0.01
rs227028	G/T	0.21	0.22	0.22	0.75	0.45	0.76
rs227011	A/G	0.21	0.22	0.24	0.43	0.35	0.29
rs227019	C/T	0.21	0.22	0.21	0.20	0.13	0.65
rs12587781	G/T	0.22	0.22	0.19	0.09	0.46	0.03
rs227003	A/G	0.21	0.21	0.21	0.87	0.93	0.60
rs1154153	A/G	0.22	0.22	0.19	0.09	0.52	0.03
rs227008	A/C	0.21	0.21	0.21	0.93	0.71	0.98
rs1263638	C/T	0.21	0.22	0.22	0.56	0.29	0.95

rs7147412	C/T	0.21	0.22	0.21	0.79	0.94	0.51

FEV_1/FVC at age 6

Associations:

The association between FEV_1/FVC at age 6 and genotypes was tested. Figure 6 shows multiple SNPs and their association with FEV_1/FVC.

Estimates: -

Table 8: FEV_1/FVC at age 6 years estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	2.57	2.58	2.58	0.79	0.49	0.80
rs3811215	A/C	2.59	2.56	2.61	0.22	0.28	0.35
rs227869	A/G	2.61	2.56	2.59	0.06	0.04	0.60
rs10083416	A/G	2.57	2.60	2.59	0.30	0.12	0.85
rs227870	A/G	2.59	2.56	2.61	0.24	0.40	0.27
rs1483971	G/T	2.61	2.55	2.59	0.01	0.01	0.55
rs3751467	A/G	2.58	2.63		0.15		
rs2128997	A/G	2.56	2.60	2.60	0.14	0.05	0.54
rs226985	C/T	2.59	2.58	2.57	0.66	0.42	0.50
rs2170810	A/G	2.59	2.58	2.56	0.56	0.61	0.29
rs6572493	A/G	2.59	2.58	2.57	0.79	0.53	0.63
rs7147589	A/G	2.57	2.59	2.57	0.51	0.30	0.83
rs17794137	A/G	2.58	2.57		0.61		
rs3811207	G/T	2.59	2.57	2.59	0.70	0.44	0.88
rs1010116	A/G	2.58	2.60	2.56	0.52	0.37	0.71
rs226998	A/G	2.59	2.57	2.59	0.23	0.21	0.49
rs226999	A/G	2.58	2.58	2.58	0.97	0.84	0.96
rs1483973	C/T	2.59	2.58	2.58	0.92	0.71	0.96
rs1483974	A/C	2.58	2.58	2.64	0.77	0.84	0.47
rs227000	A/G	2.59	2.57	2.59	0.42	0.25	0.78
rs12887424	A/G	2.59	2.57	2.58	0.61	0.32	0.83
rs170544	C/T	2.58	2.57	2.63	0.75	0.66	0.60
rs3764159	A/C	2.58	2.59	2.61	0.80	0.62	0.58
rs227028	G/T	2.61	2.56	2.57	0.07	0.02	0.74
rs227011	A/G	2.58	2.58	2.63	0.86	0.85	0.60
rs227019	C/T	2.59	2.57	2.57	0.54	0.27	0.65
rs12587781	G/T	2.58	2.60	2.59	0.60	0.31	0.93
rs227003	A/G	2.57	2.58	2.60	0.62	0.47	0.39

rs1154153	A/G	2.58	2.59	2.59	0.64	0.34	0.93
rs227008	A/C	2.61	2.56	2.58	0.06	0.02	0.78
rs1263638	C/T	2.60	2.56	2.60	0.03	0.04	0.34
rs7147412	C/T	2.58	2.58	2.60	0.75	0.73	0.46

MMEF at age 6

Associations:

The association between MMEF at age 6 and genotypes was tested after adjustment for height and sex. Figure 7 shows multiple SNPs and their association with MMEF.

Estimates: -

Table 9: MMEF at age 6 years estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	0.37	0.38	0.38	0.91	0.74	0.69
rs3811215	A/C	0.39	0.35	0.38	0.18	0.09	0.88
rs227869	A/G	0.37	0.32	0.33	0.05	0.02	0.80
rs10083416	A/G	0.37	0.39	0.38	0.39	0.18	0.88
rs227870	A/G	0.36	0.36	0.39	0.67	0.75	0.37
rs1483971	G/T	0.40	0.33	0.36	0.00	0.00	0.95
rs3751467	A/G	0.37	0.47	1.02	0.02		
rs2128997	A/G	0.37	0.40	0.41	0.33	0.14	0.52
rs226985	C/T	0.36	0.36	0.34	0.82	0.87	0.53
rs2170810	A/G	0.36	0.36	0.34	0.69	0.52	0.44
rs6572493	A/G	0.37	0.37	0.38	0.93	0.81	0.71
rs7147589	A/G	0.37	0.39	0.38	0.71	0.41	0.95
rs17794137	A/G	0.37	0.36	1.02	0.62		
rs3811207	G/T	0.47	0.45	0.44	0.60	0.31	0.76
rs1010116	A/G	0.37	0.39	0.45	0.43	0.30	0.28
rs226998	A/G	0.40	0.37	0.43	0.07	0.38	0.09
rs226999	A/G	0.38	0.37	0.45	0.53	0.89	0.27
rs1483973	C/T	0.36	0.38	0.36	0.72	0.69	0.59
rs1483974	A/C	0.37	0.38	0.42	0.72	0.47	0.62
rs227000	A/G	0.39	0.38	0.37	0.70	0.40	0.73
rs12887424	A/G	0.33	0.30	0.31	0.34	0.14	0.78
rs170544	C/T	0.35	0.35	0.48	0.59	0.93	0.30
rs3764159	A/C	0.37	0.38	0.33	0.57	0.52	0.51
rs227028	G/T	0.36	0.32	0.32	0.16	0.06	0.40
rs227011	A/G	0.36	0.37	0.49	0.57	0.64	0.30
rs227019	C/T	0.32	0.29	0.26	0.15	0.11	0.13

rs12587781	G/T	0.35	0.38	0.32	0.44	0.29	0.61
rs227003	A/G	0.33	0.33	0.36	0.39	0.64	0.17
rs1154153	A/G	0.35	0.37	0.32	0.42	0.28	0.61
rs227008	A/C	0.33	0.30	0.29	0.09	0.03	0.25
rs1263638	C/T	0.37	0.33	0.34	0.16	0.06	0.82
rs7147412	C/T	0.35	0.35	0.39	0.45	0.85	0.21

sRaw profiles

Associations:

The association between log-transformed values of sRaw profiles and genotypes was tested. Figure 8 shows multiple SNPs and their association with sRaw profiles.

2.2.8 Eosinophilocytes at age 6

Associations:

The association between log-transformed values of eosinophilocytes at age 6 and genotypes was tested. Figure 9 shows multiple SNPs and their association with eosinophilocytes at age 6 years

Estimates: -

Table 10: Eosinophilocytes at age 6 years estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	0.31	0.33	0.33	0.89	0.66	0.74
rs3811215	A/C	0.32	0.32	0.35	0.85	0.88	0.57
rs227869	A/G	0.33	0.32	0.32	0.96	0.79	0.93
rs10083416	A/G	0.32	0.33	0.30	0.90	0.99	0.66
rs227870	A/G	0.33	0.31	0.36	0.74	0.84	0.52
rs1483971	G/T	0.33	0.31	0.33	0.80	0.69	0.73
rs3751467	A/G	0.32	0.31		0.82		
rs2128997	A/G	0.32	0.32	0.34	0.95	0.87	0.77
rs226985	C/T	0.33	0.32	0.31	0.86	0.71	0.61
rs2170810	A/G	0.33	0.32	0.34	0.90	0.81	0.79
rs6572493	A/G	0.33	0.33	0.30	0.82	0.76	0.53
rs7147589	A/G	0.33	0.32	0.31	0.93	0.76	0.76
rs17794137	A/G	0.32	0.39		0.17		
rs3811207	G/T	0.32	0.32	0.31	0.96	0.98	0.77
rs1010116	A/G	0.32	0.36	0.38	0.47	0.22	0.60
rs226998	A/G	0.31	0.31	0.41	0.07	0.33	0.02

rs226999	A/G	0.32	0.33	0.44	0.56	0.57	0.30
rs1483973	C/T	0.32	0.33	0.32	0.98	0.96	0.85
rs1483974	A/C	0.32	0.32	0.31	0.99	0.89	0.96
rs227000	A/G	0.32	0.35	0.25	0.23	0.69	0.14
rs12887424	A/G	0.32	0.33	0.32	0.97	0.98	0.81
rs170544	C/T	0.33	0.30	0.57	0.46	0.69	0.27
rs3764159	A/C	0.34	0.28	0.58	0.05	0.13	0.11
rs227028	G/T	0.31	0.32	0.34	0.82	0.67	0.56
rs227011	A/G	0.33	0.30	0.57	0.44	0.66	0.27
rs227019	C/T	0.33	0.32	0.31	0.88	0.64	0.75
rs12587781	G/T	0.34	0.28	0.58	0.13	0.14	0.26
rs227003	A/G	0.31	0.35	0.30	0.36	0.28	0.61
rs1154153	A/G	0.34	0.28	0.58	0.16	0.17	0.26
rs227008	A/C	0.32	0.32	0.33	0.95	0.90	0.76
rs1263638	C/T	0.33	0.32	0.34	0.88	0.75	0.77
rs7147412	C/T	0.30	0.36	0.29	0.17	0.24	0.32

Asthma status at age 6

Associations:

The association between the asthma status at age 6 and the genotypes was tested. Figure 10 shows multiple SNPs and their association with the asthma status at age 6 years

Estimates: -

Table 11: Asthma status at age 6: hazard-ratio estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	1.00	1.01	0.85	0.91	0.90	0.67
rs3811215	A/C	1.00	0.84	0.00	0.18	0.29	0.09
rs227869	A/G	1.00	1.85	0.86	0.14	0.24	0.25
rs10083416	A/G	1.00	0.75	0.23	0.20	0.22	0.11
rs227870	A/G	1.00	1.11	2.13	0.38	0.45	0.17
rs1483971	G/T	1.00	1.21	0.74	0.61	0.84	0.40
rs3751467	A/G	1.00	2.04		0.25		
rs2128997	A/G	1.00	0.72	0.51	0.47	0.26	0.44
rs226985	C/T	1.00	0.65	1.99	0.07	0.91	0.04
rs2170810	A/G	1.00	1.18	2.38	0.19	0.29	0.07
rs6572493	A/G	1.00	1.72	3.23	0.05	0.05	0.03
rs7147589	A/G	1.00	0.88	0.00	0.28	0.42	0.14

rs17794137	A/G	1.00	0.48	0.48	0.28		
rs3811207	G/T	1.00	1.15	2.84	0.18	0.36	0.07
rs1010116	A/G	1.00	1.12	0.00	0.84	0.92	1.00
rs226998	A/G	1.00	0.85	1.49	0.50	0.97	0.27
rs226999	A/G	1.00	1.59	1.05	0.50	0.27	0.95
rs1483973	C/T	1.00	1.64	1.12	0.42	0.34	0.58
rs1483974	A/C	1.00	1.35	0.00	0.63	0.46	1.00
rs227000	A/G	1.00	1.27	2.24	0.37	0.31	0.22
rs12887424	A/G	1.00	0.54	0.98	0.20	0.23	0.38
rs170544	C/T	1.00	1.10	7.84	0.40	0.63	0.18
rs3764159	A/C	1.00	0.63	0.97	0.54	0.30	0.95
rs227028	G/T	1.00	0.73	0.67	0.62	0.33	0.64
rs227011	A/G	1.00	1.29	8.06	0.35	0.40	0.18
rs227019	C/T	1.00	0.55	1.10	0.20	0.18	0.49
rs12587781	G/T	1.00	0.54	1.34	0.36	0.21	0.72
rs227003	A/G	1.00	1.17	1.74	0.55	0.45	0.31
rs1154153	A/G	1.00	0.54	1.34	0.34	0.20	0.72
rs227008	A/C	1.00	0.78	0.66	0.67	0.41	0.55
rs1263638	C/T	1.00	0.88	1.01	0.92	0.78	0.86
rs7147412	C/T	1.00	1.00	1.78	0.44	0.64	0.20

Wheezy categories

Associations:

The association between wheezy categorization and genotypes was tested using chi-square tests. Contingency tables were derived from the cross-tabulation tables between wheezy categorization and genotypes for the co-dominant, dominant, recessive and overdominant model. The allele count (frequent vs. rare allele) was used in the log-additive model. Figure 11 shows multiple SNPs and their association with wheezy categorization.

Estimates: -

Table 12: Wheezy categorization: chi-square contributions and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	0.73	0.69	1.41	0.83	0.82	0.61
rs3811215	A/C	0.43	1.36	2.80	0.61	0.43	0.80
rs227869	A/G	2.54	3.38	1.39	0.30	0.29	0.64
rs10083416	A/G	2.79	3.43	4.79	0.09	0.15	0.13

rs227870	A/G	0.64	1.67	3.41	0.46	0.30	0.73
rs1483971	G/T	0.47	1.32	1.34	0.80	0.85	0.66
rs3751467	A/G	0.75	10.19		0.02		
rs2128997	A/G	3.09	5.17	1.38	0.14	0.08	0.70
rs226985	C/T	1.89	1.63	5.80	0.15	0.07	0.33
rs2170810	A/G	1.71	0.54	6.06	0.21	0.38	0.07
rs6572493	A/G	3.38	2.22	5.60	0.08	0.07	0.16
rs7147589	A/G	3.54	5.86	2.01	0.08	0.58	0.04
rs17794137	A/G	0.17	1.68		0.60	0.60	0.60
rs3811207	G/T	0.68	0.84	6.84	0.21	0.72	0.05
rs1010116	A/G	0.33	2.43	2.69	0.47	0.44	0.58
rs226998	A/G	0.52	2.70	3.78	0.32	0.21	0.85
rs226999	A/G	0.15	1.50	2.16	0.71	0.57	0.89
rs1483973	C/T	0.47	1.89	2.44	0.58	0.89	0.37
rs1483974	A/C	1.32	3.57	0.90	0.41	1.00	0.23
rs227000	A/G	0.98	0.26	2.61	0.70	0.42	0.57
rs12887424	A/G	0.94	1.89	1.29	0.67	0.73	0.68
rs170544	C/T	0.26	0.59	3.87	0.56	0.73	0.23
rs3764159	A/C	0.44	0.87	1.92	0.80	0.65	0.57
rs227028	G/T	2.76	0.68	1.31	0.58	0.66	0.25
rs227011	A/G	0.20	0.94	3.87	0.52	0.23	0.78
rs227019	C/T	1.31	1.66	1.39	0.64	0.49	0.70
rs12587781	G/T	0.61	1.55	1.75	0.66	0.54	0.50
rs227003	A/G	0.52	0.14	2.32	0.82	0.45	0.81
rs1154153	A/G	0.68	1.62	1.69	0.65	0.45	0.54
rs227008	A/C	3.37	1.47	0.71	0.48	0.82	0.17
rs1263638	C/T	4.58	1.61	1.58	0.26	0.06	0.62
rs7147412	C/T	0.38	0.10	2.29	0.84	0.46	0.88

Bacterial colonization

Associations:

The association between bacterial colonization (among Pneumococcae, *Haemophilus influenzae* and Moraxella) at age 4 weeks and the SNPs polymorphism was tested. Figure 12 shows multiple SNPs and their association with bacterial colonization.

Estimates: -

Table 13: Bacterial colonization at age 4 weeks: hazard-ratio estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	1.00	2.07	3.89	0.00	0.01	0.01
rs3811215	A/C	1.00	0.84	0.43	0.45	0.38	0.26
rs227869	A/G	1.00	0.61	0.25	0.01	0.02	0.01
rs10083416	A/G	1.00	0.72	0.27	0.10	0.11	0.07
rs227870	A/G	1.00	2.43	6.95	0.00	0.00	0.00
rs1483971	G/T	1.00	0.71	0.33	0.06	0.08	0.04
rs3751467	A/G	1.00	0.00		0.05		
rs2128997	A/G	1.00	0.74	0.14	0.04	0.10	0.02
rs226985	C/T	1.00	2.08	4.27	0.00	0.00	0.00
rs2170810	A/G	1.00	2.15	4.62	0.00	0.00	0.00
rs6572493	A/G	1.00	1.48	2.83	0.03	0.05	0.02
rs7147589	A/G	1.00	0.86	0.00	0.07	0.27	0.03
rs17794137	A/G	1.00	0.99	0.99	0.99		
rs3811207	G/T	1.00	2.51	6.29	0.00	0.00	0.00
rs1010116	A/G	1.00	0.27	0.00	0.03	0.00	0.59
rs226998	A/G	1.00	1.24	2.19	0.17	0.22	0.08
rs226999	A/G	1.00	0.38	0.00	0.04	0.01	0.35
rs1483973	C/T	1.00	0.62	0.51	0.22	0.10	0.33
rs1483974	A/C	1.00	1.96	0.00	0.05	0.04	1.00
rs227000	A/G	1.00	1.32	1.67	0.49	0.27	0.45
rs12887424	A/G	1.00	0.68	0.29	0.03	0.07	0.02
rs170544	C/T	1.00	0.45	0.00	0.16	0.05	1.00
rs3764159	A/C	1.00	1.03	0.54	0.82	0.94	0.53
rs227028	G/T	1.00	1.20	0.39	0.05	0.87	0.02
rs227011	A/G	1.00	0.30	0.00	0.03	0.01	1.00
rs227019	C/T	1.00	0.98	0.29	0.16	0.55	0.06
rs12587781	G/T	1.00	1.04	0.00	0.63	0.84	0.35
rs227003	A/G	1.00	4.30	2.23	0.00	0.00	0.88
rs1154153	A/G	1.00	1.04	0.00	0.63	0.84	0.35
rs227008	A/C	1.00	1.41	0.43	0.03	0.76	0.02
rs1263638	C/T	1.00	1.24	0.59	0.30	0.79	0.17
rs7147412	C/T	1.00	5.55	2.54	0.00	0.00	0.70

JJJ

Associations:

The association between the time to JJJ events and genotypes was tested. Figure 13 shows multiple SNPs and their association with JJJ.

Estimates: -

Table 14: Time to JJJ events: hazard-ratio estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	1.00	1.02	1.11	0.93	0.86	0.71
rs3811215	A/C	1.00	1.13	0.82	0.79	0.74	0.64
rs227869	A/G	1.00	1.60	1.03	0.13	0.15	0.39
rs10083416	A/G	1.00	1.17	0.28	0.10	0.98	0.04
rs227870	A/G	1.00	1.10	0.91	0.87	0.78	0.74
rs1483971	G/T	1.00	1.22	0.78	0.40	0.70	0.29
rs3751467	A/G	1.00	3.41		0.00		
rs2128997	A/G	1.00	1.20	0.47	0.25	0.75	0.14
rs226985	C/T	1.00	0.77	0.79	0.52	0.25	0.72
rs2170810	A/G	1.00	0.84	0.83	0.73	0.43	0.75
rs6572493	A/G	1.00	1.28	1.13	0.63	0.38	0.95
rs7147589	A/G	1.00	1.39	0.22	0.06	0.39	0.06
rs3811207	G/T	1.00	0.86	0.99	0.80	0.56	0.88
rs1010116	A/G	1.00	1.24	0.00	0.29	0.86	0.17
rs226998	A/G	1.00	0.81	0.78	0.62	0.33	0.67
rs226999	A/G	1.00	1.20	0.42	0.51	0.73	0.35
rs1483973	C/T	1.00	1.21	0.87	0.49	0.71	0.34
rs1483974	A/C	1.00	1.49	0.00	0.15	0.11	0.39
rs227000	A/G	1.00	1.27	1.11	0.59	0.33	0.99
rs12887424	A/G	1.00	0.81	0.92	0.72	0.48	0.85
rs170544	C/T	1.00	0.84	1.82	0.69	0.63	0.53
rs3764159	A/C	1.00	0.74	0.81	0.55	0.28	0.84
rs227028	G/T	1.00	0.88	0.91	0.88	0.62	0.95
rs227011	A/G	1.00	0.76	1.80	0.58	0.48	0.53
rs227019	C/T	1.00	0.72	1.06	0.32	0.25	0.54
rs12587781	G/T	1.00	0.78	1.08	0.67	0.42	0.86
rs227003	A/G	1.00	1.25	1.62	0.31	0.21	0.22
rs1154153	A/G	1.00	0.77	1.08	0.64	0.39	0.85
rs227008	A/C	1.00	0.84	0.86	0.76	0.46	0.87
rs1263638	C/T	1.00	0.86	0.94	0.82	0.57	0.94
rs7147412	C/T	1.00	1.08	1.55	0.38	0.46	0.17

J45

Associations:

The association between the time to J45 events and genotypes was tested. Figure 14 shows multiple SNPs and their association with J45.

Estimates: -

Table 15: Time to J45 events: hazard-ratio estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	1.00	0.91	0.93	0.95	0.75	0.98
rs3811215	A/C	1.00	0.94	0.46	0.55	0.58	0.28
rs227869	A/G	1.00	1.41	0.84	0.26	0.42	0.29
rs10083416	A/G	1.00	1.09	0.34	0.24	0.85	0.10
rs227870	A/G	1.00	1.20	1.18	0.76	0.46	0.84
rs1483971	G/T	1.00	1.04	0.69	0.57	0.81	0.29
rs3751467	A/G	1.00	3.29		0.00		
rs2128997	A/G	1.00	1.09	0.57	0.55	0.99	0.30
rs226985	C/T	1.00	0.85	0.93	0.83	0.58	0.99
rs2170810	A/G	1.00	1.12	1.05	0.91	0.70	0.99
rs6572493	A/G	1.00	1.51	1.47	0.34	0.14	0.65
rs7147589	A/G	1.00	1.33	0.26	0.15	0.52	0.12
rs3811207	G/T	1.00	1.09	1.37	0.78	0.63	0.52
rs1010116	A/G	1.00	1.00	0.00	0.46		
rs226998	A/G	1.00	0.91	0.71	0.69	0.55	0.44
rs226999	A/G	1.00	1.09	0.00	0.35		
rs1483973	C/T	1.00	1.25	0.61	0.14	0.85	0.07
rs1483974	A/C	1.00	1.94	0.00	0.02		
rs227000	A/G	1.00	1.13	1.08	0.89	0.64	0.96
rs12887424	A/G	1.00	0.62	0.84	0.21	0.12	0.69
rs170544	C/T	1.00	0.76	0.00	0.59		
rs3764159	A/C	1.00	0.81	1.01	0.77	0.51	0.93
rs227028	G/T	1.00	0.83	0.83	0.76	0.46	0.80
rs227011	A/G	1.00	0.75	0.00	0.58		
rs227019	C/T	1.00	0.65	1.11	0.20	0.19	0.41
rs12587781	G/T	1.00	0.77	1.31	0.64	0.49	0.65
rs227003	A/G	1.00	1.18	2.00	0.11	0.22	0.05
rs1154153	A/G	1.00	0.76	1.31	0.61	0.46	0.64
rs227008	A/C	1.00	0.81	0.78	0.69	0.39	0.71
rs1263638	C/T	1.00	0.80	1.02	0.66	0.52	0.67
rs7147412	C/T	1.00	1.02	1.91	0.10	0.44	0.03

Exacerbation

Associations:

The association between the time to exacerbation events and genotypes was tested. Figure 15 shows multiple SNPs and their association with exacerbation.

Estimates: -

Table 16: Time to exacerbation events: hazard-ratio estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	1.00	0.75	0.78	0.64	0.34	0.82
rs3811215	A/C	1.00	1.09	0.27	0.35	0.87	0.16
rs227869	A/G	1.00	1.45	0.51	0.05	0.57	0.04
rs10083416	A/G	1.00	1.08	0.83	0.88	0.90	0.68
rs227870	A/G	1.00	1.39	1.03	0.47	0.30	0.79
rs1483971	G/T	1.00	1.20	0.40	0.10	0.91	0.04
rs3751467	A/G	1.00	2.46		0.02		
rs2128997	A/G	1.00	0.83	1.05	0.77	0.58	0.79
rs226985	C/T	1.00	0.91	1.02	0.94	0.82	0.86
rs2170810	A/G	1.00	1.19	1.52	0.53	0.38	0.34
rs6572493	A/G	1.00	1.16	1.26	0.81	0.55	0.66
rs7147589	A/G	1.00	1.03	1.32	0.87	0.81	0.60
rs3811207	G/T	1.00	1.24	1.71	0.46	0.32	0.32
rs1010116	A/G	1.00	1.60	0.00	0.19	0.35	0.26
rs226998	A/G	1.00	1.02	1.09	0.98	0.90	0.85
rs226999	A/G	1.00	1.97	0.00	0.02	0.06	0.20
rs1483973	C/T	1.00	2.13	0.95	0.02	0.10	0.11
rs1483974	A/C	1.00	1.73	0.00	0.09	0.05	0.47
rs227000	A/G	1.00	0.99	1.22	0.91	0.93	0.66
rs12887424	A/G	1.00	1.15	1.23	0.86	0.60	0.73
rs170544	C/T	1.00	1.06	0.00	0.82	0.93	0.55
rs3764159	A/C	1.00	1.21	0.66	0.72	0.61	0.64
rs227028	G/T	1.00	1.14	1.04	0.91	0.71	0.91
rs227011	A/G	1.00	1.06	0.00	0.82	0.95	0.55
rs227019	C/T	1.00	1.16	1.42	0.69	0.49	0.49
rs12587781	G/T	1.00	1.17	0.85	0.86	0.65	0.84
rs227003	A/G	1.00	0.82	1.06	0.73	0.61	0.70
rs1154153	A/G	1.00	1.15	0.85	0.88	0.69	0.84
rs227008	A/C	1.00	1.17	1.00	0.83	0.68	0.78

rs1263638	C/T	1.00	1.38	1.21	0.56	0.31	0.99
rs7147412	C/T	1.00	0.69	1.05	0.39	0.34	0.54

Exacerbation incidences

Associations:

The association between exacerbation incidences and genotypes was tested. Figure 16 shows multiple SNPs and their association with exacerbation incidences.

Eczema

Associations:

The association between the time to eczema events and genotypes was tested. Figure 17 shows multiple SNPs and their association with eczema.

Estimates: -

Table 17: Time to eczema events: hazard-ratio estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	1.00	1.45	1.24	0.17	0.09	0.86
rs3811215	A/C	1.00	0.85	0.80	0.57	0.29	0.62
rs227869	A/G	1.00	1.22	0.99	0.43	0.39	0.54
rs10083416	A/G	1.00	0.92	0.52	0.17	0.28	0.07
rs227870	A/G	1.00	1.20	1.46	0.26	0.15	0.22
rs1483971	G/T	1.00	1.08	0.94	0.81	0.80	0.64
rs3751467	A/G	1.00	1.01		0.98		
rs2128997	A/G	1.00	0.96	0.84	0.85	0.70	0.62
rs226985	C/T	1.00	1.19	1.33	0.38	0.19	0.34
rs2170810	A/G	1.00	1.10	1.13	0.81	0.52	0.72
rs6572493	A/G	1.00	1.09	1.38	0.32	0.34	0.15
rs7147589	A/G	1.00	0.89	0.85	0.76	0.46	0.74
rs3811207	G/T	1.00	1.05	1.07	0.95	0.76	0.87
rs1010116	A/G	1.00	0.60	0.51	0.10	0.03	0.39
rs226998	A/G	1.00	0.89	0.89	0.76	0.46	0.80
rs226999	A/G	1.00	0.80	0.39	0.23	0.14	0.19
rs1483973	C/T	1.00	0.91	0.73	0.38	0.34	0.20
rs1483974	A/C	1.00	1.21	0.63	0.44	0.29	0.60
rs227000	A/G	1.00	0.97	1.05	0.97	0.93	0.83
rs12887424	A/G	1.00	0.74	0.74	0.19	0.07	0.58

rs170544	C/T	1.00	0.64	0.00	0.07		
rs3764159	A/C	1.00	1.56	0.46	0.01	0.03	0.19
rs227028	G/T	1.00	0.85	0.87	0.62	0.33	0.82
rs227011	A/G	1.00	0.78	0.00	0.30		
rs227019	C/T	1.00	0.96	0.94	0.96	0.79	0.89
rs12587781	G/T	1.00	1.54	0.63	0.03	0.02	0.40
rs227003	A/G	1.00	0.98	1.15	0.80	0.89	0.51
rs1154153	A/G	1.00	1.57	0.63	0.02	0.02	0.40
rs227008	A/C	1.00	0.81	0.85	0.48	0.23	0.84
rs1263638	C/T	1.00	0.88	1.05	0.66	0.60	0.58
rs7147412	C/T	1.00	1.03	1.06	0.96	0.80	0.84

Age at onset of asthma related events (composite score)

Associations:

The association between the age at onset of asthma related events and genotypes was tested. Figure 18 shows multiple SNPs and their association with asthma related events.

Estimates: -

Table 18: Time to asthma related events: hazard-ratio estimates and significance

	Allele	Ref. homoz.	Heteroz.	Rare homoz.	Cod.	Dom	Rec.
rs996165	C/T	1.00	0.96	0.96	0.98	0.85	0.94
rs3811215	A/C	1.00	1.07	0.83	0.87	0.89	0.65
rs227869	A/G	1.00	1.44	0.80	0.09	0.32	0.14
rs10083416	A/G	1.00	1.27	0.49	0.12	0.54	0.10
rs227870	A/G	1.00	1.18	0.86	0.62	0.61	0.54
rs1483971	G/T	1.00	1.23	0.68	0.17	0.75	0.11
rs3751467	A/G	1.00	2.55		0.00		
rs2128997	A/G	1.00	1.23	0.85	0.51	0.46	0.54
rs226985	C/T	1.00	0.89	0.73	0.59	0.39	0.39
rs2170810	A/G	1.00	0.96	0.78	0.76	0.66	0.48
rs6572493	A/G	1.00	1.21	0.92	0.52	0.59	0.46
rs7147589	A/G	1.00	1.37	0.79	0.24	0.21	0.48
rs3811207	G/T	1.00	0.96	0.96	0.98	0.84	0.95
rs1010116	A/G	1.00	1.43	0.00	0.12	0.49	0.12
rs226998	A/G	1.00	0.75	0.80	0.40	0.18	0.81
rs226999	A/G	1.00	1.43	0.35	0.15	0.29	0.24
rs1483973	C/T	1.00	1.47	0.91	0.11	0.30	0.18
rs1483974	A/C	1.00	1.50	0.00	0.10	0.08	0.33
rs227000	A/G	1.00	1.18	1.04	0.72	0.46	0.92

rs12887424	A/G	1.00	1.08	1.16	0.88	0.68	0.69
rs170544	C/T	1.00	0.94	1.49	0.89	0.88	0.68
rs3764159	A/C	1.00	0.89	0.67	0.77	0.53	0.60
rs227028	G/T	1.00	1.08	1.03	0.94	0.77	0.95
rs227011	A/G	1.00	0.91	1.49	0.86	0.79	0.68
rs227019	C/T	1.00	0.98	1.25	0.77	0.90	0.47
rs12587781	G/T	1.00	0.94	0.89	0.95	0.76	0.89
rs227003	A/G	1.00	1.03	1.22	0.81	0.73	0.53
rs1154153	A/G	1.00	0.92	0.89	0.94	0.72	0.90
rs227008	A/C	1.00	1.02	0.96	0.98	0.98	0.84
rs1263638	C/T	1.00	1.08	1.03	0.94	0.75	0.95
rs7147412	C/T	1.00	0.89	1.15	0.67	0.79	0.48

Table 19 summarizes the data above and shows the association of SNP markers in the vicinity of the TCRA gene on chr 14 with bacterial colonization of the newborn/infant airway. Data presented show chr location, basepair location, allele involved, frequency in affected vs frequency in unaffected controls, chisquare association value, P value for all bacteria, odds ratios for all bacteria, P values and odds ratios for individual bacterias, including staphylococcus, pneumococcus, moraxella, hemophilus and hemophilus/moraxella combined. As shown the association is strongest to "all" bacteria, and to hemphilus, moraxella and hemophilus/moraxella combined.

Table 19

CHR	SNP	BP	ALL P	ALL OR	Staph P	Staph OR	Pneumo P	Pneumo OR	Mora P	Mora OR	Haem P	Haem OR	Haem_mora P	Haem_mora OR
14	rs996165	22024524	0.000758	2.021	0.6935	0.9351	0.5126	1.203	0.01294	2.157	0.03202	1.903	0.001588	2.1
14	rs3811215	22024617	0.2473	0.7476	0.8305	0.9581	0.1614	0.5923	0.4469	1.286	0.2608	0.6539	0.6418	0.8804
14	rs227869	22026550	0.00371	0.5305	0.1748	1.269	0.01251	0.4605	0.2433	0.6995	0.03249	0.4997	0.02544	0.5813
14	rs10083416	22026757	0.04922	0.6178	0.5364	0.8895	0.8598	1.057	0.1133	0.5517	0.2781	0.6833	0.06126	0.5936
14	rs227870	22027011	4.63E-07	2.826	0.3277	0.8346	0.01062	2.051	0.008789	2.161	0.0008443	2.595	3.79E-05	2.553
14	rs1483971	22027328	0.02136	0.6067	0.07509	1.368	0.02051	0.4815	0.4152	0.7764	0.1032	0.5977	0.09267	0.6664
14	rs3751467	22029432	0.03872	0	0.3372	0.6171	0.1946	0	0.224	0	0.2039	0	0.08277	0
14	rs2128997	22030942	0.02963	0.5741	0.2079	0.7847	0.9018	0.9608	0.04796	0.4451	0.4333	0.7586	0.04728	0.5625
14	rs226985	22033486	4.33E-05	2.305	0.5479	0.8987	0.03062	1.831	0.09135	1.648	0.003007	2.34	0.001108	2.094
14	rs2170810	22042582	6.14E-05	2.268	0.2316	0.8091	0.01621	1.955	0.05124	1.775	0.003549	2.307	0.000603	2.174
14	rs6572493	22042795	0.006596	1.739	0.6204	0.9184	0.07702	1.641	0.1538	1.523	0.04201	1.802	0.012	1.768
14	rs7147589	22042826	0.09047	0.6422	0.5984	0.8995	0.7765	0.9076	0.1157	0.5217	0.5045	0.7834	0.09886	0.6106
14	rs17794137	22043207	0.9782	0.9873	0.4055	1.405	0.2493	0.3247	0.3017	1.77	0.697	0.7481	0.4542	1.423
14	rs3811207	22044096	5.37E-05	2.37	0.2457	0.8031	0.1359	1.585	0.003918	2.33	0.000324	2.793	5.50E-06	2.818
14	rs1010116	22047509	0.006757	0.2226	0.2451	0.7086	0.05666	0.1777	0.08289	0.202	0.07304	0.1933	0.01705	0.2056
14	rs226998	22048356	0.07323	1.444	0.003117	0.5988	0.3513	1.302	0.09133	1.644	0.8144	1.072	0.1459	1.394
14	rs226999	22048417	0.006883	0.3217	0.1772	0.7081	0.1082	0.3895	0.1729	0.4449	0.05829	0.2729	0.01399	0.2934
14	rs1483973	22048600	0.1088	0.7189	0.03688	1.43	0.06478	0.5858	0.4212	0.7864	0.6282	0.868	0.1997	0.7449
14	rs1483974	22048615	0.09275	1.546	0.1526	0.7199	0.09769	1.744	0.3677	1.398	0.4472	1.326	0.1441	1.518

14	rs227000	22049188	0.241	1.295	0.4628	0.8705	0.4389	1.263	0.4817	1.251	0.4033	1.296	0.5293	1.169
14	rs12887424	22050248	0.01852	0.6084	0.5137	0.8942	0.1161	0.6305	0.2429	0.6996	0.08974	0.5955	0.05657	0.6381
14	rs170544	22053063	0.0723	0.4567	0.2768	0.7301	0.3083	0.5411	0.4286	0.6179	0.169	0.3773	0.08547	0.4109
14	rs3764159	22054426	0.8271	0.9381	0.2667	1.317	0.735	1.139	0.5922	0.7854	0.5181	0.7479	0.6543	0.861
14	rs227028	22054846	0.1895	0.7588	0.9396	1.013	0.3747	0.7717	0.6236	0.8614	0.3073	0.7336	0.3602	0.8071
14	rs227011	22060602	0.01904	0.3066	0.1425	0.6487	0.1604	0.3704	0.228	0.4219	0.2029	0.4033	0.04773	0.3191
14	rs227019	22064646	0.2337	0.7619	0.7596	1.058	0.7226	0.8957	0.2861	0.694	0.3544	0.7349	0.3009	0.7674
14	rs12587781	22069457	0.6122	0.8562	0.09802	1.535	0.8931	1.056	0.2204	0.524	0.635	0.8073	0.3899	0.7352
14	rs227003	22069902	0.004063	1.807	0.9191	0.982	0.1876	1.454	0.0844	1.666	0.01651	1.994	0.00929	1.81
14	rs1154153	22071387	0.6122	0.8562	0.09802	1.535	0.8931	1.056	0.2204	0.524	0.635	0.8073	0.3899	0.7352
14	rs227008	22074550	0.3098	0.8081	0.8844	0.9752	0.6183	0.8666	0.9129	0.9671	0.2079	0.6771	0.4616	0.8414
14	rs1263638	22079839	0.6491	0.9083	0.7565	1.056	0.7298	0.9038	0.9808	1.007	0.6152	0.858	0.8486	0.9561
14	rs7147412	22082384	0.002951	1.837	0.7866	0.9534	0.06287	1.684	0.08637	1.658	0.01795	1.976	0.00876	1.814

Figure 19 shows a significant association detected with haplotypes formed by block 1 which includes the most highly associated SNP rs227870. A single haplotype (the second most common in Europeans) is highly associated with the colonization phenotype.

Haplotype association analysis with the colonization data

Table 20. Associations of SNP haplotypes that reside at various haplotype blocks across the TCRA gene with bacterial colonization

Block	Haplotype	Freq. Case, Control	Ratio Counts Case,Control	Frequencies	Chi Square	P Value
Block 1						
CCACACAACCACCA	**0.213**	**44.4 : 73.0, 81.3 : 372.0**		**0.378, 0.179**	**21.493**	**3.5515E-6**
ACAACCACAACACC	0.211	16.8 : 100.6, 101.6 : 351.8		0.143, 0.224	3.719	0.0538
CACCCAAAAACCCC	0.150	15.3 : 102.1, 70.9 : 382.5		0.130, 0.156	0.492	0.4832
ACCCCAAAAACCCC	0.100	6.7 : 110.7, 51.8 : 401.6		0.057, 0.114	3.361	0.0668
Block 2						
CCC	0.608	64.9 : 53.1, 284.1 : 169.9		0.550, 0.626	2.258	0.1329
CAC	**0.262**	**47.1 : 70.9, 107.8 : 346.2**		**0.399, 0.237**	**12.379**	**4.0E-4**
Block 3						
AC	0.475	48.0 : 70.0, 222.0 : 232.0		0.407, 0.489	2.54	0.111
CC	0.371	46.0 : 72.0, 167.0 : 287.0		0.390, 0.368	0.194	0.6598
CA	0.153	24.0 : 94.0, 65.0 : 389.0		0.203, 0.143	2.585	0.1079
Block 4						
AC	0.540	76.0 : 42.0, 245.0 : 209.0		0.644, 0.540	4.147	0.0417
CC	0.362	36.0 : 82.0, 163.0 : 291.0		0.305, 0.359	1.201	0.273
Block 5						
AA	0.440	59.0 : 59.0, 193.0 : 261.0		0.500, 0.425	2.131	0.1443
AC	0.414	43.0 : 75.0, 193.0 : 261.0		0.364, 0.425	1.424	0.2327
CA	0.145	16.0 : 102.0, 68.0 : 386.0		0.136, 0.150	0.15	0.6981
Block 6						

CCAAACCC	0.340	54.0 : 64.0, 144.0 : 310.0	0.458, 0.317	8.162	0.0043
CAACAAAA	0.302	30.9 : 87.1, 134.9 : 319.1	0.262, 0.297	0.564	0.4525
CCCCCCCA	0.131	12.0 : 106.0, 62.9 : 391.1	0.102, 0.139	1.131	0.2877

Table 21

Chr	BP	A1	F_A	F_U	P-val age 7	OR age 7	P-val age 6	OR age 6	P-val age 5	OR age 5
rs996165	22024524	2	0.5264	0.4897	0.02052	1.158	0.03194	1.155	0.0336	1.167
rs3811215	22024617	1	0.264	0.2581	0.669	1.031	0.7244	0.9732	0.5919	0.9562
rs227869	22026550	2	0.4363	0.4621	0.1062	0.9013	0.05626	0.8778	0.315	0.9286
rs10083416	22026757	1	0.2826	0.2948	0.4013	0.9427	0.8219	0.9834	0.175	0.8955
rs227870	22027011	1	0.2871	0.2447	0.002097	1.243	0.004719	1.235	0.002309	1.277
rs1483971	22027328	1	0.3802	0.3951	0.3366	0.9392	0.1806	0.9112	0.3882	0.9374
rs3751467	22029432	2	0.0495	0.0605	0.1401	0.8089	0.1819	0.8157	0.4478	0.8848
rs2128997	22030942	2	0.2847	0.2985	0.3399	0.9353	0.4985	0.951	0.1408	0.8874
rs226985	22033486	2	0.3234	0.3034	0.1705	1.098	0.1601	1.106	0.08426	1.143
rs2170810	22042582	2	0.3234	0.2978	0.07841	1.127	0.06643	1.141	0.04883	1.165
rs6572493	22042795	1	0.4125	0.3967	0.3079	1.068	0.3997	1.059	0.2536	1.088
rs7147589	22042826	1	0.2236	0.2335	0.4604	0.9455	0.701	0.9697	0.2094	0.8953
rs17794137	22043207	1	0.06518	0.05303	0.09369	1.245	0.2317	1.182	0.2686	1.181
rs1010116	22047509	1	0.08663	0.09895	0.1892	0.8638	0.3411	0.8943	0.1059	0.809
rs226998	22048356	1	0.3828	0.3622	0.1769	1.092	0.1732	1.099	0.34	1.074
rs226999	22048417	1	0.118	0.1354	0.1049	0.854	0.1764	0.8702	0.04074	0.7911
rs1483973	22048600	1	0.4414	0.4916	0.001535	0.8171	0.02316	0.8579	0.01384	0.8355
rs1483974	22048615	1	0.2046	0.1774	0.02615	1.193	0.02974	1.2	0.04394	1.2

Table 21. Association of the TCRA SNPs (that associate with bacterial colonization) with asthma at the age of 5, 6 and 7 years for age. P values and odds ratios are presented.

Association of TCRA with lung function:

Association was also detected between rs1483971 and lung function measurements, including FEV0.5 (P=1.4 x 102); FEV1/FVC (P=5.5 x 10-3) and MMEF (P=1.3 x 10-3), lending further support for the role of the TCRA gene in asthma.

CONCLUSION

The screening of the association between 32 TCRA SNPs and various endpoints revealed some very strong associations with bacterial colonization. Six SNPs showed a particularly high association: *rs227870*, *rs226985*, *rs2170810*, *rs3811207*, *rs227003* and *rs7147412* (with a respective significance level of $2'10^{-5}$, $1'10^{-3}$, $1'10^{-3}$, $3'10^{-4}$, $5'10^{-5}$, $3'10^{-6}$, in the codominant model). The SNP *rs3751467* was particularly associated with asthma symptoms-related endpoints such as J45 (p=$4.5'10^{-4}$), JJJ (p=$6'10^{-5}$) and the composite score of asthma-related events (p=$5'10^{-3}$). Finally, SNP *rs1483971* was found to be associated with some lung function measurements such as $FEV_{0.5}$ (p=$1.4'10^{-2}$), FEV_1/FVC (p=$5.5'10^{-3}$) and MMEF (p=$1.3'10^{-3}$).

These TCRa novel variants that predispose to bacterial colonization of the airways of newborns and infants and associate robustly with the development of asthma in later life provide new targets for the development of agents to prevent and assess risk for the development of asthma.

REFERENCES

1. C. Ober, S. Hoffjan, *Genes Immun***7**, 95 (2006).

2. E. Halapi, H. Hakonarson, *CurrOpinPulm Med***10**, 22 (2004).

3. M. F. Moffatt *et al.*, *Nature***448**, 470 (2007).

4. *Nature***447**, 661 (2007).

5. E. Zeggini *et al.*, *Nat Genet***40**, 638 (2008).

6. M. N. Weedon *et al.*, *Nat Genet***40**, 575 (2008).

7. D. Falush, M. Stephens, J. K. Pritchard, *Genetics***164**, 1567 (2003).

8. H. Bisgaard, *Ann Allergy Asthma Immunol***93**, 381 (2004).

9. A. L. Price *et al.*, *Nat Genet***38**, 904 (2006).

10. S. Hoffjan, C. Ober, *CurrOpinImmunol***14**, 709 (2002).

11. R. Tavendale, D. F. Macgregor, S. Mukhopadhyay, C. N. Palmer, *J Allergy ClinImmunol***121**, 860 (2008).

12. J. Galanter *et al.*, *Am J RespirCrit Care Med* (2008).

13. I. Gosens, A. I. den Hollander, F. P. Cremers, R. Roepman, *Exp Eye Res* (2008).

14. A. I. den Hollander *et al.*, *Nat Genet***23**, 217 (1999).

15. A. I. den Hollander *et al.*, *Hum Mol Genet***10**, 2767 (2001).

16. K. Del Villar, C. A. Miller, *Proc NatlAcadSci U S A***101**, 4210 (2004).

17. M. Lindstedt, K. Lundberg, C. A. Borrebaeck, *J Immunol***175**, 4839 (2005).

18. Y. J. Liu, *Cell***106**, 259 (2001).

19. T. Willinger, T. Freeman, H. Hasegawa, A. J. McMichael, M. F. Callan, *J Immunol***175**, 5895 (2005).

20. M. Wills-Karp, S. L. Ewart, *Nat Rev Genet***5**, 376 (2004).

21. K.L. Gunderson, *et al., Nat Genet***37**, 549 (2005)

22. F.J. Steemers, *et al., Nat Methods***3**, 31(2006)

23. H. Hakonarson *et al., Nature***448**, 591(2007)

While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set forth in the following claims.

What is claimed is:

1. A method for detecting a propensity for increased bacterial colonization of the airway and an increased risk for developing asthma, the method comprising: detecting the presence of at least one single nucleotide polymorphism on chromosome 14 in a target TCR alpha encoding polynucleotide provided in Table 1, wherein if said single nucleotide polymorphism is present, said patient has an increased risk for increased bacterial airway colonization and an increased risk for developing asthma

2. A method as claimed in claim 1, wherein the target nucleic acid is amplified prior to detection.

3. The method of claim 1, wherein the step of detecting the presence of said single nucleotide polymorphism further comprises the step of analyzing a polynucleotide sample to determine the presence of said single nucleotide polymorphism by performing a process selected from the group consisting of detection of specific hybridization, measurement of allele size, restriction fragment length polymorphism analysis, allele-specific hybridization analysis, single base primer extension reaction, and sequencing of an amplified polynucleotide.

4. A method as claimed in claim 1, wherein in the target nucleic acid is DNA.

5. The method of claim 1, wherein nucleic acids comprising said polymorphism are obtained from an isolated cell of the human subject.

6. A single nucleotide polymorphism associated with an increased risk of developing asthma selected from the group consisting of those set forth in Table 1.

7. A method for identifying agents which modulate TCRa signaling or function which contributes to the asthmatic phenotype:

a) providing T cells in host animals expressing a single nucleotide polymorphism selected from the group consisting of those set forth in Table 1;

b) providing T cells in host animals which express the cognate sequences which lack the polymorphisms of step a);

c) contacting the cells of steps a) and b) with a test agent and

d) analyzing whether said agent alters an asthma associated parameter in cells contacted in step a) relative to those of step b) in said animals, thereby identifying agents which modulate the asthmatic phenotype.

8. The method of claim 7, wherein said asthma associated parameter is selected from the group consisting of $FEV_{0.5}$, $PD_{15}TCO_2$, PD_{20} metacholine, FEV_1, $FEV_{1/FVC}$, MMEF, wheeziness, bacterial colonization, and eczema.

9. The method of claim 8, wherein said cells are selected from the group consisting of T lymphocytes and recombinant cells expressing TCRa.

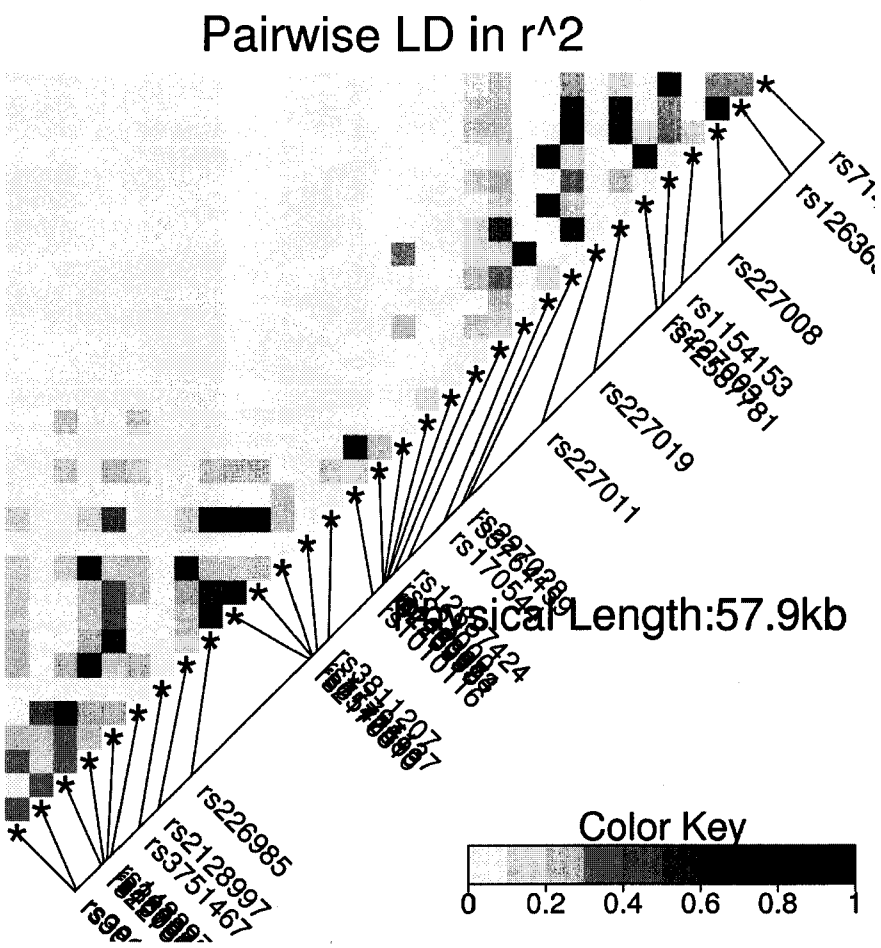

Pairwise LD in r^2
rs227008
rs1154153
rs227019
rs227011
rs226985
rs2128997
rs3751467
Color Key
0
0.2
0.4
0.6
0.8
1

Figure 1

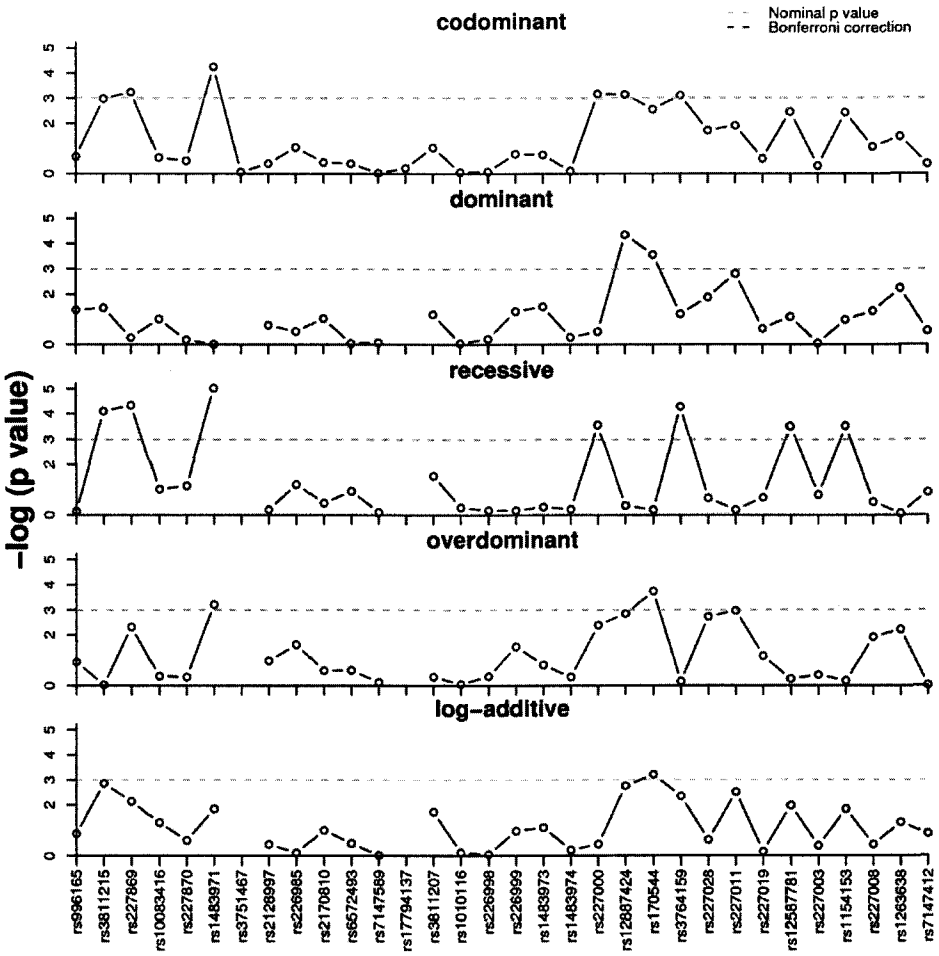

codominant
dominant
recessive
overdominant
log-additive
Nominal p value
Bonferroni correction
-log (p value)
0
1
2
3
4
5
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs17794137
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 2

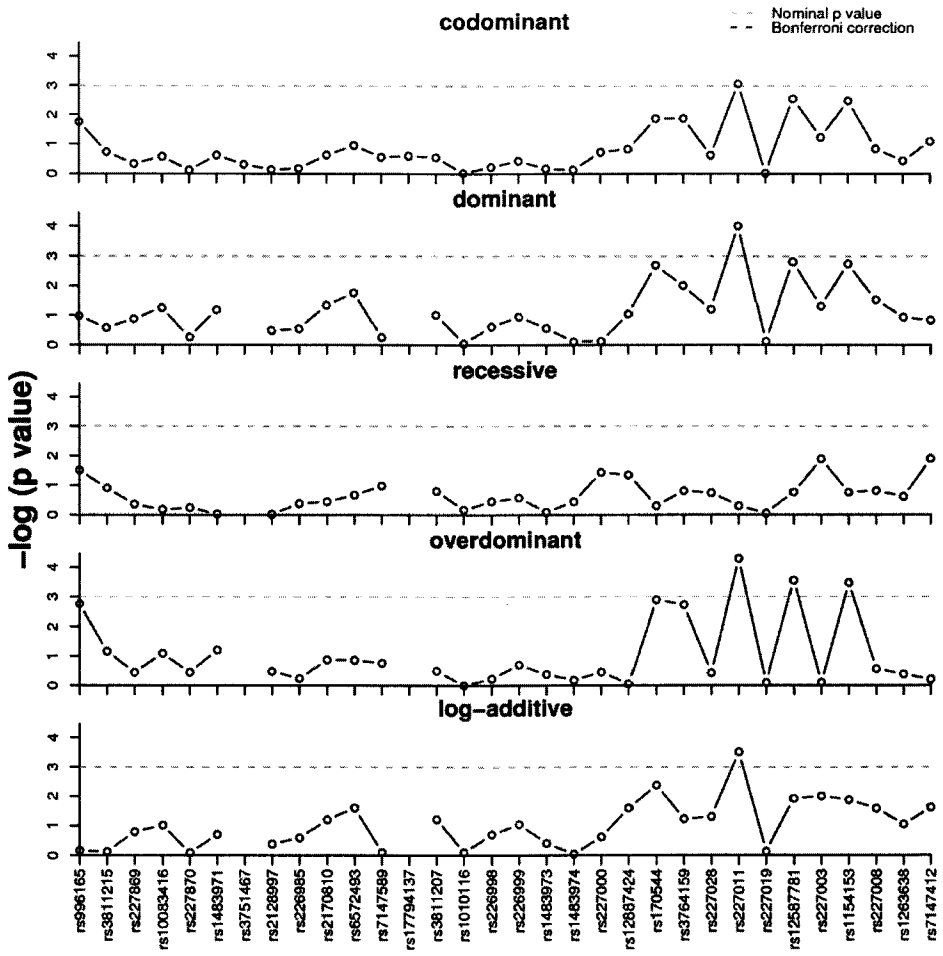

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0 1 2 3 4
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs14839971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs17794137
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12687424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 3

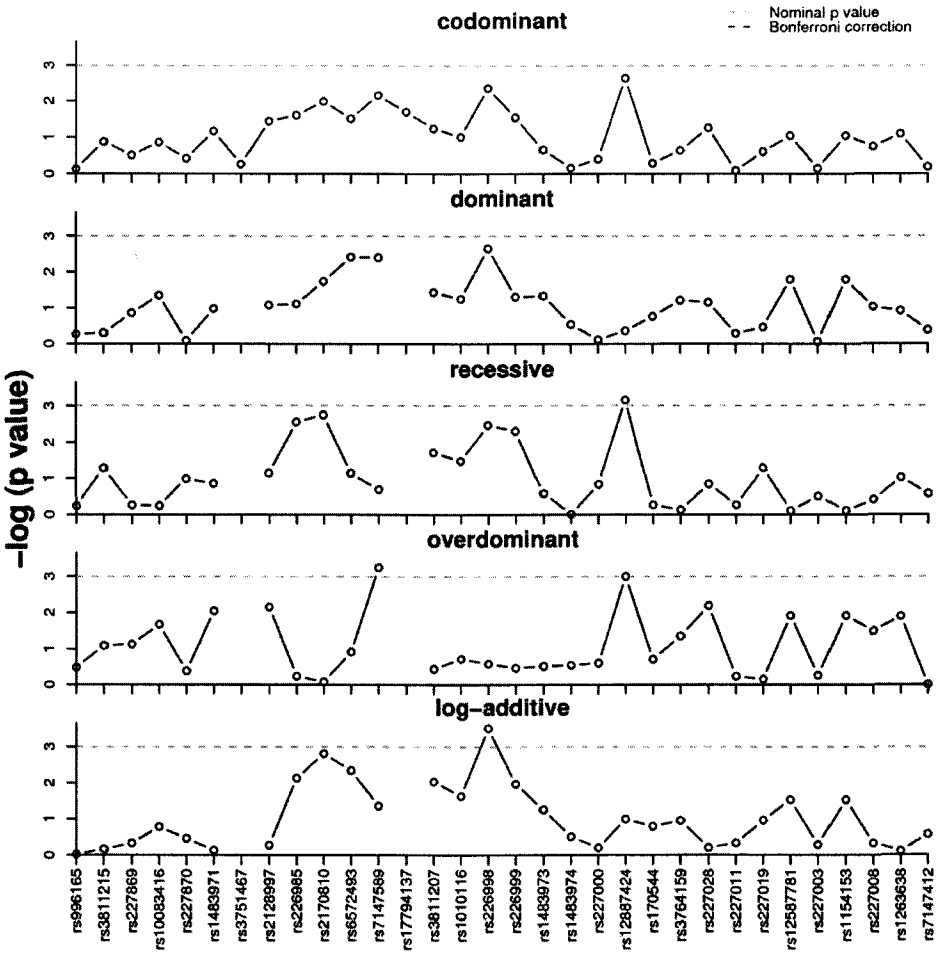

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572483
rs7147589
rs17794137
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs1170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 4

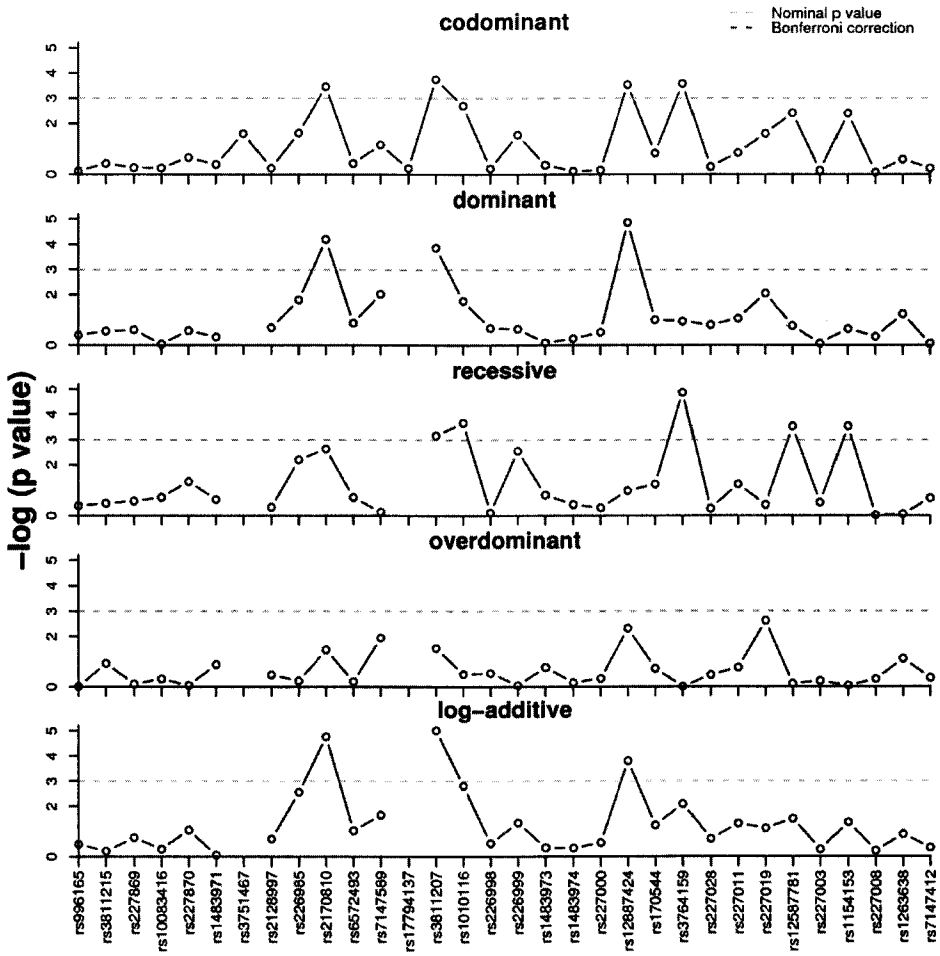

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
−log (p value)
0
1
2
3
4
5
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs17794137
rs3811207
rs10101116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 5

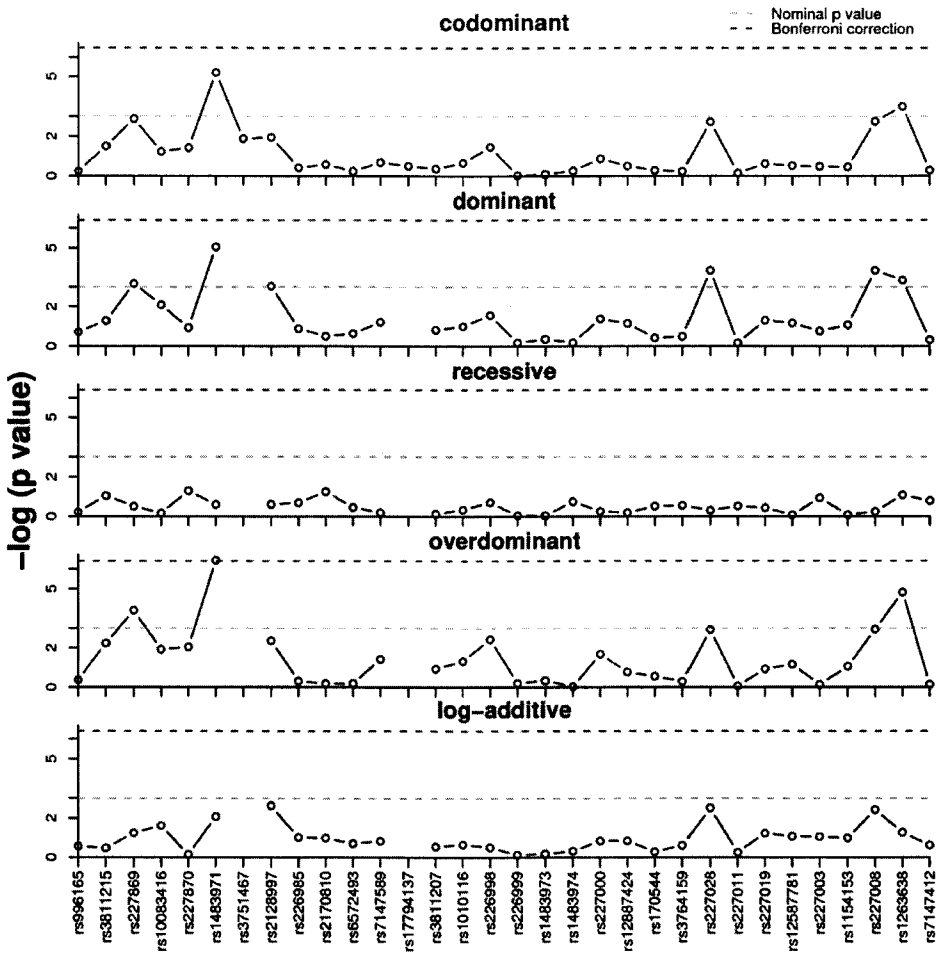

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
−log (p value)
0
2
5
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572483
rs7147589
rs17794137
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs1705544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 6

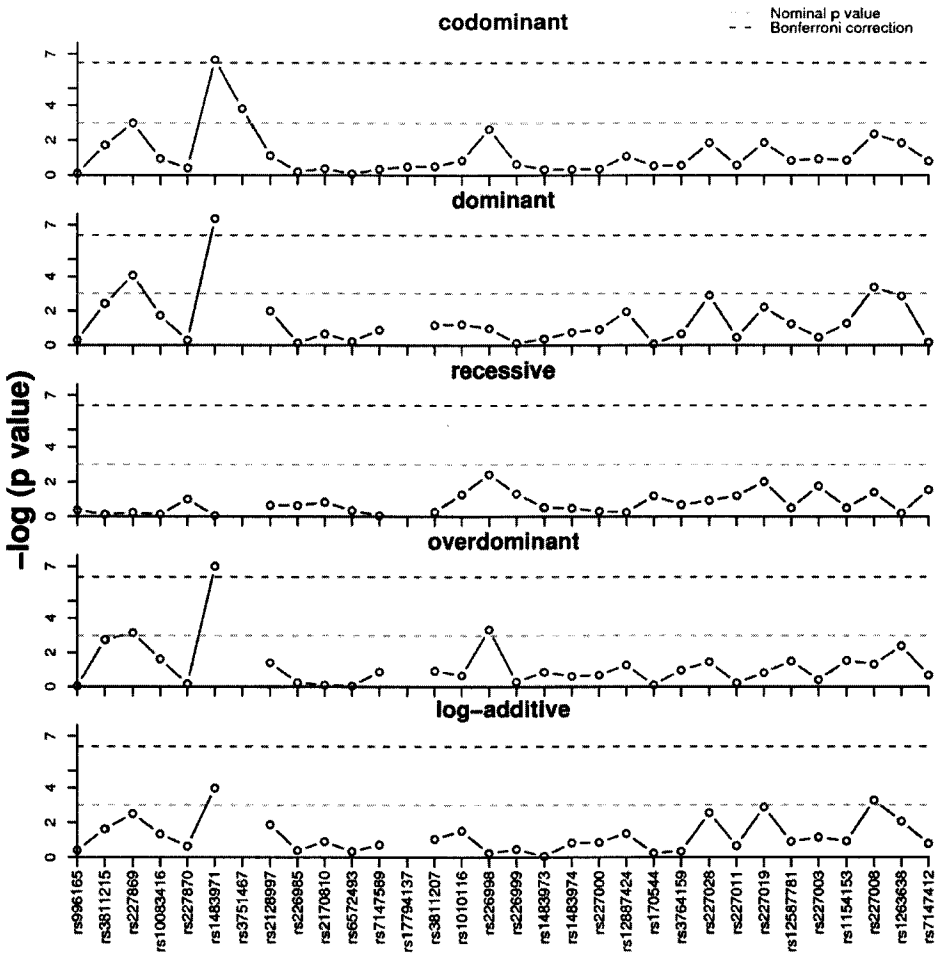

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0
2
4
7
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs17794137
rs3811207
rs10101116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 7

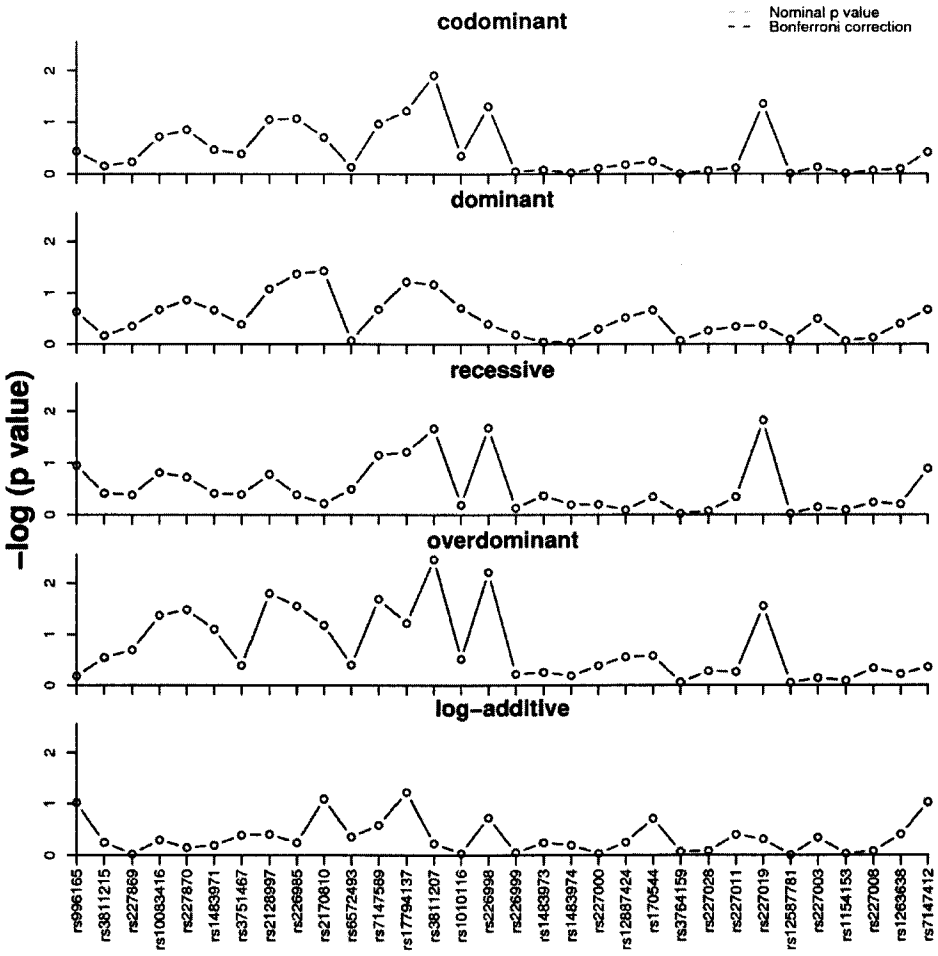

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0
1
2
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs21708610
rs6572493
rs7147589
rs17794137
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs1705444
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 8

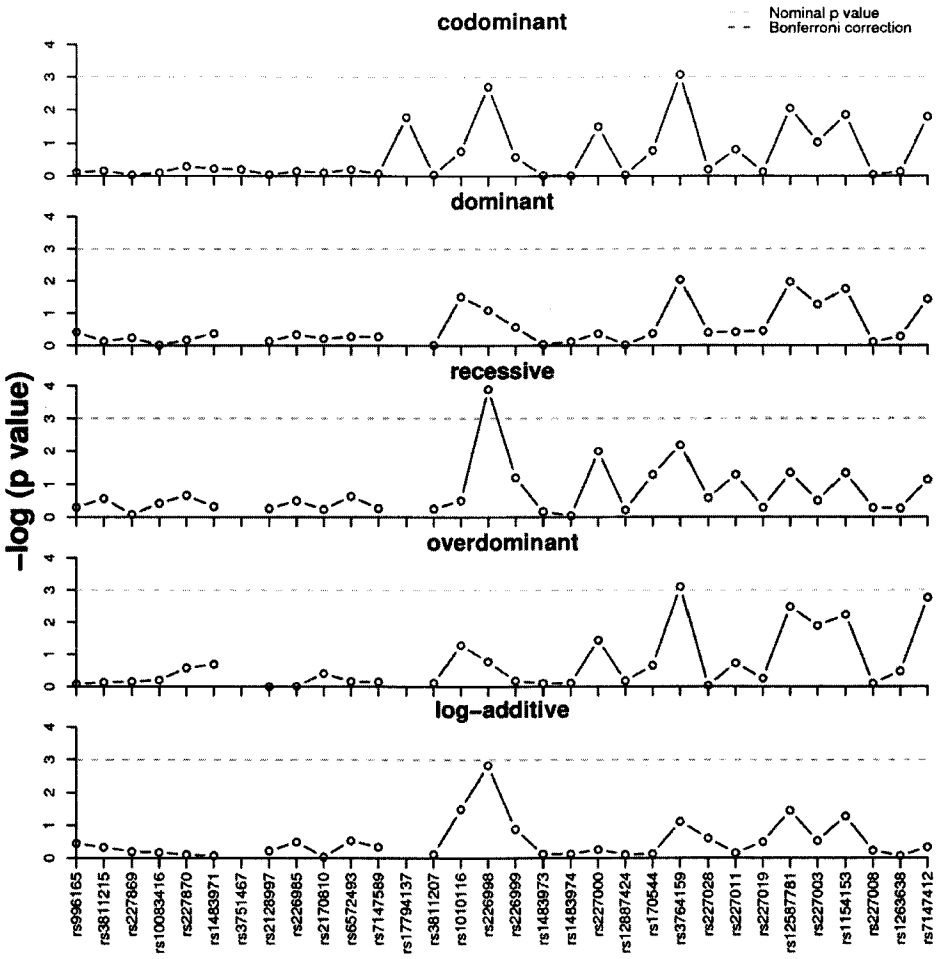

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170610
rs6572493
rs7147589
rs17794137
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 9

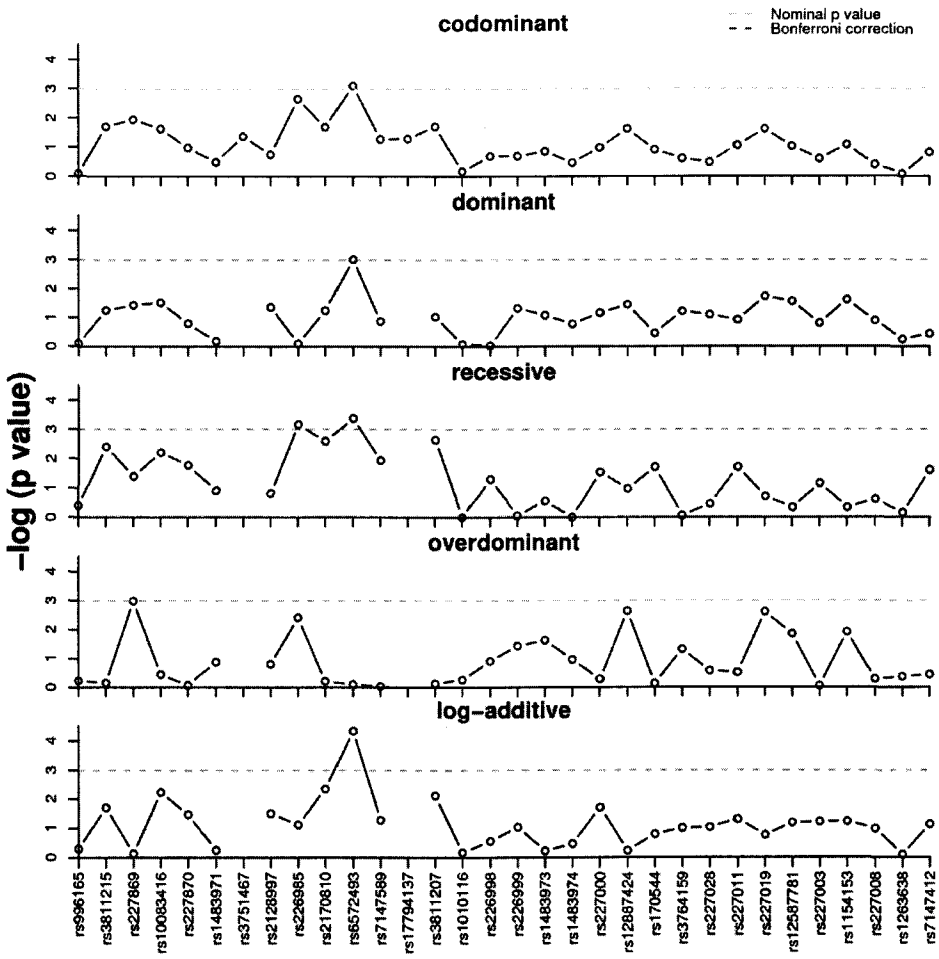

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0 1 2 3 4
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs17794137
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs1705544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 10

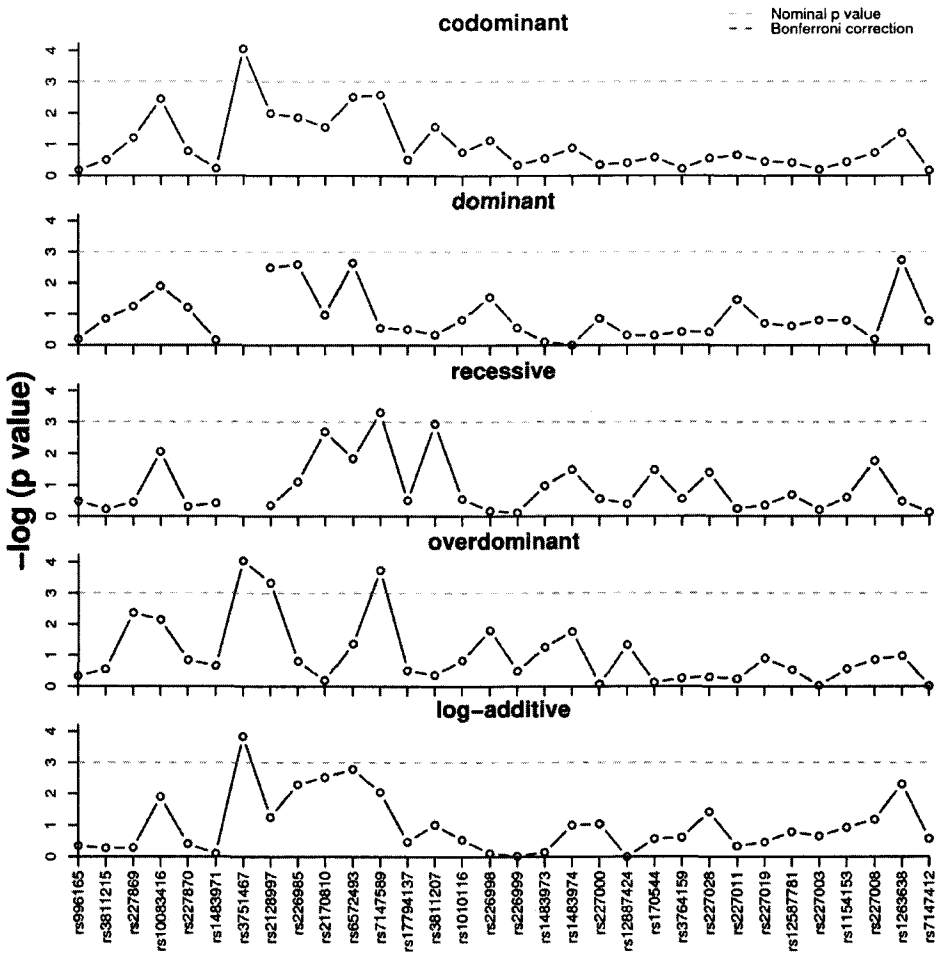

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0 1 2 3 4
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572483
rs7147589
rs17794137
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs1705544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1283638
rs7147412

Figure 11

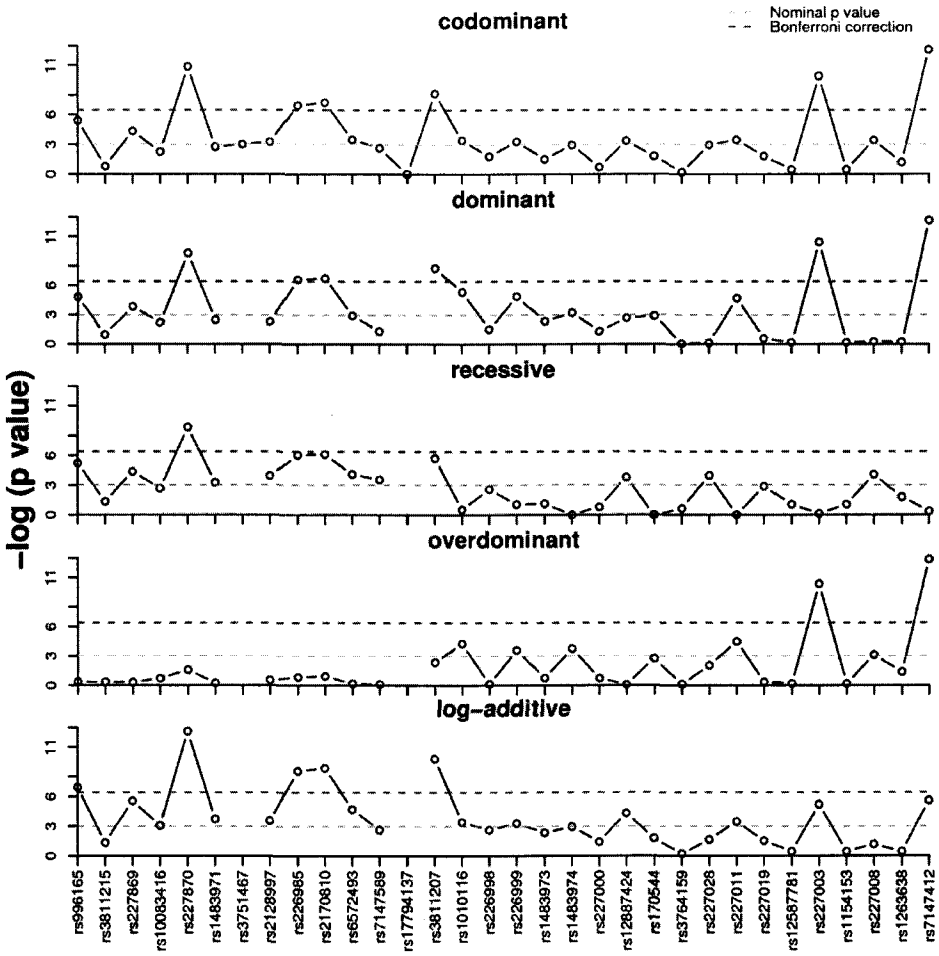

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0
3
6
11
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs17794137
rs3811207
rs10101116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs1170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 12

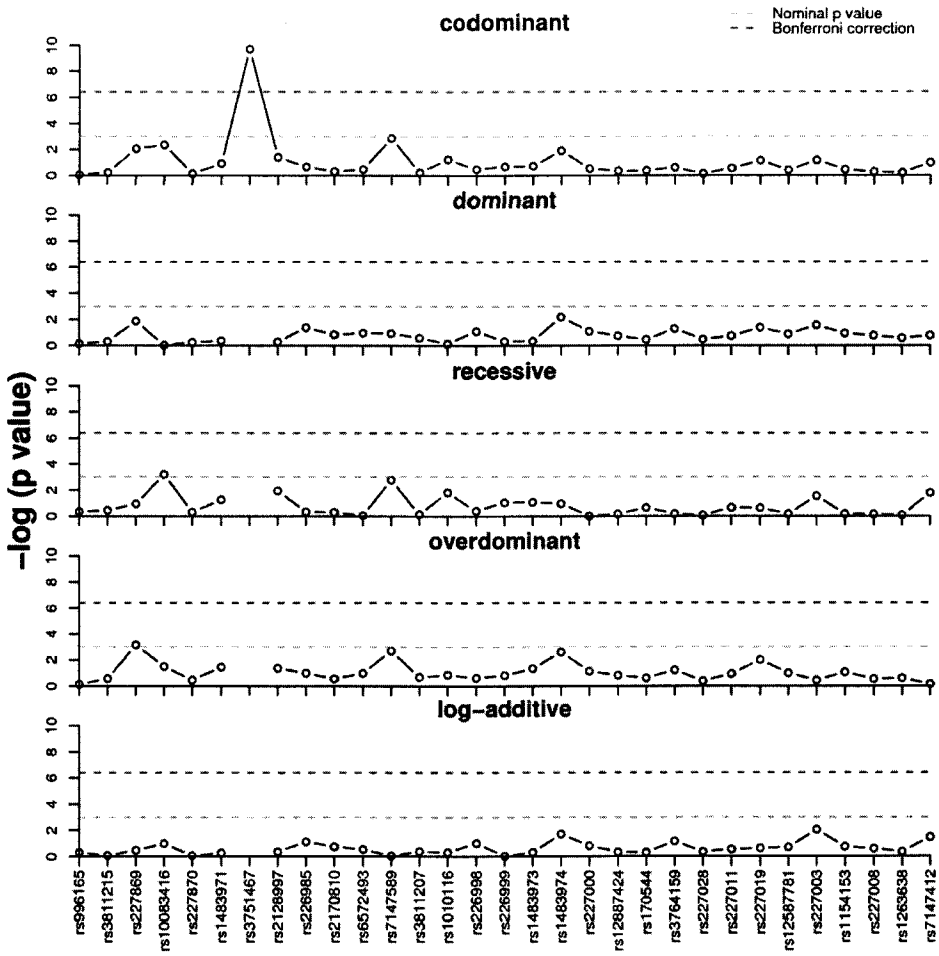

codominant
dominant
recessive
overdominant
log-additive
Nominal p value
Bonferroni correction
-log (p value)
0 2 4 6 8 10
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170610
rs6572493
rs7147589
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 13

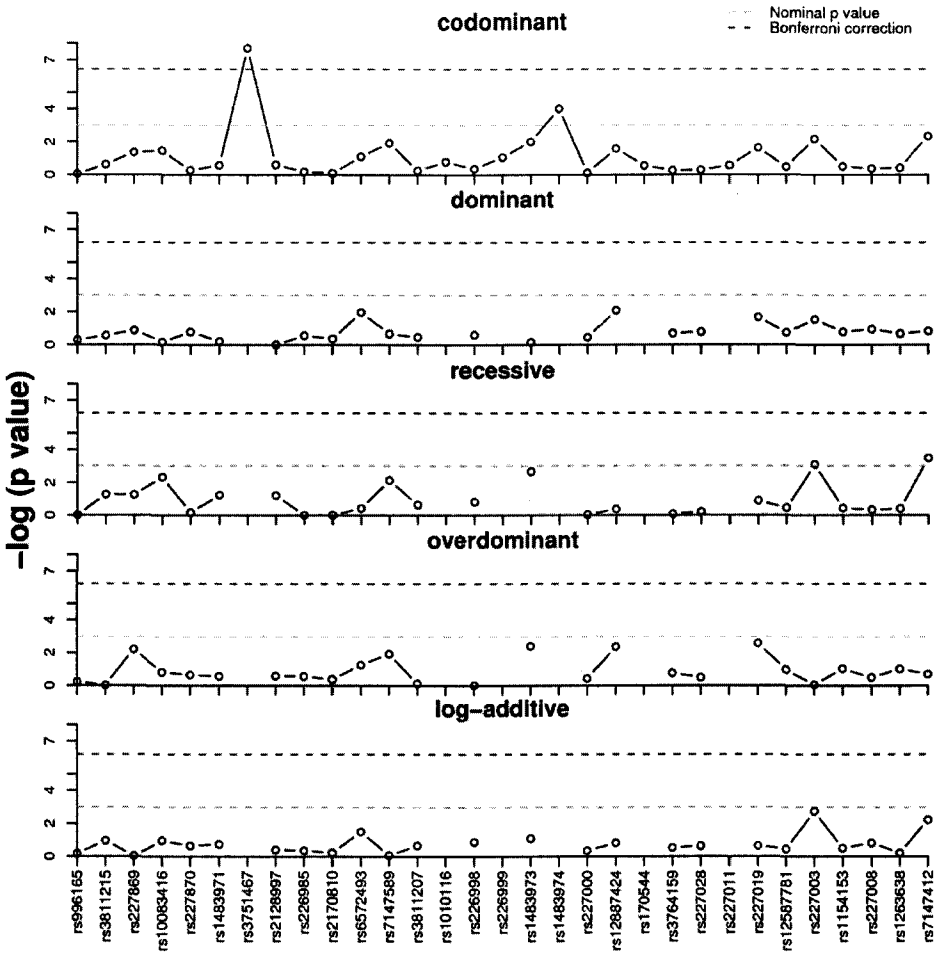

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
−log (p value)
0
2
4
7
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs3811207
rs10101116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12687424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 14

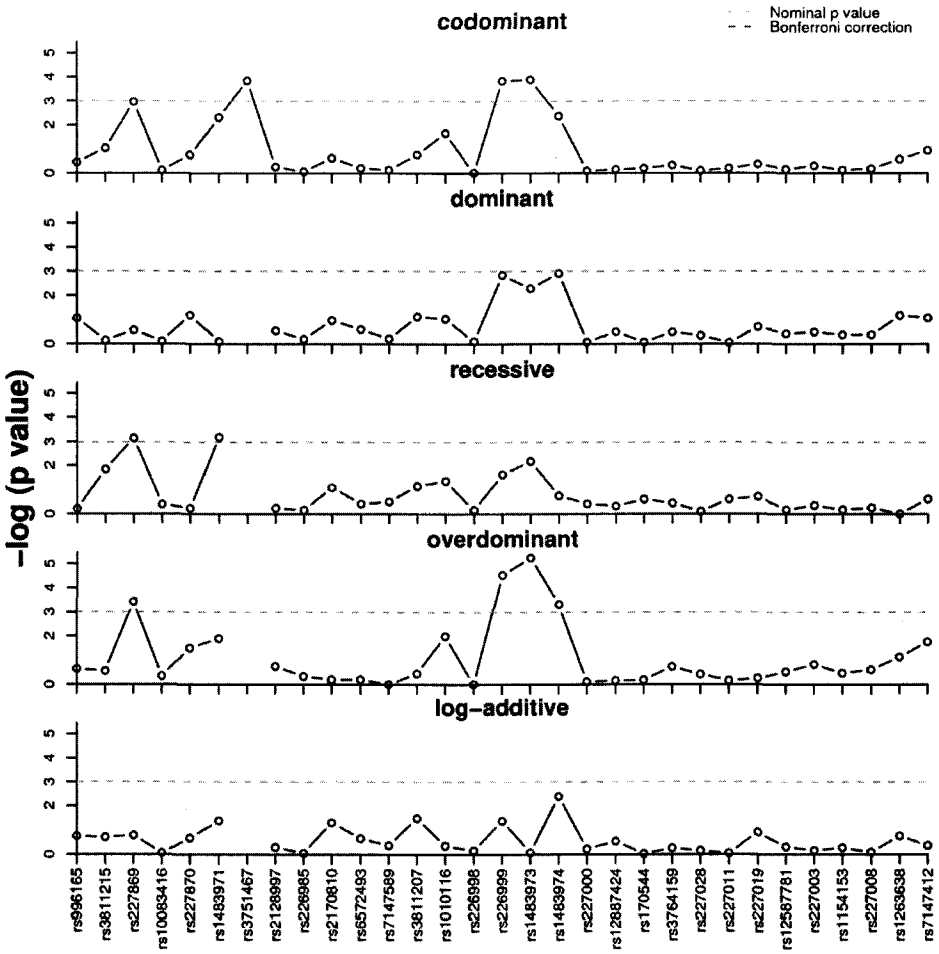

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0
1
2
3
4
5
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs3811207
rs10101116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12567781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 15

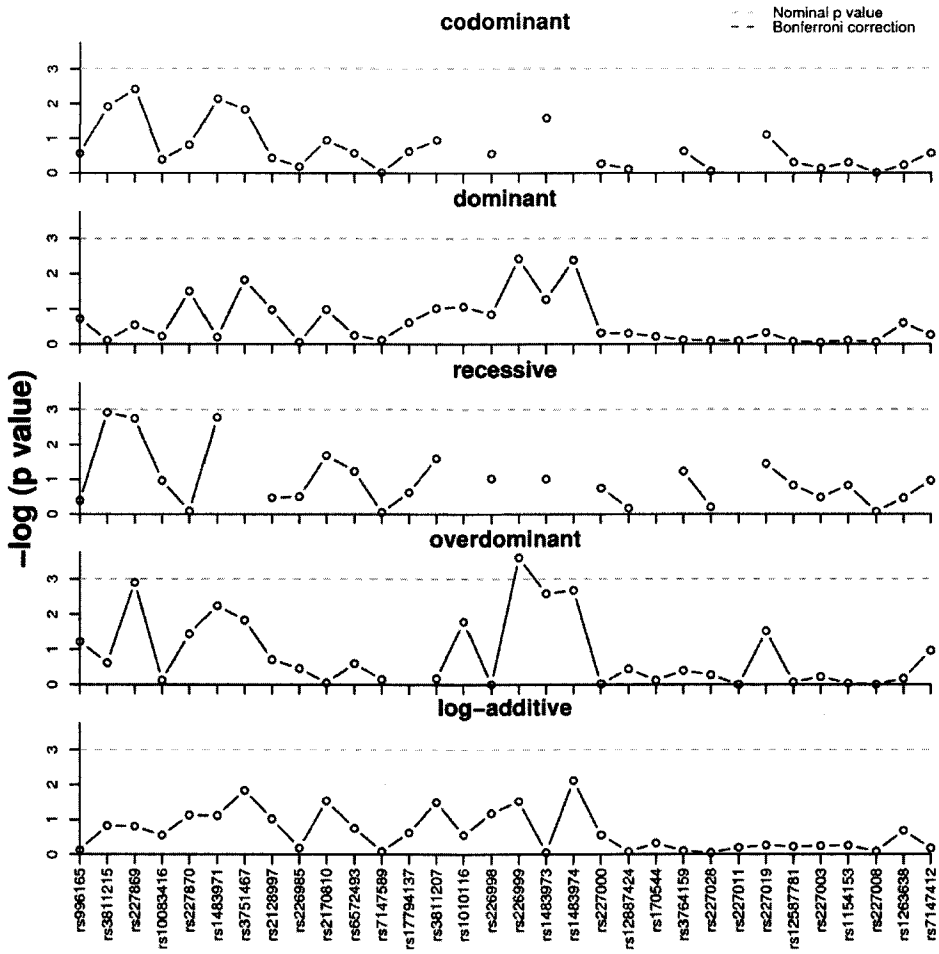

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0
1
2
3
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs17794137
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs1705441
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 16

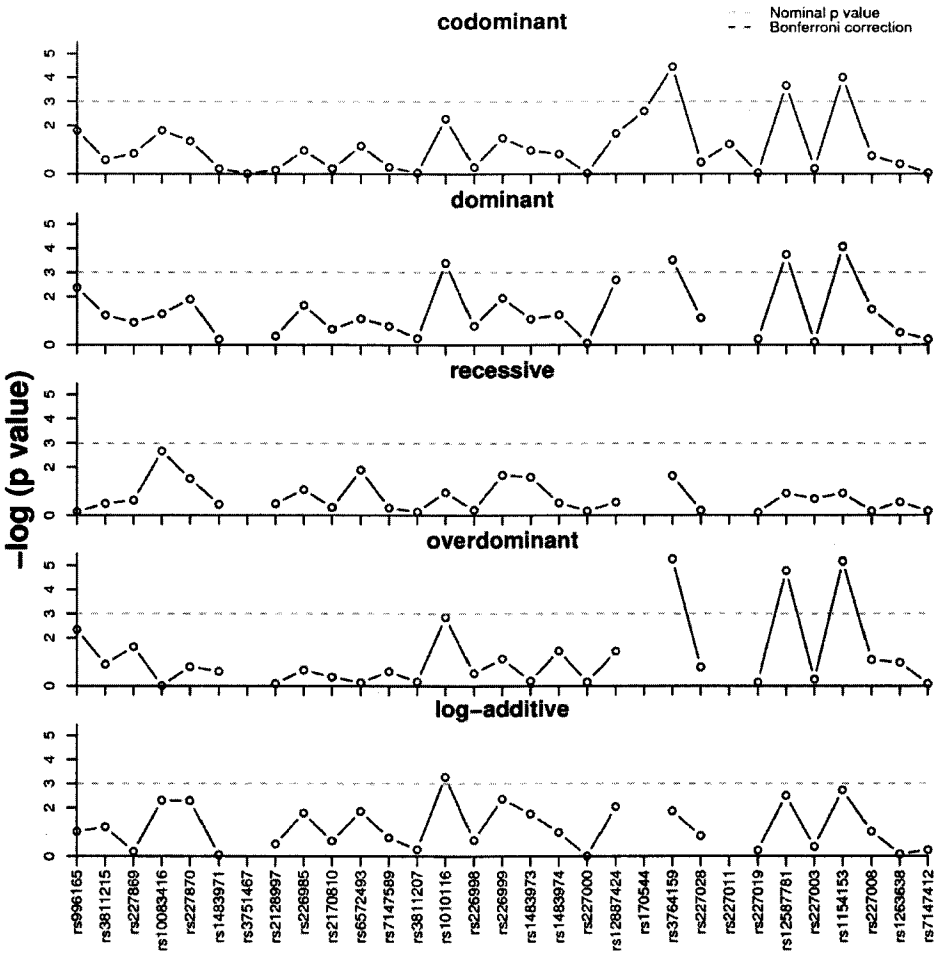

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0
2
3
4
5
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs3751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs1154153
rs227008
rs1263638
rs7147412

Figure 17

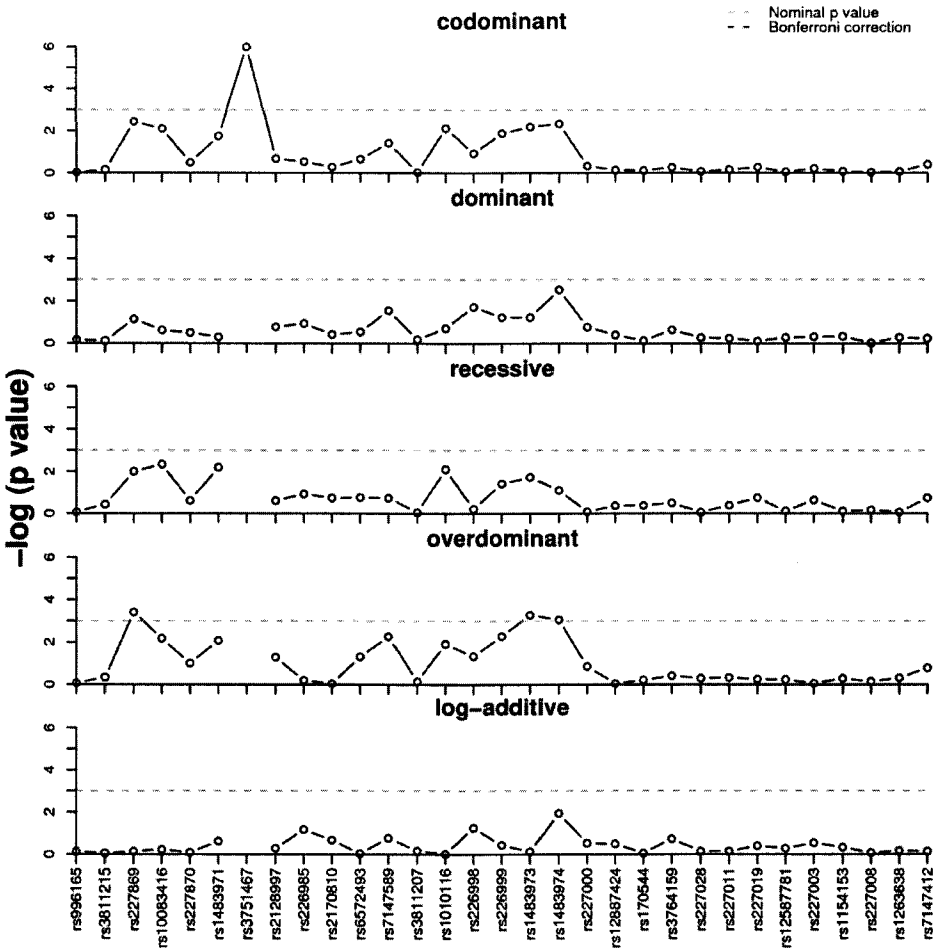

codominant
Nominal p value
Bonferroni correction
dominant
recessive
overdominant
log-additive
-log (p value)
0
2
4
6
rs996165
rs3811215
rs227869
rs10083416
rs227870
rs1483971
rs37751467
rs2128997
rs226985
rs2170810
rs6572493
rs7147589
rs3811207
rs1010116
rs226998
rs226999
rs1483973
rs1483974
rs227000
rs12887424
rs170544
rs3764159
rs227028
rs227011
rs227019
rs12587781
rs227003
rs11154153
rs227008
rs1263838
rs7147412

Figure 18

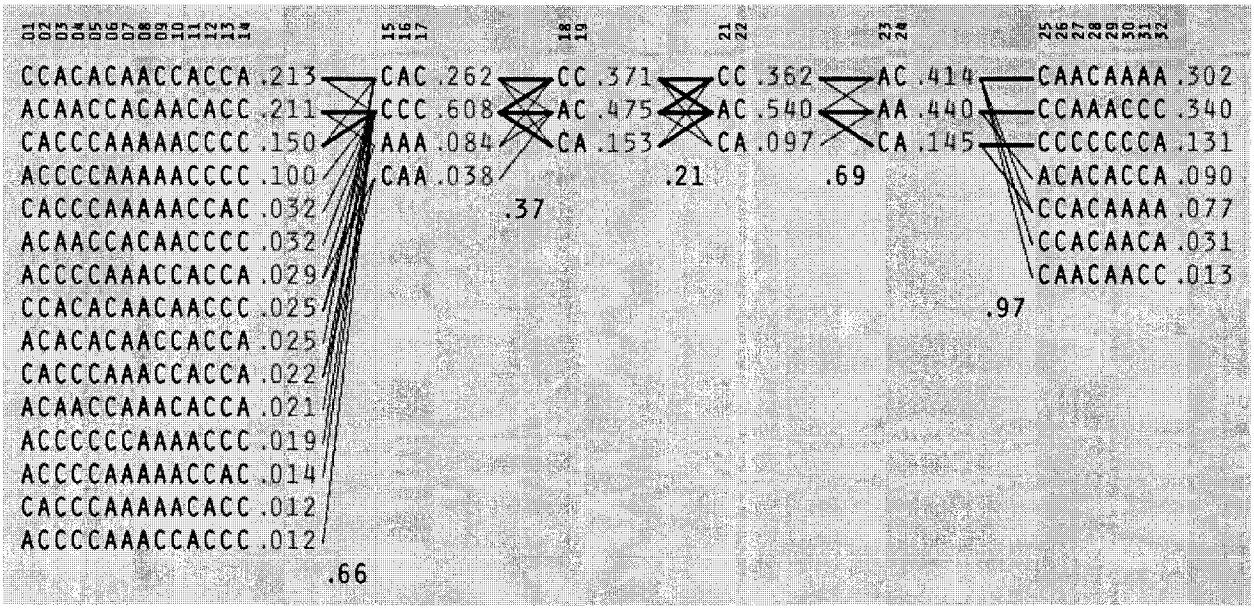

01 02 03 04 05 06 07 08 09 10 11 12 13 14
CCACACAACCACCA .213
ACAACCACAACACC .211
CACCCAAAAACCCC .150
ACCCCAAAAACCCC .100
CACCCAAAAACCAC .032
ACAACCACAACCCC .032
ACCCCAAACCACCA .029
CCACACAACAACCC .025
ACACACAACCACCA .025
CACCCAAACCACCA .022
ACAACCAAACACCA .021
ACCCCCCAAAACCC .019
ACCCCAAAAACCAC .014
CACCCAAAAACACC .012
ACCCCAAACCACCC .012
.66
15 16 17
CAC .262
CCC .608
AAA .084
CAA .038
.37
18 19
CC .371
AC .475
CA .153
.21
21 22
CC .362
AC .540
CA .097
.69
23 24
AC .414
AA .440
CA .145
25 26 27 28 29 30 31 32
CAACAAAA .302
CCAAACCC .340
CCCCCCCA .131
ACACACCA .090
CCACAAAA .077
CCACAACA .031
CAACAACC .013
.97

Figure 19

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 09/66980

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68; C07H 21/04 (2010.01)
USPC - 435/6, 536/23.1
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)
USPC: 435/91.2, 91.1, 536/22.1

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)
Electronic Databases Searched: PubWEST DB=PGPB,USPT,USOC,EPAB,JPAB; PLUR=NO; OP=ADJ, Google Scholar, Google Patent
Search Terms Used: asthma, TCR, T cell receptor, bacter$, coloniz$, grow$, polymorphism, SNP$, rs996165, rs3811215, rs227869, rsl0083416, rs227870, rs1483971, rs3751467, rs2128997, rs226985, rs2170810, rs6572493, rs7147589, [all othe

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	Vercelli. Discovering susceptibility genes for asthma and allergy. Nature Reviews: Immunology March 2008, volume 8, pg 169-182. entire document, esp: abstract, pg 172 section entitled "TH2-cell differentiation and effector function"; pg 170 section entitled "Known susceptibility genes in asthma and allergy"; pg 173 section entitled "Epithelial cells"; pg 177 section entitled "Gene-environment interactions in the population"; Fig. 1, Fig. 2, Fig. 5.	7-9 1-6
A	Moffatt et al. Single nucleotide polymorphism and linkage disequilibrium within the TCR a/g locus. Human Molecular Genetics, 2000, Vol 9, No. 7 pg 1011-1019. entire document, esp: abstact, pg 1011 section entitled "Introduction"; pg. 1012 section entitled "Measurement of Linkage Disequilibrium"; pg 1016 section entitled "Materials and Methods", Table 1.	1-6
A	Datasheet entitled: T-cell receptor alpha locus". PharmGKB. 2001. online at <https://www.pharmgkb.org/do/serve?objId=PA36704&objCls=Gene#tabview=tab0>	1-6

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
22 January 2010 (22.01.2010)	14 APR 2010
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

CORRECTED VERSION