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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 August 2009 (27.08.2009)

PCT

(10) International Publication Number
WO 2009/103992 A1

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

(21) **International Application Number**:
PCT/GB2009/000475

(22) **International Filing Date**:
19 February 2009 (19.02.2009)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
0803004.1 19 February 2008 (19.02.2008) GB

(71) **Applicants** *(for all designated States except US)*: **QUEEN MARY & WESTFIELD COLLEGE** [GB/GB]; Mile End Road, London E1 4NS (GB). **UNIVERSITAIR MEDISCH CENTRUM GRONINGEN** [NL/NL]; Hanzeplein 1, NL-9713 GZ Groningen (NL). **THE PROVOST FELLOWS AND SCHOLARS OF THE COLLEGE OF QUEEN ELIZABETH NEAR DUBLIN** [IE/IE]; College Green, Dublin 2 (IE). **GENOME RESEARCH LIMITED** [GB/GB]; 215 Euston Road, London W1N 2BE (GB).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **VAN HEEL, David** [GB/GB]; Centre for Gastroenterology, Institute of Cell & Molecular Science, 4 Newark Street, London E1 2AT (GB). **HUNT, Karen** [GB/GB]; Centre for Gastroenterology, Institute of Cell & Molecular Science, 4 Newark Street, London E1 2AT (GB). **WIJMENGA, Cisca** [NL/NL]; Universitair Medisch Centrum Groningen, Hanzeplein 1, NL-9713 GZ Groningen (NL). **MCMANUS, Owen, Ross** [IE/IE]; School of Medicine, Trinity College, Dublin 2 (IE). **DELOUKAS, Panagiotis** [GR/GB]; Genome Research Limited, 215 Euston Road, London W1N 2BE (GB).

(74) **Agents**: **BRAMLEY, Sarah, Jane** et al.; Kilburn & Strode LLP, 20 Red Lion Street, London WC1R 4PJ (GB).

(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, 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, 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, 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))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2009/103992 A1

(54) **Title**: GENETIC VARIATION ASSOCIATED WITH COELIAC DISEASE

(57) **Abstract**: The present invention provides a method of diagnosing coeliac disease, said method comprising analysing a sample of nucleic acid from a human subject to determine the presence or absence of one or more single nucleic polymorphisms (SNPs) in one or more human chromosomal regions selected from the group consisting of Iq31, 2ql l-2ql2, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24.

GENETIC VARIATION ASSOCIATED WITH COELIAC DISEASE

The present invention relates to the diagnosis of coeliac disease, and in particular to human chromosomal regions and specific single nucleotide polymorphisms (SNPs) within those regions which are associated with coeliac disease. These human chromosomal regions and SNPs can thus be used to predict the occurrence of coeliac disease in a patient.

Coeliac disease (alternative spelling: celiac disease) is a common heritable condition with a prevalence of approximately 1% in Western populations. In coeliac disease, HLA-DQ2 or –DQ8 restricted T cell responses occur to dietary proteins (glutens) in cereals such as wheat, rye and barley. This leads to small intestinal inflammation and intestinal villous atrophy, with consequent clinical features such as chronic diarrhoea and fatigue. Sufferers have serum antibodies to gluten and show delayed hypersensitivity to gluten.

Several recent immunological advances have identified dominant epitopes, the role of tissue transglutaminase, and the crystal structures of DQ molecules binding wheat peptides as factors predisposing to coeliac disease. However, the primary factors, both environmental and genetic, predisposing to disease in the ~30% of the population carrying susceptible HLA-DQ types are mostly unknown.

A single nucleotide polymorphism (SNP) is a DNA sequence variation which occurs when a single nucleotide of the genome differs between members of a species or between different chromosomes in an individual. When a SNP occurs in a protein coding sequence, it may give rise to a change in amino acid and thus affect the polypeptide sequence encoded. Such SNPs are termed non-synonymous, whereas synonymous SNPs do not lead to a change in the encoded polypeptide sequence.

The present inventors have previously carried out a genome-wide association study and identified a number of SNPs in the chromosomal region harbouring the *IL2* and *IL21* genes which are associated with susceptibility to coeliac disease in humans (van

Heel *et al.*, Nature Genetics **39**, 827-829, 2007). This indicates that genetic variation in this region of chromosome 4q27 predisposes to coeliac disease. However, it is expected that the *IL2-IL21* locus explains less than 1% of the increased familial risk for coeliac disease.

There thus exists a need in the art to identify methods for accurately identifying individuals at risk for coeliac disease.

The present inventors have now identified seven further human chromosomal regions, and a number of SNPs within these regions, which correlate with increased risk for coeliac disease.

According to a first aspect of the invention there is thus provided a method of diagnosing coeliac disease, said method comprising analysing a sample of nucleic acid from a human subject to determine the presence or absence of one or more SNPs in one or more human chromosomal regions selected from the group consisting of 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24.

The present invention meets a hitherto unmet need for a method of accurately identifying individuals at risk for coeliac disease. The invention advantageously provides a number of SNPs which are associated with coeliac disease and can thus be used individually or in combination to determine whether an individual is at risk for or suffering from coeliac disease.

In humans, the diploid chromosome number is 46, with 23 chromosomes being inherited from each parent via the sperm and the egg. Homologous chromosomes form pairs with one chromosome from each parent. Human cells thus contain 23 pairs of chromosomes. Of these, the autosomes are numbered from 1 to 22, with the other pair being the sex chromosomes, either XX (female) or XY (male).

The International System for Human Cytogenetic Nomenclature (ISCN) is fixed by the Standing Committee on Human Cytogenetic Nomenclature. The basic

terminology for banded chromosomes was decided at a meeting in Paris in 1971, and is therefore often referred to as the Paris nomenclature.

Each human chromosome has a short arm and a long arm. The short arm is designated p and the long arm is designated q. Each arm of the chromosome is divided into regions labelled p1, p2, p3 etc., and q1, q2, q3, etc., counting outwards from the centromere. Regions are delimited by specific landmarks, which are consistent and distinct morphological features, such as the ends of the chromosome arms, the centromere and certain bands. Regions are further divided into bands.

The seven further human chromosomal regions that the present inventors have now identified as being associated with coeliac disease are 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24.

A single nucleotide polymorphism (SNP) is a DNA sequence variation which occurs when a single nucleotide of the genome differs between members of a species or between different chromosomes in an individual. Typically, each SNP has only two different alleles. In other words, for each SNP there is typically only two different nucleotide variants. SNPs are identified herein using by their GenBank accession number, as available in the National Center for Biotechnology Information (NCBI) database. The GenBank accession number "rs....." indicates the position of a SNP within the human genome and the sequence surrounding the SNP, and may be readily identified by a person skilled in the art using the NCBI database. The specific sequences corresponding to the rs number of the SNP within the NCBI database may change over time, but this will be indicated on the NCBI database and will be readily identified by a person skilled in the art.

The sequences including the SNPs for use in the present invention are also identified herein by their sequence identifiers, SEQ ID NOs: 1 to 21. These sequences are shown in Figure 7.

The nucleotide sequences of SEQ ID NOs: 1 to 21 are polymorphic sequences. A polymorphic sequence is a polynucleotide sequence including a polymorphic site representing a SNP. The sequences set out in SEQ ID NOs: 1 to 21 include both alleles of the SNP. The SNP alleles are shown in square brackets in the sequences of SEQ ID NOs: 1 to 21 shown in Figure 7.

The SNPs within the seven new human chromosomal regions that the present inventors have identified as being associated with coeliac disease are rs2816316 (SEQ ID NO: 1), rs13015714 (SEQ ID NO: 2), rs917997 (SEQ ID NO: 3), rs6441961 (SEQ ID NO: 4), rs17810546 (SEQ ID NO: 5), rs9811792 (SEQ ID NO: 6), rs9851967 (SEQ ID NO: 7), rs13076312 (SEQ ID NO: 8), rs1464510 (SEQ ID NO: 9), rs1559810 (SEQ ID NO: 10), rs1738074 (SEQ ID NO: 11), rs3184504 (SEQ ID NO: 12) and rs653178 (SEQ ID NO: 13). Accordingly, in the methods of the invention, the one or more SNPs are typically selected from the group consisting of the SNPs present in the sequences of SEQ ID NOs: 1 to 13.

Within these 13 SNPs, the following alleles are indicative of coeliac disease (i.e. more common in coeliac disease than in control subjects):

T in SEQ ID NO: 1 (rs2816316)
G in SEQ ID NO: 2 (rs13015714)
A in SEQ ID NO: 3 (rs917997)
T in SEQ ID NO: 4 (rs6441961)
G in SEQ ID NO: 5 (rs17810546)
C in SEQ ID NO: 6 (rs9811792)
C in SEQ ID NO: 7 (rs9851967)
T in SEQ ID NO: 8 (rs13076312)
T in SEQ ID NO: 9 (rs1464510)
T in SEQ ID NO: 10 (rs1559810)
A in SEQ ID NO: 11 (rs1738074)
T in SEQ ID NO: 12 (rs3184504)
G in SEQ ID NO: 13 (rs653178)

Accordingly, the risk of coeliac disease is increased in individuals possessing one of more of the above alleles. In other words, the risk of coeliac disease is increased in individuals who are homozygous for one or more of the above alleles or who are heterozygous.

The human chromosomal regions now identified by the inventors as being associated with coeliac disease are in addition to the region 4q27 that was previously identified by the inventors (van Heel *et al.*, *supra*). Accordingly, in some embodiments of the invention, the methods of the invention further comprise analysing a sample of nucleic acid from the human subject to determine the presence or absence of one or more SNPs in the human chromosomal region 4q27.

The SNPs within the 4q27 region that were identified as being associated with coeliac disease are rs11938795 (SEQ ID NO: 14), rs13151961 (SEQ ID NO: 15), rs13119723 (SEQ ID NO: 16), rs11734090 (SEQ ID NO: 17), rs7684187 (SEQ ID NO: 18), rs12642902 (SEQ ID NO: 19), rs6822844 (SEQ ID NO: 20) and rs6840978 (SEQ ID NO: 21). Accordingly, in these embodiments of the invention, the one or more SNPs are typically selected from the group consisting of the SNPs present in the sequences of SEQ ID NOs: 14 to 21.

Within these 8 SNPs, the following alleles are indicative of coeliac disease (i.e. more common in coeliac disease than in control subjects):

T in SEQ ID NO: 14 (rs11938795)
A in SEQ ID NO: 15 (rs13151961)
A in SEQ ID NO: 16 (rs13119723)
T in SEQ ID NO: 17 (rs11734090)
A in SEQ ID NO: 18 (rs7684187)
G in SEQ ID NO: 19 (rs12642902)
G in SEQ ID NO: 20 (rs6822844)
C in SEQ ID NO: 21 (rs6840978)

Accordingly, the risk of coeliac disease is increased in individuals possessing one or more of the above alleles. In other words, the risk of coeliac disease is increased in individuals who are homozygous for one or more of the above alleles or who are heterozygous.

The region 4q27 and the SNPs rs11938795 (SEQ ID NO: 14), rs13151961 (SEQ ID NO: 15), rs13119723 (SEQ ID NO: 16), rs11734090 (SEQ ID NO: 17), rs7684187 (SEQ ID NO: 18), rs12642902 (SEQ ID NO: 19), rs6822844 (SEQ ID NO: 20) and rs6840978 (SEQ ID NO: 21) can therefore be used in the methods of the invention in addition to the seven new human chromosomal regions and 13 new SNPs which have been identified by the inventors.

The SNPs for use in the present invention and their cytogenetic locations are set out in the following Table:

Table 1

Gen Bank accession no. of SNP in NCBI	**Chromosome region**	**Sequence identifier**
rs2816316	1q31	SEQ ID NO: 1
rs13015714	2q11 - 2q12	SEQ ID NO: 2
rs917997	2q11 - 2q12	SEQ ID NO: 3
rs6441961	3p21	SEQ ID NO: 4
rs17810546	3q25 - 3q26	SEQ ID NO: 5
rs9811792	3q25 - 3q26	SEQ ID NO: 6
rs9851967	3q28	SEQ ID NO: 7
rs13076312	3q28	SEQ ID NO: 8
rs1464510	3q28	SEQ ID NO: 9
rs1559810	3q28	SEQ ID NO: 10
rs1738074	6q25	SEQ ID NO: 11
rs3184504	12q24	SEQ ID NO: 12

Gen Bank accession no. of SNP in NCBI	Chromosome region	Sequence identifier
rs653178	12q24	SEQ ID NO: 13
rs11938795	4q27	SEQ ID NO: 14
rs13151961	4q27	SEQ ID NO: 15
rs13119723	4q27	SEQ ID NO: 16
rs11734090	4q27	SEQ ID NO: 17
rs7684187	4q27	SEQ ID NO: 18
rs12642902	4q27	SEQ ID NO: 19
rs6822844	4q27	SEQ ID NO: 20
rs6840978	4q27	SEQ ID NO: 21

The present invention provides a method of diagnosing coeliac disease, said method comprising analysing a sample of nucleic acid from a human subject to determine the presence or absence of one or more SNPs in one or more human chromosomal regions selected from the group consisting of 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24.

Accordingly, in one embodiment of the invention, the one or more SNPs are selected from the group consisting of the SNPs present in the following sequences: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12 and SEQ ID NO: 13.

As set out above, the seven new human chromosomal regions and 13 new SNPs within those chromosomal regions which have been identified by the inventors as being associated with coeliac disease can be used in the methods of the invention in combination with the previously reported genetic region 4q27 and the SNPs within it.

In this embodiment, the method of the invention further comprises analysing a sample of nucleic acid from the human subject to determine the presence or absence of one or

more SNPs in the human chromosomal region 4q27. In this embodiment of the invention, the presence or absence of one or more SNPs in one or more human chromosomal regions selected from the group consisting of 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24 is determined, in addition to determining the presence or absence of one or more SNPs in the human chromosomal region 4q27.

Typically, in this embodiment of the invention, the human chromosomal region 4q27 comprises one or more SNPs selected from the group consisting of the SNPs present in the following sequences: SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20 and SEQ ID NO: 21.

The present invention can also be used in combination with previously reported HLA typing to diagnose coeliac disease. HLA typing can be done, for example, by serology or genetic methods well known in the art, or by genotyping an HLA marker, for example the marker rs2187668 (van Heel *et al., supra*).

In the methods of the present invention, one or more SNPs selected from the group consisting of the SNPs present in SEQ ID NOs: 1 to 13 are used. Typically, more than one of these SNPs is used. In this embodiment of the invention, any combination of the listed SNPs can be used. For example, a combination of any 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or even all 13 of these SNPs is used in a method of the invention. Each additional SNP provides additional information on the risk of coeliac disease.

Typically, at least one SNP from each chromosomal region is used, i.e. at least one SNP from each of the chromosomal regions 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24 which have been identified by the inventors as being associated with coeliac disease. In this embodiment, the SNPs present in SEQ ID NO: 1, SEQ ID NO: 2 and/or SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 and/or SEQ ID NO: 6, SEQ ID NO: 7 and/or SEQ ID NO: 8 and/or SEQ ID NO: 9 and/or SEQ ID NO: 10, SEQ ID NO: 11, and SEQ ID NO: 12 and/or SEQ ID NO: 13 are typically used in combination.

For example, the SNPs present in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 9, SEQ ID NO: 11 and SEQ ID NO: 13 are used in combination.

In one embodiment of the present invention, one or more SNPs selected from the group consisting of the SNPs present in SEQ ID NOs: 14 to 21 is used in addition to one or more of the SNPs present in SEQ ID NOs: 1 to 13. In this embodiment, any one of the SNPs present in SEQ ID NOs: 14 to 21 or a combination of any 2, 3, 4, 5, 6, 7 or even all 8 of these SNPs is used in addition to one or more of the SNPs present in SEQ ID NOs: 1 to 13. Any combination of the SNPs present in SEQ ID NOs: 14 to 21 can thus be used together with any combination of the SNPs present in SEQ ID NOs: 1 to 13. Accordingly, any 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or even all 21 of the SNPs present in SEQ ID NOs: 1 to 21 can be used in a method of the invention. Each additional SNP provides additional information on the risk of coeliac disease.

Typically, at least one SNP from each chromosomal region is used, i.e. at least one SNP from each of the chromosomal regions 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 4q27, 6q25 and 12q24. In this embodiment, the SNPs present in SEQ ID NO: 1, SEQ ID NO: 2 and/or SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 and/or SEQ ID NO: 6, SEQ ID NO: 7 and/or SEQ ID NO: 8 and/or SEQ ID NO: 9 and/or SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12 and/or SEQ ID NO: 13, SEQ ID NO: 14 and/or SEQ ID NO: 15 and/or SEQ ID NO: 16 and/or SEQ ID NO: 17 and/or SEQ ID NO: 18 and/or SEQ ID NO: 19 and/or SEQ ID NO: 20 and/or SEQ ID NO: 21 are typically used in combination.

For example, the SNPs present in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13 and SEQ ID NO: 20 are used in combination.

The above-mentioned combinations of SNPs can also be used in combination with previously reported HLA typing to diagnose coeliac disease. HLA typing can be

done, for example, by serology or genetic methods well known in the art, or by genotyping an HLA marker, for example the marker rs2187668 (van Heel *et al., supra*).

In the methods of the invention, a sample of nucleic acid from a human subject is analysed to determine the presence or absence of one or more SNPs in one or more of the above-mentioned human chromosomal regions, typically to determine whether one or more of the above-mentioned SNPs is present in the nucleic acid sample.

The sample of nucleic acid used in the invention is typically a sample of DNA or RNA. DNA includes cDNA synthesized from mRNA.

The sample of nucleic acid can be derived from any biological sample which contains the subject's nucleic acid. For example, the biological sample can be a sample of whole blood, plasma, serum, urine, sputum or lymph and can be obtained by any suitable means.

The genetic regions identified by the inventors are within the human genome and therefore the subject is a human subject.

The methods of the present invention are typically carried out on a sample of nucleic acid that has previously been taken from a human subject, and thus the taking of the nucleic acid sample does not typically form part of the methods of the invention. In some embodiments of the invention, however, the method also comprises taking the nucleic acid sample from the subject, for example by a cheek swab, by taking a urine sample or by taking a blood sample.

In some embodiments of the invention, it will be necessary to compare the sample taken from the subject to a control sample. The control sample is typically a nucleic acid sample taken from a subject who is known not to be suffering from coeliac disease. However, it may not be necessary to compare the sample to a control sample,

since the SNP, and in particular the disease-associated allele, can be detected on its own, for example by sequencing of the nucleic acid sample.

To determine the presence or absence of one or more SNPs in one or more of the above-mentioned human chromosomal regions, and typically to determine whether one or more of the above-mentioned SNPs is present in the nucleic acid sample, the nucleic acid will typically be isolated from the subject.

Nucleic acid isolation can be carried out using any appropriate method known in the art. For example, DNA can be directly purified from tissues or cells or a specific region can be amplified using, for example, the polymerase chain reaction (PCR) and isolated.

The isolated nucleic acid is then analysed to determine the presence or absence of one or more SNPs in one or more of the above-mentioned human chromosomal regions.

In one embodiment, the nucleic acid is analysed by sequencing of the isolated nucleic acid. Again, sequencing of the isolated nucleic acid can be carried out by any suitable method known in the art.

In the methods of the present invention, one or more SNPs is detected. SNPs can be detected by a variety of methods known in the art. For example, the nucleic acid can be sequenced to determine whether a disease-causing allele is present. Hybridization-based methods can also be used to identify SNPs. The nucleotides of the polymorphic site can be identified by hybridizing the DNA with a probe containing the sequence of the SNP site or a complementary probe thereof, and examining the degree of hybridization. Enzyme-based methods such as restriction fragment length polymorphism, allele-specific polymerase chain reaction (PCR) and primer extension (for example the Infinium assay used in the Examples herein) can also be used.

The present invention provides a method for the diagnosis of coeliac disease. In one embodiment, the invention is used to diagnose individuals with one or more medical

symptoms of coeliac disease, such as diarrhoea and/or fatigue. In other embodiments, the invention is used to diagnose individuals who do not currently have medical symptoms of coeliac disease, but are related to a known coeliac disease sufferer, or are concerned about the risk of coeliac disease for some other reason. These embodiments of the invention allow preventative measures to be taken to avoid the subject suffering from symptoms of coeliac disease, for example by modification of the subject's diet. The invention can also be used to diagnose coeliac disease in cases where diagnosis has previously been unclear.

The invention provides methods of diagnosing coeliac disease. The methods of the invention can be used to diagnose coeliac disease in a human subject if it is found that there are one or more SNPs in one or more human chromosomal regions selected from the group consisting of 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24 in a sample of nucleic acid from that subject. Typically, the presence of one or more SNPs in one or more of said human chromosomal regions is indicative of coeliac disease.

Typically, the one or more SNPs are selected from the group consisting of the SNPs present in SEQ ID NOs: 1 to 13. In this embodiment of the invention, the presence of one or more of the following alleles is indicative of coeliac disease:

T in SEQ ID NO: 1 (rs2816316)
G in SEQ ID NO: 2 (rs13015714)
A in SEQ ID NO: 3 (rs917997)
T in SEQ ID NO: 4 (rs6441961)
G in SEQ ID NO: 5 (rs17810546)
C in SEQ ID NO: 6 (rs9811792)
C in SEQ ID NO: 7 (rs9851967)
T in SEQ ID NO: 8 (rs13076312)
T in SEQ ID NO: 9 (rs1464510)
T in SEQ ID NO: 10 (rs1559810)
A in SEQ ID NO: 11 (rs1738074)
T in SEQ ID NO: 12 (rs3184504)

G in SEQ ID NO: 13 (rs653178)

In this embodiment of the invention, at least one SNP from each chromosomal region is typically used, and in this embodiment the presence of the following alleles is indicative of coeliac disease: T in SEQ ID NO: 1, G in SEQ ID NO: 2 and/or A in SEQ ID NO: 3, T in SEQ ID NO: 4, G in SEQ ID NO: 5 and/or C in SEQ ID NO: 6, C in SEQ ID NO: 7 and/or T in SEQ ID NO: 8 and/or T in SEQ ID NO: 9 and/or T in SEQ ID NO: 10, A in SEQ ID NO: 11, and T in SEQ ID NO: 12 and/or G in SEQ ID NO: 13.

In some embodiments of the invention, the method further comprises analysing a sample of nucleic acid from the human subject to determine the presence or absence of one or more SNPs in the human chromosomal region 4q27. In this embodiment of the invention, the presence of one or more SNPs in this chromosomal region is indicative of coeliac disease.

Typically, in this embodiment of the invention, the presence of one or more of the following alleles is indicative of coeliac disease:

T in SEQ ID NO: 14 (rs11938795)
A in SEQ ID NO: 15 (rs13151961)
A in SEQ ID NO: 16 (rs13119723)
T in SEQ ID NO: 17 (rs11734090)
A in SEQ ID NO: 18 (rs7684187)
G in SEQ ID NO: 19 (rs12642902)
G in SEQ ID NO: 20 (rs6822844)
C in SEQ ID NO: 21 (rs6840978)

In this embodiment, at least one of the above alleles is detected in combination with at least one of the above-mentioned alleles of SEQ ID NOs: 1 to 13.

In a second aspect, the present invention provides a method of testing for coeliac disease, said method comprising analysing a sample of nucleic acid from a human

subject to determine the presence or absence of one or more SNPs in one or more human chromosomal regions selected from the group consisting of 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24.

In a third aspect, the present invention provides a method of identifying one or more SNPs in a sample of nucleic acid, said method comprising analysing said sample of nucleic acid to determine the presence or absence of one or more SNPs in one or more human chromosomal regions selected from the group consisting of 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24.

In a fourth aspect, the present invention provides a method of diagnosing coeliac disease, said method comprising analysing a sample of nucleic acid from a subject to determine the presence or absence of one or more SNPs in one or more human chromosomal regions selected from the group consisting of 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24, wherein the presence of one or more SNPs in one or more of said human chromosomal regions is indicative of coeliac disease;
thereby diagnosing coeliac disease in the subject.

Preferred features for the second, third and fourth aspects of the invention are as for the first aspect *mutatis mutandis.*

The invention will now be further described by way of reference to the following Examples and Figures which are provided for the purposes of illustration only and are not to be construed as limiting on the invention. Reference is made to a number of Figures, in which:

Figure 1 shows quantile-quantile (Q-Q) plots for association results in follow-up samples. Figure 1a is a Q-Q plot of association results (Cochran-Mantel-Haenszel test) for 1020 non-*HLA* SNPs in UK2, IRISH and DUTCH follow-up samples. Data points in light grey indicate SNPs shown in Table 2 with *P* overall $< 5x10^{-7}$ in all samples including UKGWAS. Straight line indicates expected results under null hypothesis. Figure 1b is a Q-Q plot of

residual association results (Cochran-Mantel-Haenszel test) for 992 non-*HLA* SNPs, excluding 28 SNPs mapping to the eight coeliac associated regions described in Table 2, in UK2, IRISH and DUTCH follow-up samples.

Figure 2 shows linkage disequilibrium structure and association results for eight non-*HLA* coeliac disease associated regions. Chromosomal positions based on NCBI build 36 coordinates, showing Ensembl (release 48) genes. Allele count *P* values are shown for SNPs analysed in UKGWAS samples (diamonds), and Cochran-Mantel-Haenszel association test *P* values (circles) for SNPs analysed in all UKGWAS, UK2, IRISH and DUTCH samples.

Figure 3 shows correlation of rs917997 genotype with whole blood *cis* *IL18RAP* mRNA expression. Expression levels (bars show group means) were determined in samples from gluten-free diet treated coeliac individuals.

Figure 4 shows cluster plots for most significantly associated SNPs. R, theta plots for the most significantly associated SNP from each of eight loci (Table 2) from Infinium (UKGWAS) and GoldenGate (UK2, Irish, Dutch) datasets. Genotype call rates (all 7238 samples) were 99.88% for rs917997 and rs1464510, 99.97% for rs1738074 and rs653178, and complete (no missing data) for rs2816316, rs6441961, rs17810546 and rs6822844.

Figure 5 shows expression of putative candidate genes in small intestinal tissue. *IL1RL1* (two probes), *SLC9A4*, *CCR1*, *CCR3* (Illumina probe 1980750), *SCHIP1*, *IL2*, *IL21*, *TAGAP* (two probes) and *Tenr* were not detected above background in intestinal tissue and are not shown. NC: healthy normal controls, M0: treated coeliac disease with (Marsh 0) normal intestinal histology, MIII: untreated coeliac disease with villous atrophy (Marsh III).

Figure 6 shows expression profiling of T cell subsets in murine intestine, thymus and spleen. Representative (n>3) semi-quantitative RT-PCR

experiments for RGS-1, CCR5, CCR9 and β-actin on small intestinal intraepithelial lymphocytes (IELs); TCRγδ$^+$CD4$^-$CD8$^-$ [γδ DN]; TCRγδ$^+$CD4$^-$ CD8 αα$^+$ [γδ CD8αα]; TCRαβ$^+$CD4$^-$CD8 αα$^+$ [αβ CD8αα]; TCRαβ$^+$CD4$^-$ CD8αβ$^+$ [αβ CD8αβ]; TCR γδ$^+$ cells [γδ] from thymus and spleen, CD4$^+$CD8αβ$^+$ [DP], TCRαβ$^+$CD4$^+$CD8αβ$^-$ [SP CD4] and TCRαβ$^+$CD4$^-$ CD8αβ$^+$ [SP CD8] from thymus, and TCRαβ$^+$CD4$^-$CD8αβ$^+$ [αβ CD8] from spleen. cDNA was normalised to β -actin. Negative control [-] is reaction in absence of cDNA. Positive control [+] is reaction from mixed bulk IEL and splenic cDNA in excess.

Figure 7 shows the nucleotide sequences of the SNPs for use in the invention. The sequences are shown with reference to the NCBI accession number (rs.....)

Examples

INTRODUCTION

To identify additional coeliac disease susceptibility genes, the inventors recently tested 310,605 SNPs in a genome wide association study of 778 coeliac cases and 1,422 population controls from the United Kingdom (UKGWAS), using the Illumina HumanHap300 BeadChip (van Heel *et al.*, *supra*). The only SNP outside the *HLA* region demonstrating genome-wide significance was rs13119723 on 4q27, located in a ~500 kb block of linkage disequilibrium (LD) containing the *IL2* and *IL21* genes. Independent replication of SNPs from the *IL2-IL21* region was established in both Dutch and Irish collections of coeliac patients and controls. It is estimated, using the current markers, that the *IL2-IL21* region explains less than 1% of the increased familial risk to coeliac disease. Since a greater number of significantly associated SNPs in the UKGWAS was observed than would be expected by chance, the inventors proceeded to study >1,000 of the most significant UKGWAS association results in a further 1,643 coeliac cases and 3,406 controls from three independent European coeliac disease collections. This two-stage strategy, involving a joint

analysis of all data, substantially reduces the genotyping requirements versus performing whole genome genotyping on all samples and has been shown to maintain sufficient statistical power.

METHODS

Subjects. Detailed characteristics of UKGWAS, IRISH and DUTCH samples are provided in Table 3 below, and the inventors' previously published study (van Heel *et al.*, *supra*).

All subjects were of white Northern European origin. UK2 subjects were recruited similarly to the previously described UKGWAS subjects (van Heel *et al.*, *supra*) - cases were recruited from hospital clinics and controls from the British 1958 Birth Cohort, except for n=374 cases and n=176 unrelated controls recruited direct through Coeliac UK advertisement. For DNA procedures see Supplementary Methods below. It was noted that *HLA-DQ2.5cis* genotype frequencies (Table 3 below, inferred from rs2187668) were similarly high across the coeliac collections (carriage rate range 87.4% to 89.1%), and similarly low across the control populations (carriage rate range 25.5% to 33.8%), suggesting broadly comparable phenotypic ascertainment across the four collections. Informed consent was obtained from all subjects. Ethical approval was from Oxfordshire REC B or East London and the City REC 1 (UKGWAS, UK2), the Medical Ethical Committee of the University Medical Center Utrecht (DUTCH), and the Institutional Ethics Committee of St James's Hospital (IRISH).

Marker selection and genotyping. Single non-*HLA* SNPs with two-tailed allele count χ^2 test $P<0.00275$, or if nsSNPs $P<0.01$, were selected from the inventors' published coeliac disease genome wide association study (van Heel *et al.*, *supra*) of 310,605 post-quality control markers (Hardy-Weinberg equilibrium P>0.0001 in controls). Genotyping data and clustering of SNP genotypes was managed in BeadStudio. Samples with <95% call rate over 1025 SNPs, SNPs with <95% call rate over remaining samples, SNPs with poor amplification or poor genotype cloud clustering were excluded.

Quality control steps. Using the 1025 SNP dataset, pairwise comparisons of identity by-descent were made for all samples (UKGWAS, UK2, IRISH, DUTCH) using PLINK (Purcell *et al.*, Am J Hum Genet **81**, 559-75, 2007). A higher proportion of 1st degree relatives in the all sample dataset (98 pairs) was detected than in the initial UK coeliac GWAS (11 pairs), and therefore in the current analyses the lowest call rate sample from each pair of 1° relatives from the entire study dataset was excluded. Minor but insignificant changes are therefore present in the UKGWAS dataset results compared to the inventors' previous publication (van Heel *et al.*, *supra*).

Potential ethnic outlier samples were excluded (n=3, outliers had previously been excluded from the UKGWAS dataset) using the nearest neighbour allele sharing method in PLINK (samples with Z scores <-3 with >1 of 5 nearest neighbours excluded). A filter for SNP selection was applied for follow up GoldenGate genotyping based on Hardy Weinberg (HWE) equilibrium P>0.0001 in Infinium genotyped controls from the UKGWAS. HWE P values for each of the coeliac disease associated SNPs and 5 HLA tag markers in each of the follow-up UK2, IRISH and Dutch collections are shown in Table 4. Because of the prior filter step, no follow-up SNPs were specifically excluded based on HWE analysis. All of the top association findings were in HWE in controls.

Table 4 shows detailed association results. Table 4 indicates the alleles for the SNPs associated with coeliac disease. For rs2816316 (SEQ ID NO: 1), rs13015714 (SEQ ID NO: 2), rs6441961 (SEQ ID NO: 4), rs9811792 (SEQ ID NO: 6), rs9851967 (SEQ ID NO: 7), rs13076312 (SEQ ID NO: 8), rs1464510 (SEQ ID NO: 9), rs1559810 (SEQ ID NO: 10), rs3184504 (SEQ ID NO: 12), rs11938795 (SEQ ID NO: 14), rs11734090 (SEQ ID NO: 17), rs6822844 (SEQ ID NO: 20) and rs6840978 (SEQ ID NO: 21) the alleles shown in Table 4 do not correspond to the disease associated alleles specified in the description, claims and Figures herein. This is because the double stranded DNA sequence can be read in either strand direction. Therefore if a SNP is described as, for example, having A and C alleles, this identical to describing it as having T and G alleles. This is due to the complementarity of nucleic acid based

pairing, in which the complement of A is T and vice versa and the complement of C is G and vice versa, which is well understood by a person skilled in the art. The alleles for the SNPs shown in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 20 and SEQ ID NO: 21 are shown in this way in Table 4 as the complementary strand of the sequence shown in Figure 7. However, the alleles are in fact the same.

Genotype statistical analysis. Cochran-Mantel-Haenszel allele count chi-squared association tests were performed using PLINK (Purcell *et al.*, *supra*) with 4 clusters: UKGWAS (Infinium assay), UK2 (GoldenGate assay), IRISH and DUTCH collections. All *P*-values are two-tailed. Minimal evidence for bias was observed in the Cochran-Mantel-Haenszel test statistics due to population stratification or other factors (see λ_{GC} values ≤ 1.03 in RESULTS), and present uncorrected statistics. To test for differences in association across the datasets as manifested by heterogeneity of odds ratios at disease-associated regions, the Breslow-Day test was applied to SNPs in Table 2 (below). These SNPs, and *HLA-DQ2.5cis* (inferred by rs2187668 genotype), were also evaluated for possible epistatic (gene-gene) interaction in predisposing to coeliac disease. Breslow-Day and epistatic interaction tests were carried out as implemented in PLINK. Epistatic interactions were also tested for using the model of Howard *et al* (Am J Hum Genet **70**, 230-6, 2002). For the PLINK analysis of interaction, genotypes at the SNPs are evaluated as allele "doses" (e.g. 0, 1, or 2) whereas for the model of Howard *et al* (supra) heterozygotes and homozygotes for the higher-risk allele are pooled together and compared against the lower-risk homozygote genotype. The PLINK interaction analysis and our analysis of the model of Howard *et al* (implemented in R statistical software, www.r-project.org) both tested for statistically significant departure from the model of log-additive odds ratios as evaluated by logistic regression.

The possible influence of stratification was also investigated by an alternative principle components approach in which each dataset was separately evaluated by

EIGENSTRAT software (Price *et al.* Nat Genet **38**, 904-9, 2006). After eliminating SNPs in the 9 regions showing association with coeliac disease, EIGENSTRAT was applied to the remaining SNPs (drawn from 310,605 SNPs for UKGWAS, or from 1025 SNPs for UK2, IRISH, DUTCH) to determine eigenvectors that correct for possible population stratification in cases and controls. For each case-control set, EIGENSTRAT calculates the Armitage Trend test chi-square statistic for each genotyped SNP and then applies a SNP-specific correction based on the eigenvectors that individually adjusts the value of Armitage chi-square. Following EIGENSTRAT correction, we also corrected with the residual genomic control factor for each collection and then combined the post-correction chi-squares across the 4 studies according to the meta-analysis method of Stouffer (Rosenthal, R. *Meta-Analytic Procedures for Social Research*, Sage Publications Inc., 1991). This enabled us to achieve a corrected, study-wide Z-score and *P* value for each SNP that can be compared to the Cochran-Mantel-Haenszel *P* value and shows that all 9 regions remain significant at $P < 5 \times 10^{-7}$ after correction by this alternative approach (Table 4). The corrected *P* values for SNPs not in the 9 coeliac-associated regions also generate a QQ-plot that shows no evidence of overdispersion and is almost identical to the QQ-plot (Figure 1b) of the Cochran-Mantel-Haenszel *P* values.

Calculation of familial clustering for each region is described in Supplementary Methods, below.

***P* value threshold for genomewide statistical significance.** For statistical significance of disease association results in the combined genome scan (UKGWAS) and follow-up datasets (UK2, DUTCH, IRISH), the same genome-wide significance threshold ($P < 5 \times 10^{-7}$) adopted by the WTCCC (Wellcome Trust Case Control Consortium, Nature **447**, 661-78, 2007) was applied. It should be noted that this threshold is close to the conservative Bonferroni corrected significance threshold as suggested by Skol *et al.* (Nat Genet **38**, 209-13, 2006) for a two-stage study such as this in which genome-scan and subsequent results from stage-2 followup samples are combined. Since 307,411 non-*HLA* SNPs were tested in this genome scan, the

genomewide significance threshold based on Skol *et al.* criteria would be approximately $0.05/307,411 \approx 2 \times 10^{-7}$.

Whole blood expression analysis. PAXgene blood RNA sample was collected from unselected Illumina Hap300 genotyped UKGWAS coeliac cases. All were selected on the basis of gluten free diet treatment for >6 months, to avoid possible bias in gene expression levels due to active inflammation. Notable features of this experimental design were: use of primary cells, analysis of expression from unstimulated leucocytes, and the possibility that the samples would be enriched for disease causal genetic variants compared to healthy controls. RNA was extracted using the PAXgene protocol, hybridised to Illumina HumanRef-8v2 expression arrays, scanned and processed in BeadStudio. Data were cubic spline normalised.

Whole blood PAXgene expression data were obtained from 114 unique coeliac individuals who had been genotyped in the UKGWAS. To assess the quality of the RNA for each sample, 6,298 genes were assessed with two different probes present on the Illumina HumanRef8v2 chip. Intensities for all probes for each sample (Median Pearson correlation all samples = 0.22) were correlated and removed n=2 outlier samples that had low correlation (Pearson correlation < 0.10). Analysis of transcripts mapping to the non-pseudoautosomal region of chromosome Y, allowed for determining sex, resulting in the removal of n=2 samples with mismatched gender. n=1 further sample was removed where genotypes did not correlate as expected with *cis* expression level across the genome. Data from 109 individuals were analysed after applying these quality control criteria. ANOVA tests were performed to investigate possible *cis* effects of the strongest SNP associations on candidate gene expression from disease associated regions.

Intestinal biopsy expression analysis. Duodenal tissue biopsies were collected in RNAlater (Ambion), RNA was extracted using TRIzol (Invitrogen) and glass beads, hybridised to HumanRef-8v2 arrays and analysed as for the PAXgene samples. Expression profiling of T cell subsets is described in Supplementary Methods, below.

Table 2 Genomic regions with the strongest association signals for coeliac disease

Cytogenetic location	Position (bp)	SNP	Rank UKGWAS	P UKGWAS	*P* follow-up	Odds ratio[a] *follow-up* [95% CI]	P *overall*	Putative candidate genes[b]
Sample size				n=2189	n=5049		n=7238	
Assay chemistry				Infinium	GoldenGate			
1q31	190803436	rs2816316	382	0.0012	5.11×10^{-9}	0.71 [0.63 - 0.80]	2.58×10^{-11}	RGS1
2q11-2q12	102338297	rs13015714	149	0.00041	2.66×10^{-6}	1.27 [1.15 - 1.40]	4.37×10^{-9}	IL1RL1, IL18R1, IL18RAP, SLC9A4
	102437000	rs917997	41	9.06×10^{-5}	1.97×10^{-6}	1.27 [1.15 - 1.40]	8.49×10^{-10}	
3p21	46327388	rs6441961	1004	0.0028	3.21×10^{-5}	1.21 [1.10 - 1.32]	3.14×10^{-7}	CCR1, CCR3
3q25 - 3q26	161147744	rs17810546	111	0.00034	7.77×10^{-7}	1.34 [1.19 - 1.51]	1.07×10^{-9}	IL12A, SCHIP1
	161179692	rs9811792	898	0.0028	5.42×10^{-6}	1.21 [1.12 - 1.32]	5.24×10^{-8}	
3q28	189570322	rs9851967	453	0.00091	7.13×10^{-6}	0.82 [0.76 - 0.90]	2.45×10^{-8}	LPP
	189571948	rs13076312	186	0.00034	1.21×10^{-5}	1.21 [1.11 - 1.31]	1.78×10^{-8}	
	189595248	rs1464510	52	7.65×10^{-5}	1.21×10^{-5}	1.21 [1.11 -1.31]	5.33×10^{-9}	
	189607048	rs1559810	59	0.00010	3.28×10^{-5}	1.19 [1.10 - 1.30]	1.99×10^{-8}	
4q27	123292459	rs11938795	88	0.00026	1.07×10^{-5}	0.80 [0.73 - 0.89]	1.30×10^{-8}	IL2, IL21
	123334952	rs13151961	10	5.48×10^{-6}	4.32×10^{-8}	0.72 [0.65 - 0.81]	1.53×10^{-12}	
	123437763	rs13119723	1	2.51×10^{-7}	1.23×10^{-7}	0.73 [0.65 - 0.82]	5.94×10^{-13}	
	123447563	rs11734090	106	0.00034	2.95×10^{-5}	0.81 [0.74 - 0.90]	4.76×10^{-8}	
	123560609	rs7684187	245	0.00091	3.11×10^{-5}	0.82 [0.75 - 0.90]	1.13×10^{-7}	
	123727951	rs12642902	25	4.59×10^{-5}	7.35×10^{-6}	0.82 [0.75 - 0.89]	2.12×10^{-9}	
	123728871	rs6822844	7	4.80×10^{-6}	9.84×10^{-9}	0.71 [0.63 - 0.80]	2.82×10^{-13}	
	123774157	rs6840978	31	4.79×10^{-5}	2.32×10^{-7}	0.75 [0.68 - 0.84]	5.53×10^{-11}	
6q25	159385965	rs1738074	503	0.0016	1.19×10^{-5}	1.21 [1.11 - 1.31]	6.71×10^{-8}	TAGAP
12q24	110368991	rs3184504[c]	318	0.0011	3.22×10^{-5}	1.19 [1.10 - 1.30]	1.33×10^{-7}	SH2B3(LNK), ATXN2
	110492139	rs653178	207	0.00065	2.98×10^{-5}	1.19 [1.10 - 1.30]	8.00×10^{-8}	

Cochran-Mantel-Haenszel association analysis of 1020 SNPs (excluding *HLA* region). SNPs with *P* overall < 5 x 10-7 (the WTCCC threshold for reporting possible associations Wellcome Trust Case Control Consortium, *supra*) are shown. Association statistics are reported for twenty one SNPs from eight regions, including the previously reported IL2/IL21 region. Chromosomal positions based on NCBI build-36 coordinates. Data from the UKGWAS was re-analysed with removal of a small number of first degree relative compared to previous publication (van Heel *et al.*, *supra*). [a]Odds ratios for the UKGWAS ,and overall study are reported in Table 4. [b]Named genes either map to the same strong LD block as associated SNPs, or the SNPs physically map within the gene – causality is not proven. [c]rs3184504 is a non-synonymous SNP in SH2B3 (R262

WO 2009/103992 PCT/GB2009/000475

SUPPLEMENTARY METHODS

Subject DNA. DNA was extracted from whole blood, except for 1958 Cohort control samples which were lymphoblastoid cell line DNA, and 374 cases and 176 controls from the UK2 collection which were Oragene saliva DNA. Whole genome amplified (WGA) blood DNA was used for 194 Irish cases and 18 Dutch cases. Genotype cluster theta values for WGA DNA were similar to blood DNA, for a small fraction of markers intensity (R) was lower. One marker (rs641941) failed in Irish but not Dutch WGA samples.

Genotype concordance. Concordance between saliva DNA GoldenGate genotypes and blood DNA Infinium genotypes was 99.85% in n=4 subjects genotyped for 1025 SNPs on both platforms. A control DNA sample was included on 96 well sample plates genotyped in both London (UK2, Irish collections) and the Netherlands (Dutch collections). Concordance between plates for this sample was 99.94% for 45 replicates of 1025 SNPs. A further 9 control samples were genotyped once in both London and the Netherlands, concordance was 99.90% over 1025 SNPs.

Calculation of familial clustering due to the identified loci. For the most significant SNP in each locus in Table 2, genotype relative risk (GRR) was calculated for the observed allelic odds ratio assuming a multiplicative mode of inheritance and population frequency of the disease-associated allele equal to the frequency observed in controls (Risch & Teng, Genome Res **8**, 1273-88, 1998; McGinnis, Am J Hum Genet **67**, 1340-7, 2000). The calculated GRR and population frequency of the disease-associated allele were then used to calculate familial clustering in terms of the sibling relative risk (λ_s) contributed by the locus (McGinnis, *supra*). Familial clustering contributed by the 8 loci in Table 2 and by HLA were combined according to the multilocus multiplicative model of Risch (Am J Hum Genet **46**, 222-8, 1990) and thus $\sum\log(\lambda_{si})/\log(\lambda_s)$ is the proportion of the total disease clustering (λ_s) contributed by the identified loci (λ_{si}). For this calculation, epidemiological estimates of 30 for λ_s and 3.3 for the λ_{si} contributed by the HLA region (Bevan *et al.*, J Med Genet **36**, 687-90, 1999) were used.

Expression profiling of T cell subsets. T cell populations from the small intestine, thymus and spleen of healthy adult C57Bl/6 mice were stained with antibodies and FACS sorted on a MoFlo machine (Cytomation) as previously described (Shires *et al.*, Immunity **15**, 419-34, 2001). Total RNA preparation, reverse transcription, and semi-quantitative PCR were also undertaken as previously described (Shires *et al., supra*). Primer sequences were:

CCR5-F	5'-GGTACTTGGCTATTGTCCATGCTG-3'	(SEQ ID NO: 22)
CCR5-R	5'-ATGACAAGTAGAGGCAGGATCAGG-3'	(SEQ ID NO: 23)
CCR7-F	5'-ATCATCCGTACCTTGCTCCAGGCAC-3'	(SEQ ID NO: 24)
CCR7-R	5'-TGTCAACCTGACTGGCCAGAATTGC-3'	(SEQ ID NO: 25)
RGS-1-F	5'-ACCTGAGATCGATGATCCCACATCT-3'	(SEQ ID NO: 26)
RGS-1-R	5'-CTGTCGATTCTCGAGTATGGAAGTC-3'	(SEQ ID NO: 27)
β-actin-F	5'-TCCCTGTATGCCTCTGGTCGTACCAC-3'	(SEQ ID NO: 28)
β-actin-R	5'-CAGGATCTTCATGAGGTAGTCTGTCAG-3'	(SEQ ID NO: 29)

Table 3 | Subjects

	UKGWAS	UK2	Irish	Dutch	**TOTAL**
Coeliac Cases (n)	767	719	416	508	2410
HLA DQ2.5 positive	89.1%	87.6%	88.2%	87.4%	
Male/Female (n)	213/554	166/553	139/277	170/338	
Controls (n)	1422	1561	957	888	4828
HLA DQ2.5 positive	25.5%	27.7%	33.8%	28.6%	
Male/Female	720/702	584/977	283/674	540/348	
TOTAL	**2189**	**2280**	**1373**	**1396**	**7238**

Data from the UKGWAS was re-analysed with removal of a small number of first degree relatives compared to previous publication (van Heel *et al., supra*). HLA-DQ2.5*cis* status (positive, one or two copies) genotype inferred from rs2187668

RESULTS

1,164 non-*HLA* SNPs were initially selected from the UKGWAS for follow up, comprising 1,088 single SNPs with association results of P <0.00275 and 76 additional non-synonymous SNPs (nsSNP) with association results between $P \geq 0.00275$ and $P<0.01$. After exclusion of SNPs unlikely to be successful based on Illumina GoldenGate assay design criteria, and inspection of actual genotype clusters, 1,025 SNPs were analysed in follow-up collections. These collections comprised coeliac cases and controls of Northern European origin and of similar phenotypes to the UKGWAS samples. This design represents an unbiased search of the genome for susceptibility variants. The markers included 8 SNPs from the *IL2-IL21* region that were reported to be associated to coeliac disease in the inventors' previous study (van Heel *et al.*, *supra*), and 5 SNPs were additionally selected to tag coeliac disease associated *HLA-DQ2/8* haplotypes (de Bakker *et al.*, Nat Genet **38**, 1166-72, 2006). Samples failing quality control criteria were excluded (see Methods, above), and 719 cases and 1,561 population controls from the UK (UK2 collection), 416 cases and 957 bloodbank controls from Ireland (IRISH), 508 cases and 888 bloodbank controls from the Netherlands (DUTCH) were analysed.

Observed association statistics for the UKGWAS SNPs in the follow-up collections markedly deviate from expected findings (Figure 1a). The inventors reported 21 non-*HLA* SNPs from 8 distinct chromosomal regions meeting a genome wide significance threshold in all 7,238 samples of P overall $< 5 \times 10^{-7}$ (Table 2 above). Results from the WTCCC (Wellcome Trust Case Control Consortium, *supra*) and other recent GWA studies have shown that the majority of markers at a $P < 5 \times 10^{-7}$ "genome-wide" significance level will be true findings, although independent replication by other investigators is necessary for definitive validation. Only one of these eight regions, the *IL2-IL21* region, has been previously reported in coeliac disease (van Heel *et al.*, *supra*). Breslow-Day tests were non-significant for each of the eight regions implying consistent effect sizes and direction across the four collections, and accuracy of the reported Cochran-Mantel-Haenszel test odds ratios. The observation of generally weaker association evidence in the IRISH dataset (Table 4) is therefore likely to be a reflection of the smaller sample-size of this collection, rather than ethnic heterogeneity.

No evidence was observed for gene-gene interactions (departure from log-additive effects) between these regions, nor between the *HLA* and these regions. Association statistics, linkage disequilibrium plots, and Ensembl genes for each of the eight regions are shown in Table 2 and Figure 2, and more detailed statistics are shown in Table 4. Coeliac disease HLA-DQ associations reflect the ability of antigen presenting cells to present toxic cereal epitopes to T cells. Remarkably, seven of the eight identified non-HLA regions also contain biologically plausible candidate genes involved in the immune response.

The data was inspected for possible bias due to population differences between cases and controls, genotyping artefact, missing genotype data or other factors. The overall genotype call rate across all samples and SNPs was high at 99.94% (details for each of the coeliac disease associated SNPs and 5 HLA tag markers in Table 5). Inspection of cluster plots for the most associated SNPs from each of the eight non-HLA coeliac associated regions (Figure 4), consistent findings from two assay chemistries (Infinium and GoldenGate) in the UK populations, and multiple associated SNPs for some regions suggested that results were not due to laboratory generated false positive findings. The median distribution of test statistics was assessed using genomic control (Devlin & Roeder, Biometrics **55**, 997-1004, 1999) (λ_{GC} = 1.0 indicates a null distribution with no inflation of test statistics). In the full UKGWAS genome wide association scan dataset (using 767 non-first degree related cases and 1422 controls genotyped for 307,411 non-*HLA* SNPs) there was minimal evidence for inflation of test statistics (λ_{GC} = 1.03). In the UK2 IRISH DUTCH follow-up samples, genotyped for 1020 non-HLA SNPs there was also little evidence for inflation (λ_{GC} = 1.02). When 28 SNPs mapping to the eight non-HLA coeliac associated regions were also excluded there was no evidence for test statistic bias (λ_{GC} = 1.00) in the follow-up samples, and the residual Q-Q plot showed little deviation from expected results (Figure 1b). Furthermore, when an alternative principal components approach was used, each of the eight non-HLA regions again met the genome wide significance threshold (*P* overall < 5×10^{-7}) when corrected by EIGENSTRAT analysis (Table 6).

Cis gene expression in whole blood RNA samples (n=109) was correlated with SNPs from the eight non-HLA coeliac associated regions, to test for possible functional effects of these markers. Ei ght putative candidate genes from the regions were expressed above background levels in these samples. 14 pairs of SNPs and expressed gene probes were analysed (Table 7). Tissue and cell type expression were analysed based on published literature and the GNF SymAtlas database (Su *et al.*, Proc Natl Acad Sci USA **101**, 6062-7, 2004) for putative candidate genes from each region. Gene expression in small intestinal tissue (where coeliac disease manifests) from healthy controls, and treated and untreated coeliac disease individuals was analysed. Each region is summarised below.

4q27

A previously identified SNP rs6822844 (van Heel *et al.*, *supra*), located ~24 kb 3' of *IL21*, showed the strongest association with coeliac disease in the current study (*P* overall=2.82 x 10^{-13}). Independent replication, with the same allele and direction, of the inventors' previous report (van Heel *et al.*, *supra*) is provided by the UK2 collection (Table 4, rs6822844 *P* UK2=0.0017). This SNP is among a cluster of 8 associated SNPs that are in a block of strong linkage disequilibrium containing four genes (*KIAA1109-ADAD1-IL2-IL21*). Both *IL2* and *IL21* are strong candidate genes because of their role in T cell activation. A SNP allele from this region was also reported in a recent type 1 diabetes GWAS (conferring susceptibility), and Graves' disease (conferring reduced risk) (Todd *et al.*, Nat Genet **39**, 857-64 2007). A further Dutch study suggested association of rs6822844 with type 1 diabetes and rheumatoid arthritis, in the same direction as the coeliac disease data (Zhernakova *et al.*, Am J Hum Genet **81**, 1284-8, 2007). These findings suggest the 4q27 region might represent a more general autoimmune locus, although whether effects are due to one or multiple causal variants and the exact nature of these effects is currently unclear.

1q31

The most significant SNP outside the *HLA* and *IL2-IL21* regions is rs2816316 (*P* overall = 2.58 x 10^{-11}) located within a ~70kb LD block containing the gene *RGS1* (regulator of G-protein signaling 1). rs2816316 maps 8kb distal to the 5' end of

RGS1. *RGS* family genes attenuate the signalling activity of G-proteins by acting as GTPase activating proteins. RGS1 acts to regulate chemokine receptor signaling and is known to be involved in B-cell activation and proliferation. *RGS1*$^{-/-}$ mice have enhanced B cell movement into and out of lymph nodes (Han *et al.*, Immunity **22**, 343-54 2005), and heightened dendritic cell migratory response to chemokines (Shi *et al.*, J Immunol **172**, 5175-84, 2004). *RGS1* was found to be expressed in human small intestinal biopsies (Figure 5). Interestingly, the present inventors previously observed *RGS1* to be strongly expressed in murine intestinal intra-epithelial lymphocytes (Pennington *et al.*, Nat Immunol **4**, 991-8, 2003), and now confirm that T cell *RGS1* expression appears to be specific to the intestinal intra-epithelial lymphocyte compartment and is not found in conventional splenic or thymic αβ T cells (Figure 6). Intestinal intra-epithelial lymphocytes play a key role in epithelial cell death and the development of villous atrophy in coeliac disease (Hue *et al.*, Immunity **21**, 367-77, 2004).

2q11 - 2q12

Two associated SNPs, rs917997 (*P* overall=8.49 x 10^{-10}) and rs13015714 (both in LD, r^2=0.95 in UKGWAS controls) map to a ~400kb linkage disequilibrium block. Interestingly, in the WTCCC GWAS (Wellcome Trust Case Control Consortium, *supra*), Crohn's disease shows modest association (proxy SNP for rs917997, P~10^{-4}, Table 8). This LD block contains four genes, two of which are receptors for the IL-18 protein (*IL18RAP* and *IL18R1*). The IL-18 pathway is highly relevant as mature IL-18 induces T cell interferon-γ synthesis, a key cytokine involved in the mucosal inflammation of coeliac disease. IL-18 binds to targeted cells through a receptor comprising an α chain (IL18Rα, IL18R1) and a β chain (IL18Rβ, AcPL, *IL18RAP*). Mature IL-18 is expressed in the intestinal mucosa of active, treated and latent coeliac patients but not in healthy controls (Salvati *et al.*, Gut **50**, 186-90, 2002).

IL18RAP is strongly expressed in unstimulated T cells and NK cells (GNF SymAtlas, Su *et al.*, *supra*), and is expressed in small intestinal biopsies (Figure 5). A large *cis* effect (ANOVA *P*= 3.2 x 10^{-5}) of rs917997 genotypes was observed on the level of *IL18RAP* mRNA expression in whole blood from treated coeliac patients (Figure 3),

accounting for 16.1% of the population variance in *IL18RAP* expression. A significant allele dosage effect on expression (*post-hoc* regression testing for linear trend $P < 0.0001$) was also observed. Individuals homozygous for the minor rs917997 A allele (which is more common in coeliac than control subjects) expressed the lowest levels of *IL18RAP* mRNA, heterozygotes intermediate and G allele homozygotes the highest levels.

The coding regions of *IL18RAP* were sequenced in 23 coeliac disease patients and 8 control individuals, and 19 variants were found, 17 of which were already in dbSNP. No variants (dbSNP or from resequencing) map to the region of the Illumina *IL18RAP* expression probe. None of the variants is predicted to have functional consequences. One new variant (c.1210+17A>G) was observed in a single control and the other new variant (c.1384+70_1384+71insT) was found in two coeliac individuals.

3p21

SNP rs6441961 (*P* overall = 3.14×10^{-7}) maps within a large cluster of chemokine receptor genes on 3p21, including *CCR1, CCR2, CCRL2, CCR3, CCR5* and *CCXCR1*. rs6441961 lies 44kb 3' of the nearest gene (*CCR3*). LD block definition is hampered by poor HapMap coverage of this region due to structural variation (Iafrate *et al.*, Nat Genet **36**, 949-51, 2004).

Chemokines and their receptors are critical for the recruitment of effector immune cells to the site of inflammation. In the WTCCC GWAS (Wellcome Trust Case Control Consortium, *supra*), type 1 diabetes shows modest association in the same direction with the same allele of SNP rs6441961 ($P \sim 10^{-5}$, Table 8) suggesting a possible common mechanism between both immune-mediated diseases.

3q25 - 3q26

Two SNPs (rs17810546, rs9811792) in a ~70kb linkage disequilibrium block show strong association to coeliac disease (rs17810546 *P* overall = 1.07×10^{-9}). This region is immediately 5' of *IL12A* (interleukin-12 A). Interestingly, the two

associated SNPs (r^2=0.19 in UKGWAS controls) may represent independent association signals. SNP rs17810546 also shows modest correlation with SNPs in both *IL12A* and *SCHIP1* (schwannomin interacting protein 1). *IL12A* encodes the IL12p35 subunit that together with IL12p40 form IL12p70, i.e. the heterodimeric IL-12 cytokine that has a broad range of biological activities on T and natural killer cells. IL-12 induces interferon-γ secreting Th1 cells, one of the immunological hallmarks of coeliac disease.

3q28

Multiple correlated SNPs (rs1464510 *P* overall = 5.33 x 10^{-9}) within a ~70 kb linkage disequilibrium block show association with coeliac disease. This block is either 5' of the RefSeq gene *LPP*, or intronic for other possible isoforms of *LPP*. The *LPP* gene shows very high expression in the small intestine (Figure 5), and may play a structural role at sites of cell adhesion in maintaining cell shape and motility. Relatively little is known about *LPP*, and how genetic variation in this region might predispose to coeliac disease is unclear.

6q25

SNP rs1738074 (*P* overall = 6.71 X 10^{-8}) on chromosome 6q25.3 maps to a ~200 kb linkage disequilibrium block containing *TAGAP* (T-cell activation GTPase activating protein), itself within a larger region of weaker LD containing *RSPH3* (radial spokehead-like 3). TAGAP is expressed in activated T cells (Mao *et al.*, Genomics **83**, 989-99, 2004), has three isoforms, and is a Rho GTPase-activating protein important for modulating cytoskeletal changes, although little is known about its role in immune function.

12q24

Association signals arise from two correlated SNPs rs653178 (*P* overall = 8 x 10^{-8}) and rs3184504 (r^2=0.99 in UKGWAS controls). These markers map in the vicinity of *SH2B3* (also known as *LNK*) and *ATXN2*, modest LD is seen over a broader region of ~1Mb containing multiple other genes. Strong association with type 1 diabetes is reported in this region, with rs3184504 (same allele and direction as coeliac disease)

entirely accounting for the association signal (Todd *et al., supra*). *SH2B3* (also known as *LNK*, lymphocyte adaptor protein) is strongly expressed in monocytes and dendritic cells, as well as to a lesser extent in resting B, T cells and NK cells (GNF SymAtlas, Su *et al., supra*). SH2B3 was found to be strongly expressed in the small intestine. Higher expression in inflamed coeliac biopsies may reflect leukocyte recruitment and activation (Figure 5). The rs3184504 marker is a non-synonymous SNP in exon 3 of *SH2B3* leading to a R262W amino acid change in the pleckstrin homology domain. This domain may be important in plasma membrane targeting. SH2B3 regulates T cell receptor, growth factor, and cytokine receptor-mediated signalling implicated in leukocyte and myeloid cell homeostasis (Fitau *et al.,* J Biol Chem **281**, 20148-59, 2006 and Li *et al.,* J Immunol **164**, 5199-206, 2000). *SH2B3*$^{-/-}$ mice have increased responses to multiple cytokines (Velazquez *et al.,* J Exp Med **195**, 1599-611, 2002).

Detailed Association Results

R	SNP	bp_b36	P ASSOC UKGWAS	OR UKGWAS	L95	U95	P_CMH UK2 IRISH DUTCH	OR_CMH UK2 IRISH DUTCH	L95	U95	P_CMH OVERALL	OR_CMH OVERALL	L95	U95
1	rs2816316	190803436	**0.001159**	0.7485	0.6283	0.8918	**5.11E-09**	0.7091	0.6317	0.7961	**2.58E-11**	0.7208	0.6545	0
2	rs13015714	102338297	**0.0004103**	1.298	1.123	1.501	**2.66E-06**	1.267	1.148	1.399	**4.37E-09**	1.277	1.177	
2	rs917997	102437000	**9.06E-05**	1.333	1.154	1.54	**1.97E-06**	1.27	1.151	1.401	**8.49E-10**	1.29	1.189	
3	rs6441961	46327388	**0.00284**	1.224	1.072	1.397	**3.21E-05**	1.206	1.104	1.317	**3.14E-07**	1.211	1.125	
3	rs17810546	161147744	**0.0003393**	1.379	1.156	1.644	**7.77E-07**	1.342	1.194	1.509	**1.07E-09**	1.353	1.228	
3	rs9811792	161179692	**0.002775**	1.209	1.068	1.37	**5.42E-06**	1.214	1.117	1.32	**5.24E-08**	1.212	1.131	
3	rs9851967	189570322	**0.0009039**	0.8069	0.7108	0.916	**7.13E-06**	0.8231	0.756	0.8962	**2.45E-08**	0.818	0.7623	0
3	rs13076312	189571948	**0.0003372**	1.255	1.108	1.421	**1.21E-05**	1.205	1.108	1.31	**1.78E-08**	1.22	1.139	
3	rs1464510	189595248	**7.65E-05**	1.285	1.135	1.456	**1.21E-05**	1.205	1.108	1.31	**5.33E-09**	1.229	1.147	
3	rs1559810	189607048	**0.0001026**	1.281	1.13	1.452	**3.28E-05**	1.194	1.098	1.299	**1.99E-08**	1.221	1.138	
4	rs11938795	123292459	**0.0002567**	0.7592	0.6547	0.8802	**1.07E-05**	0.8046	0.7303	0.8865	**1.30E-08**	0.7907	0.7291	0
4	rs13151961	123334952	**5.48E-06**	0.6618	0.5535	0.7914	**4.32E-08**	0.7245	0.6453	0.8134	**1.53E-12**	0.7053	0.64	0
4	rs13119723	123437763	**2.51E-07**	0.602	0.4956	0.7311	**1.23E-07**	0.7343	0.6547	0.8236	**5.94E-13**	0.6967	0.6311	
4	rs11734090	123447563	**0.000337**	0.7639	0.6592	0.8853	**2.95E-05**	0.8143	0.7394	0.8968	**4.76E-08**	0.7988	0.7368	
4	rs7684187	123560609	**0.0009105**	0.789	0.6858	0.9077	**3.11E-05**	0.8217	0.7491	0.9013	**1.13E-07**	0.8116	0.7513	0
4	rs12642902	123727951	**4.59E-05**	0.7554	0.66	0.8647	**7.35E-06**	0.816	0.7465	0.8919	**2.12E-09**	0.7971	0.74	0
4	rs6822844	123728871	**4.80E-06**	0.6602	0.5521	0.7895	**9.84E-09**	0.712	0.6337	0.8	**2.82E-13**	0.6961	0.6314	0
4	rs6840978	123774157	**4.79E-05**	0.715	0.6079	0.8409	**2.32E-07**	0.7549	0.6784	0.84	**5.53E-11**	0.7426	0.6792	0
6	rs7775397	32369230	**7.82E-151**	6.553	5.644	7.608	**0**	6.174	5.607	6.799	**0**	6.285	5.796	
6	rs2395182	32521295	**2.43E-17**	0.4514	0.3743	0.5444	**1.23E-38**	0.4455	0.3934	0.5045	**2.74E-54**	0.4473	0.4033	0
6	rs2187668	32713862	**1.43E-167**	7.016	6.059	8.124	**0**	6.255	5.691	6.875	**0**	6.472	5.978	
6	rs7775228	32766057	**1.27E-05**	1.461	1.232	1.734	**1.29E-13**	1.549	1.379	1.74	**9.77E-18**	1.52	1.381	
6	rs2856705	32778934	**4.84E-08**	1.642	1.372	1.964	**7.42E-21**	1.776	1.574	2.005	**3.03E-27**	1.732	1.567	
6	rs1738074	159385965	**0.001564**	1.223	1.079	1.386	**1.19E-05**	1.206	1.109	1.311	**6.71E-08**	1.211	1.13	
12	rs3184504	110368991	**0.001075**	0.8123	0.7171	0.9201	**3.22E-05**	1.194	1.098	1.297	**1.33E-07**	1.205	1.124	
12	rs653178	110492139	**0.0006489**	0.8053	0.711	0.9121	**2.98E-05**	1.194	1.099	1.298	**8.00E-08**	1.209	1.128	

Detailed Association Results

A1	F_A	F_U	A2	P ASSOC UKGWAS	P HWE CONTROLS UKGWAS	A1	F_A	F_U	A2	P ASSOC UK2	P HWE CONTROLS UK2	A1	F_A	F_U	A2	P ASSOC IRISH	P HWE CONTROLS IRISH	A1	F_A	F_U	A2	P ASSOC DUTCH
C	0.1362	0.1741	A	**0.001159**	0.5803	C	0.1293	0.1877	A	**1.05E-06**	0.04661	C	0.1611	0.187	A	**0.1027**	0.3388	C	0.1407	0.1892	A	**0.001089**
C	0.2611	0.2139	A	**0.0004103**	0.04824	C	0.2434	0.2024	A	**0.001769**	0.4806	C	0.2067	0.2017	A	**0.7621**	0.1916	C	0.2884	0.214	A	**9.72E-06**
A	0.2673	0.2148	G	**9.06E-05**	0.02747	A	0.2451	0.2078	G	**0.004778**	0.4423	A	0.2138	0.2037	G	**0.5488**	0.09061	A	0.2933	0.2151	G	**3.61E-06**
A	0.3409	0.2971	G	**0.00284**	0.7032	A	0.3421	0.3017	G	**0.006314**	0.9521	A	0.3425	0.3177	G	**0.2007**	0.4557	A	0.3799	0.3221	G	**0.00195**
G	0.1617	0.1227	A	**0.0003393**	0.9018	G	0.146	0.1268	A	**0.07603**	0.3611	G	0.1839	0.1301	A	**0.0002473**	0.6676	G	0.1713	0.1244	A	**0.0006367**
G	0.4857	0.4385	A	**0.002775**	0.8295	G	0.4826	0.4401	A	**0.007436**	0.8373	G	0.4928	0.4457	A	**0.02279**	0.8444	G	0.497	0.4392	A	**0.003167**
A	0.3851	0.437	G	**0.0009039**	0.957	A	0.372	0.4433	G	**5.92E-06**	0.5052	A	0.4255	0.4561	G	**0.1379**	0.2416	A	0.3858	0.411	G	**0.1912**
A	0.5163	0.4596	G	**0.0003372**	0.2858	A	0.5243	0.4513	G	**4.46E-06**	0.9592	A	0.476	0.4472	G	**0.1648**	0.01556	G	0.4764	0.4983	A	**0.2647**
A	0.5196	0.4569	C	**7.65E-05**	0.2614	A	0.516	0.4462	C	**1.13E-05**	0.8779	A	0.4832	0.4477	C	**0.08649**	0.01086	C	0.4793	0.5	A	**0.2932**
A	0.4732	0.4122	C	**0.0001026**	0.07049	A	0.468	0.409	C	**0.0001836**	0.7537	A	0.4459	0.4096	C	**0.0766**	0.01946	A	0.4783	0.451	C	**0.1633**
G	0.2132	0.263	A	**0.0002567**	0.7323	G	0.2309	0.2655	A	**0.01255**	0.01954	G	0.2488	0.28	A	**0.09034**	1	G	0.2234	0.2832	A	**0.0005416**
G	0.1258	0.1786	A	**5.48E-06**	0.5883	G	0.1419	0.1765	A	**0.003467**	0.02931	G	0.155	0.1944	A	**0.01424**	1	G	0.1289	0.1909	A	**2.57E-05**
G	0.1013	0.1577	A	**2.51E-07**	0.1928	G	0.1446	0.1791	A	**0.003899**	0.05855	G	0.1562	0.1964	A	**0.01253**	1	G	0.1348	0.192	A	**0.0001144**
G	0.2161	0.2651	A	**0.000337**	0.6831	G	0.233	0.2665	A	**0.0159**	0.02332	G	0.25	0.2806	A	**0.09785**	0.8101	G	0.2313	0.2866	A	**0.001474**
G	0.2536	0.301	A	**0.0009105**	0.801	G	0.2643	0.304	A	**0.006101**	0.00122	G	0.3005	0.326	A	**0.1867**	0.7136	G	0.2589	0.3142	A	**0.00202**
A	0.2855	0.346	G	**4.59E-05**	0.5979	A	0.3052	0.3514	G	**0.002188**	0.00154	A	0.3341	0.3713	G	**0.06202**	0.5333	A	0.3091	0.3604	G	**0.005959**
A	0.1258	0.179	C	**4.80E-06**	0.5882	A	0.1377	0.1746	C	**0.00173**	0.01066	A	0.1502	0.1991	C	**0.002427**	0.6127	A	0.1299	0.1864	C	**0.000114**
A	0.1636	0.2148	G	**4.79E-05**	0.5293	A	0.1732	0.2104	G	**0.003368**	0.04773	A	0.1959	0.2372	G	**0.01722**	1	A	0.1604	0.2185	G	**0.0002097**
C	0.4863	0.1262	A	**7.82E-151**	0.5484	C	0.493	0.1326	A	**7.14E-152**	0.3215	C	0.5168	0.1714	A	**2.69E-77**	0.7326	C	0.5315	0.1391	A	**6.21E-109**
C	0.1043	0.2051	A	**2.43E-17**	0.5695	C	0.0988	0.2034	A	**2.47E-18**	0.3104	C	0.1538	0.2456	A	**8.74E-08**	0.5427	C	0.08563	0.2072	A	**5.84E-17**
A	0.5296	0.1383	G	**1.43E-167**	0.3743	A	0.5292	0.1502	G	**3.89E-158**	0.6924	A	0.5421	0.1823	G	**4.92E-81**	0.2336	A	0.5659	0.1543	G	**1.44E-114**
G	0.1773	0.1285	A	**1.27E-05**	0.1233	G	0.1878	0.1265	A	**5.20E-08**	0.6465	G	0.1827	0.1385	A	**0.002993**	1	G	0.1467	0.09403	A	**2.42E-05**
A	0.1662	0.1083	G	**4.84E-08**	0.1675	A	0.1818	0.1102	G	**3.70E-11**	1	A	0.1731	0.1202	G	**0.0002034**	0.759	A	0.1378	0.07095	G	**7.10E-09**
A	0.472	0.4222	G	**0.001564**	0.277	A	0.459	0.4279	G	**0.04931**	0.9588	A	0.5192	0.4681	G	**0.0138**	1	A	0.4596	0.3947	G	**0.0008179**
G	0.4594	0.5113	A	**0.001075**	0.2227	A	0.5223	0.4718	G	**0.00154**	0.6475	A	0.5048	0.4765	G	**0.1724**	0.4372	A	0.5276	0.4786	G	**0.0128**
A	0.4576	0.5116	G	**0.0006489**	0.1518	G	0.5229	0.4715	A	**0.001238**	0.6118	G	0.506	0.4765	A	**0.1549**	0.4372	G	0.5256	0.4786	A	**0.01689**

WO 2009/103992 PCT/GB2009/000475

Detailed Call Rates

SNP	bp_b36	P_CMH OVERALL	F_MISS ALL_7238_ SAMPLES	F_MISS UKGWAS _CASES	F_MISS UKGWAS_ CONTROLS	F_MISS UK2_ CASES	F_MISS UK2_ CONTROLS	F_MISS IRISH_ CASES	F_MISS IRISH_ CONTROLS	F_MISS DUTCH_ CASES	F_MI DUT CON
rs2816316	190803436	2.58E-11	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs13015714	102338297	4.37E-09	0.00028	0.00130	0.00070	0.00000	0.00000	0.00000	0.00000	0.00000	
rs917997	102437000	8.49E-10	0.00124	0.00000	0.00000	0.00417	0.00128	0.00481	0.00209	0.00000	
rs6441961	46327388	3.14E-07	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs17810546	161147744	1.07E-09	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs9811792	161179692	5.24E-08	0.00014	0.00000	0.00000	0.00139	0.00000	0.00000	0.00000	0.00000	
rs9851967	189570322	2.45E-08	0.00028	0.00130	0.00070	0.00000	0.00000	0.00000	0.00000	0.00000	
rs13076312	189571948	1.78E-08	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs1464510	189595248	5.33E-09	0.00124	0.00130	0.00422	0.00000	0.00064	0.00000	0.00104	0.00000	
rs1559810	189607048	1.99E-08	0.00069	0.00130	0.00281	0.00000	0.00000	0.00000	0.00000	0.00000	
rs11938795	123292459	1.30E-08	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs13151961	123334952	1.53E-12	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs13119723	123437763	5.94E-13	0.00097	0.00261	0.00352	0.00000	0.00000	0.00000	0.00000	0.00000	
rs11734090	123447563	4.76E-08	0.00014	0.00130	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs7684187	123560609	1.13E-07	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs12642902	123727951	2.12E-09	0.00055	0.00000	0.00000	0.00417	0.00000	0.00000	0.00104	0.00000	
rs6822844	123728871	2.82E-13	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs6840978	123774157	5.53E-11	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs7775397	32369230	0	0.00014	0.00130	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
rs2395182	32521295	2.74E-54	0.00041	0.00000	0.00211	0.00000	0.00000	0.00000	0.00000	0.00000	
rs2187668	32713862	0	0.00111	0.00913	0.00070	0.00000	0.00000	0.00000	0.00000	0.00000	
rs7775228	32766057	9.77E-18	0.00028	0.00000	0.00141	0.00000	0.00000	0.00000	0.00000	0.00000	
rs2856705	32778934	3.03E-27	0.00055	0.00391	0.00000	0.00139	0.00000	0.00000	0.00000	0.00000	
rs1738074	159385965	6.71E-08	0.00028	0.00000	0.00070	0.00000	0.00064	0.00000	0.00000	0.00000	
rs3184504	110368991	1.33E-07	0.00069	0.00522	0.00000	0.00000	0.00064	0.00000	0.00000	0.00000	
rs653178	110492139	8.00E-08	0.00028	0.00130	0.00070	0.00000	0.00000	0.00000	0.00000	0.00000	

WO 2009/103992 PCT/GB2009/000475

Eigenstrat Corrected Results

CHR	SNP	bp_b36	P ASSOC UKGWAS	P EIGENSTRAT UKGWAS+GC	P_CMH OVERALL	P EIGENSTRAT OVERALL+GC
1	rs2816316	190803436	**0.001159**	0.001907234	**2.58E-11**	2.6401E-10
2	rs13015714	102338297	**0.0004103**	0.001482461	**4.37E-09**	6.4801E-08
2	rs917997	102437000	**9.06E-05**	0.0003485	**8.49E-10**	1.07506E-08
3	rs6441961	46327388	**0.00284**	0.003083522	**3.14E-07**	5.31281E-07
3	rs17810546	161147744	**0.0003393**	0.000871531	**1.07E-09**	3.89843E-09
3	rs9811792	161179692	**0.002775**	0.005488555	**5.24E-08**	4.9355E-07
3	rs9851967	189570322	**0.0009039**	0.001750894	**2.45E-08**	1.75678E-07
3	rs13076312	189571948	**0.0003372**	0.000736503	**1.78E-08**	1.63189E-07
3	rs1464510	189595248	**7.65E-05**	0.000147628	**5.33E-09**	5.32054E-08
3	rs1559810	189607048	**0.0001026**	0.000219321	**1.99E-08**	1.57933E-07
4	rs11938795	123292459	**0.0002567**	0.000723281	**1.30E-08**	6.32648E-08
4	rs13151961	123334952	**5.48E-06**	1.77022E-05	**1.53E-12**	1.05963E-11
4	rs13119723	123437763	**2.51E-07**	1.66977E-06	**5.94E-13**	5.9521E-12
4	rs11734090	123447563	**0.000337**	0.000885697	**4.76E-08**	2.19117E-07
4	rs7684187	123560609	**0.0009105**	0.00208332	**1.13E-07**	6.19146E-07
4	rs12642902	123727951	**4.59E-05**	0.000152461	**2.12E-09**	1.65107E-08
4	rs6822844	123728871	**4.80E-06**	1.60261E-05	**2.82E-13**	1.9728E-12
4	rs6840978	123774157	**4.79E-05**	0.000120314	**5.53E-11**	3.89402E-10
6	rs7775397	32369230	**7.82E-151**	4.193E-148	**0**	0
6	rs2395182	32521295	**2.43E-17**	2.92865E-16	**2.74E-54**	1.19476E-50
6	rs2187668	32713862	**1.43E-167**	5.8178E-163	**0**	0
6	rs7775228	32766057	**1.27E-05**	1.08263E-05	**9.77E-18**	4.65683E-17
6	rs2856705	32778934	**4.84E-08**	2.03462E-08	**3.03E-27**	2.90003E-27
6	rs1738074	159385965	**0.001564**	0.0012549	**6.71E-08**	1.33071E-07
12	rs3184504	110368991	**0.001075**	0.000775232	**1.33E-07**	4.56786E-07
12	rs653178	110492139	**0.0006489**	0.000450066	**8.00E-08**	2.95227E-07

Tables 4 to 6 - Footnotes

All outputs are from PLINK, text has been edited in Excel.

SampleName	GenotypingAssay
UKGWAS	Infinium
UK2	GoldenGate
IRISH	GoldenGate
DUTCH	GoldenGate

All samples genotyped for 1025 SNPs, including 5 HLA tag markers, only coeliac-associated SNPs
and HLA tag markers shown

Abbreviations for Table 4 - Detailed Association Results

A1 represents minor allele (whole sample) for specified collection
A2 represents minor allele (whole sample) for specified collection
F_A, F_U allele frequency in affecteds, unaffecteds
P ASSOC: two tailed P value from the 2x2 allele count chi-squared test
P CMH: two tailed P value from the 2x2xK Cochran-Mantel-Haenszel test
OR CMH: allelic odds ratio from the 2x2xK Cochran-Mantel-Haenszel test

Abbreviations for Table 5 - Detailed Call Rates

F_MISS fraction missing data (0.0=complete data, 1.0 = entirely missing data)

tic	Gene	Illumina Probe ID	Detected above background (P value*)	SNP	ANOVA p.value	Bonferroni corrected ANOVA P value	Mean AA	Mean AB	Mean BB	n AA	n AB
	ATXN2	6560692	<10E-16	rs3184504	0.57	1.00	1005.3	1006.4	1044.3	30	57
	SH2B3	1580026	<10E-16	rs3184504	0.37	1.00	1089.9	1126.3	1021.1	30	57
	IL18R1	1500328	<10E-16	rs13015714	0.49	1.00	437.6	428.7	487.6	58	43
	IL18R1	1500328	<10E-16	rs917997	0.21	1.00	506.5	423.3	437.6	9	42
	IL18RAP	5130475	<10E-16	rs13015714	1.68E-05	0.00023	2694.4	2215.9	1345.9	58	43
	IL18RAP	5130475	<10E-16	rs917997	3.15E-05	0.00044	1451.4	2214.0	2694.4	9	42
	CCR1	4250070	<10E-16	rs6441961	0.38	1.00	413.4	357.1	367.4	15	49
	CCR3	7570670	<10E-16	rs6441961	0.013	0.19	794.2	763.5	637.2	15	49
	LPP	4260731	<10E-16	rs13076312	0.87	1.00	1570.9	1601.5	1573.2	29	60
	LPP	4260731	<10E-16	rs1464510	0.85	1.00	1562.1	1600.0	1586.1	27	63
	LPP	4260731	<10E-16	rs1559810	0.79	1.00	1561.8	1602.9	1568.6	25	59
	LPP	4260731	<10E-16	rs9851967	0.93	1.00	1566.3	1597.1	1580.7	13	61
	TAGAP	3450152	<10E-16	rs1738074	0.55	1.00	583.5	584.3	612.7	24	55
	TAGAP	4250369	<10E-16	rs1738074	0.46	1.00	650.6	639.3	684.8	24	55

obes with expression below background in PAXgene whole blood samples

	RGS1	3420356	>0.99
	IL1RL1	670411	>0.99
	IL1RL1	3870753	>0.99
	SLC9A4	6020020	>0.99
	CCR3	1980750	>0.99
	IL12A	240091	>0.99
	SCHIP1	1300670	>0.99
	IL2	1990484	>0.99
	IL21	5360241	>0.99

probe, the distribution of expression values from the 109 individuals were compared by Wilcoxon Rank Sum test to the combined
ution of negative control and labelling control probes (n=714 on the Illumina Ref8v2)

: regions with same SNP or a perfect proxy SNP in WTCCC GWAS

ocus	best coeliac SNP	HapMapCEU r2 with best wtccc SNP	wtccc file/phenotype	id	rsid	pos	allele1	allele2
	rs2816316	1.0	./snptest_BD_01.txt	SNP_A-1977749	rs1323296	189269165	A	G
	rs2816316	1.0	./snptest_CAD_01.txt	SNP_A-1977749	rs1323296	189269165	A	G
	rs2816316	1.0	./snptest_CD_01.txt	SNP_A-1977749	rs1323296	189269165	A	G
	rs2816316	1.0	./snptest_HT_01.txt	SNP_A-1977749	rs1323296	189269165	A	G
	rs2816316	1.0	./snptest_RA_01.txt	SNP_A-1977749	rs1323296	189269165	A	G
	rs2816316	1.0	./snptest_T1D_01.txt	SNP_A-1977749	rs1323296	189269165	A	G
	rs2816316	1.0	./snptest_T2D_01.txt	SNP_A-1977749	rs1323296	189269165	A	G
2	rs917997	1.0	./snptest_BD_02.txt	SNP_A-4241491	rs2041756	102508428	C	T
2	rs917997	1.0	./snptest_CAD_02.txt	SNP_A-4241491	rs2041756	102508428	C	T
2	rs917997	1.0	./snptest_CD_02.txt	SNP_A-4241491	rs2041756	102508428	C	T
2	rs917997	1.0	./snptest_HT_02.txt	SNP_A-4241491	rs2041756	102508428	C	T
2	rs917997	1.0	./snptest_RA_02.txt	SNP_A-4241491	rs2041756	102508428	C	T
2	rs917997	1.0	./snptest_T1D_02.txt	SNP_A-4241491	rs2041756	102508428	C	T
2	rs917997	1.0	./snptest_T2D_02.txt	SNP_A-4241491	rs2041756	102508428	C	T
	rs6441961	same	./snptest_BD_03.txt:	SNP_A-1972591	rs6441961	46327388	C	T
	rs6441961	same	./snptest_CAD_03.txt:	SNP_A-1972591	rs6441961	46327388	C	T
	rs6441961	same	./snptest_CD_03.txt:	SNP_A-1972591	rs6441961	46327388	C	T
	rs6441961	same	./snptest_HT_03.txt:	SNP_A-1972591	rs6441961	46327388	C	T
	rs6441961	same	./snptest_RA_03.txt:	SNP_A-1972591	rs6441961	46327388	C	T
	rs6441961	same	./snptest_T1D_03.txt:	SNP_A-1972591	rs6441961	46327388	C	T
	rs6441961	same	./snptest_T2D_03.txt:	SNP_A-1972591	rs6441961	46327388	C	T
6	rs17810546	1.0	./snptest_BD_03.txt	SNP_A-4295819	rs17809756	161112786	A	G
6	rs17810546	1.0	./snptest_CAD_03.txt	SNP_A-4295819	rs17809756	161112786	A	G
6	rs17810546	1.0	./snptest_CD_03.txt	SNP_A-4295819	rs17809756	161112786	A	G
6	rs17810546	1.0	./snptest_HT_03.txt	SNP_A-4295819	rs17809756	161112786	A	G
6	rs17810546	1.0	./snptest_RA_03.txt	SNP_A-4295819	rs17809756	161112786	A	G
6	rs17810546	1.0	./snptest_T1D_03.txt	SNP_A-4295819	rs17809756	161112786	A	G
6	rs17810546	1.0	./snptest_T2D_03.txt	SNP_A-4295819	rs17809756	161112786	A	G

: regions with same SNP or a perfect proxy SNP in WTCCC GWAS

maximum or_call	controls_AA	controls_AB	controls_BB	controls_NULL	cases_AA	cases_AB	cases_BB	cases_NULL
0.998518	99	853	1980	6	71	546	1233	18
0.998647	99	853	1980	6	49	556	1307	14
0.999063	99	853	1980	6	54	496	1190	8
0.998905	99	853	1980	6	63	555	1326	8
0.998679	99	853	1980	6	64	524	1258	14
0.998822	99	853	1980	6	52	530	1370	11
0.998776	99	853	1980	6	63	558	1293	10
0.999575	1798	986	151	3	1139	644	85	0
0.999507	1798	986	151	3	1160	659	106	1
0.999343	1798	986	151	3	982	656	108	2
0.999229	1798	986	151	3	1195	662	93	2
0.9996	1798	986	151	3	1094	665	101	0
0.999579	1798	986	151	3	1185	689	87	2
0.999439	1798	986	151	3	1113	707	101	3
0.996772	1452	1236	240	10	893	773	172	30
0.998208	1452	1236	240	10	943	796	175	12
0.997866	1452	1236	240	10	855	732	147	14
0.99712	1452	1236	240	10	994	784	149	25
0.998102	1452	1236	240	10	891	773	183	13
0.998793	1452	1236	240	10	878	854	230	1
0.997625	1452	1236	240	10	916	845	146	17
0.999773	47	642	2249	0	35	403	1429	1
0.999709	47	642	2249	0	27	430	1467	2
0.999882	47	642	2249	0	13	381	1352	2
0.999782	47	642	2249	0	29	446	1476	1
0.999865	47	642	2249	0	26	400	1432	2
0.999544	47	642	2249	0	41	419	1501	2
0.999758	47	642	2249	0	31	431	1459	3

: regions with same SNP or a perfect proxy SNP in WTCCC GWAS

st_add	P value frequentist_gen	-log10 P value bayesian_add	-log10 P value bayesian_gen	P value sex_frequentist_add	P value sex_frequentist_gen	good_clustering
0.410742	0.64437	-0.773862	-0.794302	0.631235	0.88867	1
0.299281	0.269129	-0.691405	-0.795925	0.435382	0.500753	1
0.489992	0.779561	-0.810373	-0.803361	0.737712	0.833312	1
0.608233	0.876368	-0.869727	-0.875776	0.71518	0.950723	1
0.745442	0.866225	-0.897521	-0.865357	0.953969	0.990619	1
.0317126	0.0959855	0.0617038	-0.343427	0.0890103	0.264708	1
0.945628	0.986788	-0.921757	-0.902305	0.990533	0.967159	1
0.859191	0.577734	-0.945112	-0.972689	0.373542	0.515247	1
0.434015	0.733712	-0.826394	-0.87261	0.657653	0.923286	1
)0823309	0.00285024	1.44251	1.13476	0.0032407	0.0123014	1
0.817303	0.829233	-0.947099	-0.839875	0.715849	0.875981	1
0.120982	0.241305	-0.436214	-0.599261	0.265741	0.553705	1
0.942174	0.339189	-0.955521	-0.60987	0.842634	0.630336	1
.0483701	0.0609361	-0.120355	-0.245009	0.0972004	0.141692	1
0.257952	0.367921	-0.710777	-0.857753	0.295526	0.482409	1
0.502103	0.511329	-0.894687	-1.0024	0.994081	0.91612	1
0.77122	0.941312	-0.958283	-0.940101	0.746424	0.920733	1
0.187787	0.390418	-0.619932	-0.7434	0.333136	0.641689	1
0.110907	0.122485	-0.442033	-0.580176	0.00985161	0.0280635	1
1.17E-05	2.20E-05	3.11797	2.94852	6.79E-06	2.48E-05	1
0.585893	0.337135	-0.921351	-0.981531	0.180921	0.256282	1
0.838944	0.762293	-0.848081	-0.858632	0.514429	0.39512	1
0.93901	0.802895	-0.856495	-0.800658	0.626698	0.89394	1
0.208855	0.0414907	-0.511854	-0.452254	0.174328	0.0782586	1
0.568504	0.685248	-0.78962	-0.841112	0.0917301	0.247465	1
0.599951	0.820557	-0.797387	-0.711467	0.405071	0.677063	1
0.719089	0.426059	-0.837139	-0.885139	0.923164	0.704495	1
0.656972	0.888807	-0.816022	-0.847692	0.322782	0.232452	1

: regions without perfect proxy, but with imputed SNP data in WTCCC GWAS

						combined NBS and 58C imputed control data		
	wtccc file/phenotype	imputed rsid	pos	allele1	allele2	controls_AA	controls_AB	controls_BB
rs1464510	./snptest_BD_03.txt	rs1464510	189595256	A	C	591	1471	873
rs1464510	./snptest_CAD_03.txt	rs1464510	189595256	A	C	591	1471	873
rs1464510	./snptest_CD_03.txt	rs1464510	189595256	A	C	591	1471	873
rs1464510	./snptest_HT_03.txt	rs1464510	189595256	A	C	591	1471	873
rs1464510	./snptest_RA_03.txt	rs1464510	189595256	A	C	591	1471	873
rs1464510	./snptest_T1D_03.txt	rs1464510	189595256	A	C	591	1471	873
rs1464510	./snptest_T2D_03.txt	rs1464510	189595256	A	C	591	1471	873
rs6822844	./snptest_BD_04.txt	rs6822844	123867026	G	T	1886	944	102
rs6822844	./snptest_CAD_04.txt	rs6822844	123867026	G	T	1886	944	102
rs6822844	./snptest_CD_04.txt	rs6822844	123867026	G	T	1886	944	102
rs6822844	./snptest_HT_04.txt	rs6822844	123867026	G	T	1886	944	102
rs6822844	./snptest_RA_04.txt	rs6822844	123867026	G	T	1886	944	102
rs6822844	./snptest_T1D_04.txt	rs6822844	123867026	G	T	1886	944	102
rs6822844	./snptest_T2D_04.txt	rs6822844	123867026	G	T	1886	944	102
rs1738074	./snptest_BD_06.txt	rs1738074	159436386	C	T	929	1455	545
rs1738074	./snptest_CAD_06.txt	rs1738074	159436386	C	T	929	1455	545
rs1738074	./snptest_CD_06.txt	rs1738074	159436386	C	T	929	1455	545
rs1738074	./snptest_HT_06.txt	rs1738074	159436386	C	T	929	1455	545
rs1738074	./snptest_RA_06.txt	rs1738074	159436386	C	T	929	1455	545
rs1738074	./snptest_T1D_06.txt	rs1738074	159436386	C	T	929	1455	545
rs1738074	./snptest_T2D_06.txt	rs1738074	159436386	C	T	929	1455	545
rs653178	./snptest_BD_12.txt	rs653178	110470476	C	T	766	1449	718
rs653178	./snptest_CAD_12.txt	rs653178	110470476	C	T	766	1449	718
rs653178	./snptest_CD_12.txt	rs653178	110470476	C	T	766	1449	718
rs653178	./snptest_HT_12.txt	rs653178	110470476	C	T	766	1449	718
rs653178	./snptest_RA_12.txt	rs653178	110470476	C	T	766	1449	718
rs653178	./snptest_T1D_12.txt	rs653178	110470476	C	T	766	1449	718
rs653178	./snptest_T2D_12.txt	rs653178	110470476	C	T	766	1449	718

: regions without perfect proxy, but with imputed SNP data in WTCCC GWAS

_NULL	cases_AA_exp	cases_AB_exp	cases_BB_exp	cases_NULL_exp	P 2x2 allele count chi-sq (2 tailed)
3	366	918	583	1	0.333
3	381	961	582	2	0.684
3	348	886	512	2	0.919
3	387	944	620	1	0.256
3	343	932	584	1	0.107
3	417	961	583	2	0.578
3	375	951	597	1	0.348
6	1228	576	62	2	0.320
6	1254	606	63	3	0.507
6	1156	530	59	3	0.230
6	1287	587	76	2	0.441
6	1259	543	56	2	0.017
6	1320	579	62	2	0.041
6	1257	602	62	3	0.406
9	602	914	347	5	0.781
9	619	937	365	5	0.957
9	522	851	370	5	0.039
9	632	947	367	6	0.804
9	644	886	325	5	0.049
9	659	976	323	5	0.047
9	608	976	335	5	0.588
5	513	877	474	4	0.828
5	561	922	439	4	0.023
5	523	818	404	3	0.015
5	565	914	469	4	0.111
5	543	882	431	4	0.036
5	685	922	353	3	1E-13
5	506	921	493	4	0.644

CLAIMS

1. A method of diagnosing coeliac disease, said method comprising analysing a sample of nucleic acid from a human subject to determine the presence or absence of one or more single nucleic polymorphisms (SNPs) in one or more human chromosomal regions selected from the group consisting of 1q31, 2q11-2q12, 3p21, 3q25-3q26, 3q28, 6q25 and 12q24.

2. A method according to claim 1, wherein said one or more SNPs are selected from the group consisting of the SNPs present in the following sequences: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12 and SEQ ID NO: 13.

3. A method according to claim 2, wherein said one or more SNPs comprises the SNPs present in the following sequences: SEQ ID NO: 1, SEQ ID NO: 2 and/or SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 and/or SEQ ID NO: 6, SEQ ID NO: 7 and/or SEQ ID NO: 8 and/or SEQ ID NO: 9 and/or SEQ ID NO: 10, SEQ ID NO: 11, and SEQ ID NO: 12 and/or SEQ ID NO: 13.

4. A method according to any one of claims 1 to 3, wherein said method further comprises analysing a sample of nucleic acid from said human subject to determine the presence or absence of one or more SNPs in the human chromosomal region 4q27.

5. A method according to claim 4, wherein said one or more SNPs are selected from the group consisting of the SNPs present in the following sequences: SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20 and SEQ ID NO: 21.

6. A method according to claim 5, wherein said one or more SNPs comprise the SNPs present in the following sequences: SEQ ID NO: 14 and/or SEQ ID NO: 15

and/or SEQ ID NO: 16 and/or SEQ ID NO: 17 and/or SEQ ID NO: 18 and/or SEQ ID NO: 19 and/or SEQ ID NO: 20 and/or SEQ ID NO: 21.

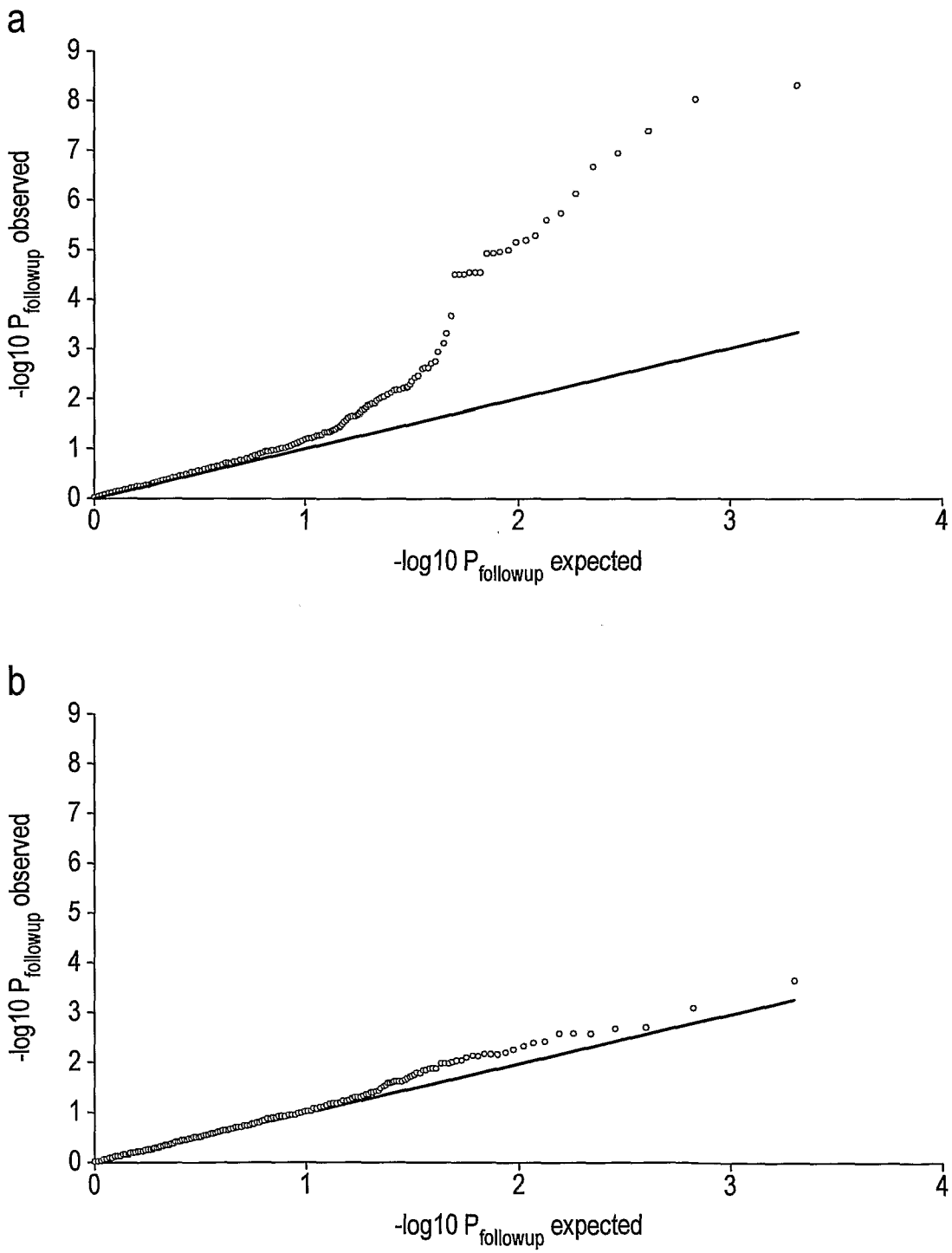
a
-log10 P followup observed
-log10 P followup expected
0
1
2
3
4
5
6
7
8
9
b
-log10 P followup observed
-log10 P followup expected

FIG. 1

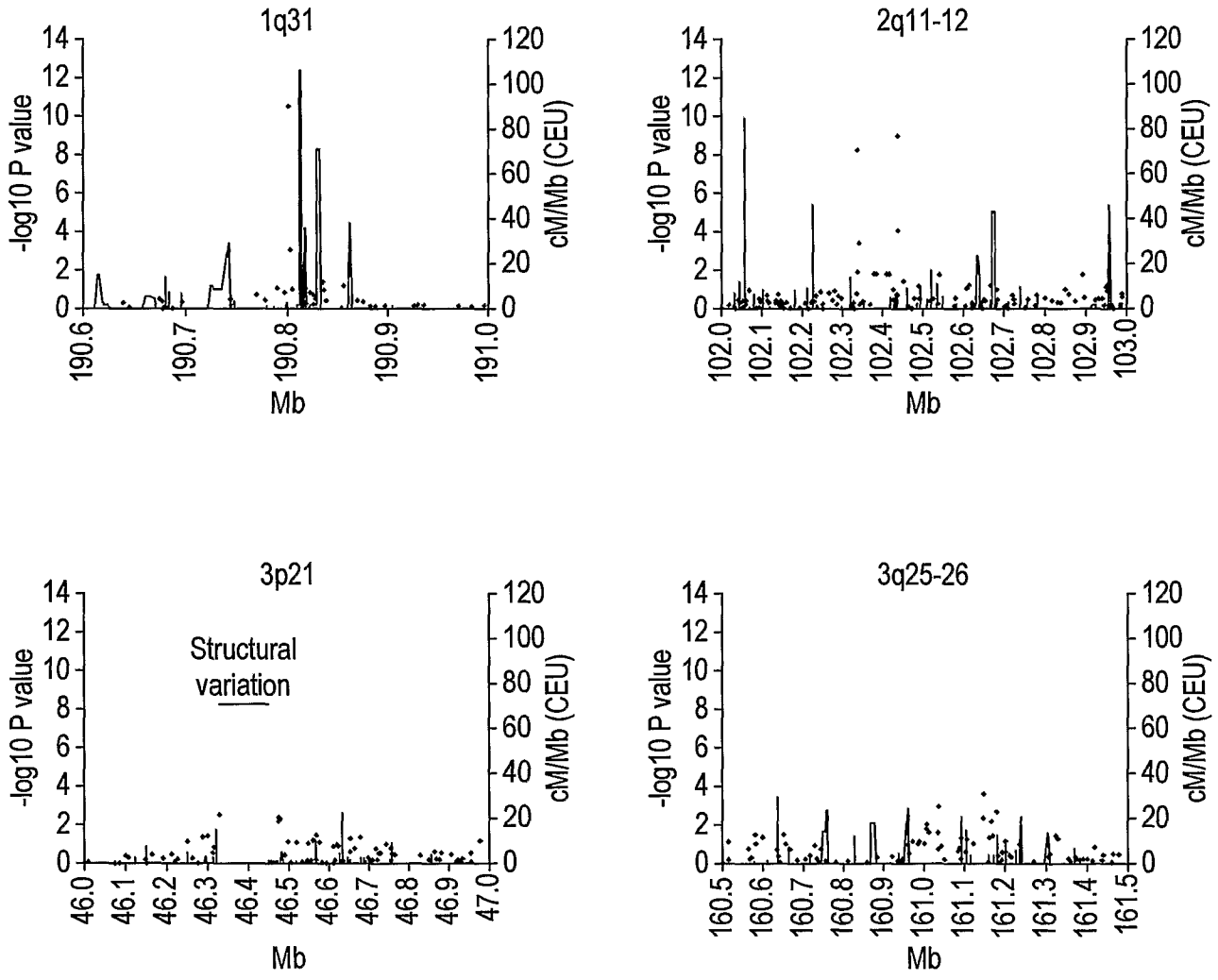
1q31
-log10 P value
cM/Mb (CEU)
Mb
2q11-12
3p21
Structural variation
3q25-26

FIG. 2

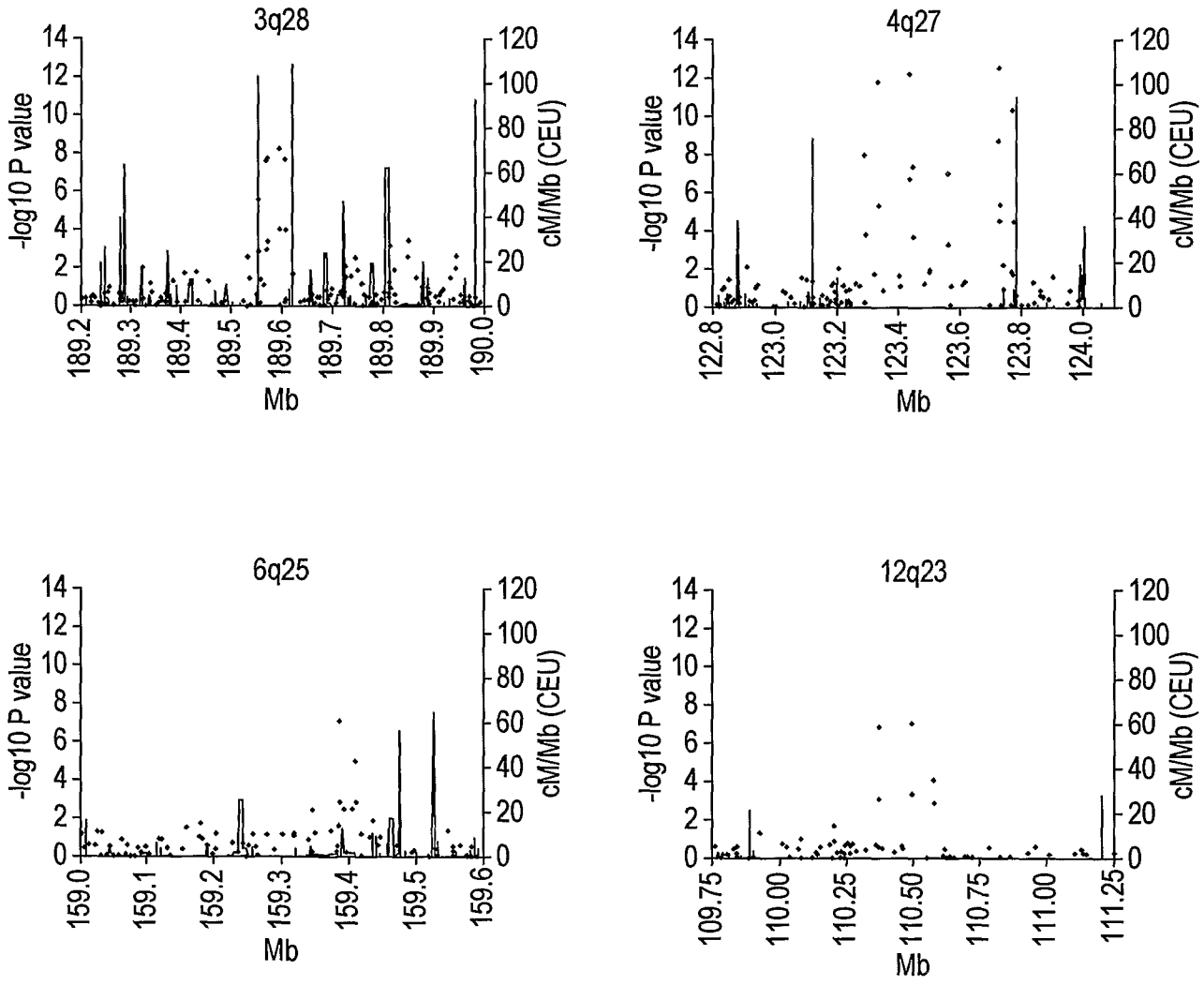
3q28
4q27
6q25
12q23
-log10 P value
cM/Mb (CEU)
Mb
189.2
189.3
189.4
189.5
189.6
189.7
189.8
189.9
190.0
122.8
123.0
123.2
123.4
123.6
123.8
124.0
159.0
159.1
159.2
159.3
159.4
159.5
159.6
109.75
110.00
110.25
110.50
110.75
111.00
111.25

FIG. 2 CONT'D

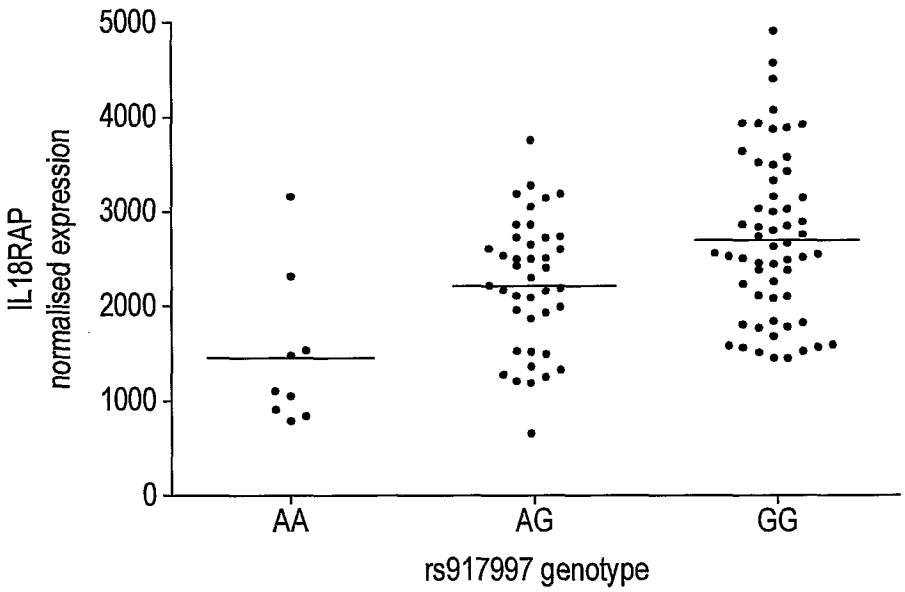
5000
4000
3000
2000
1000
0
IL18RAP
normalised expression
AA
AG
GG
rs917997 genotype

FIG. 3

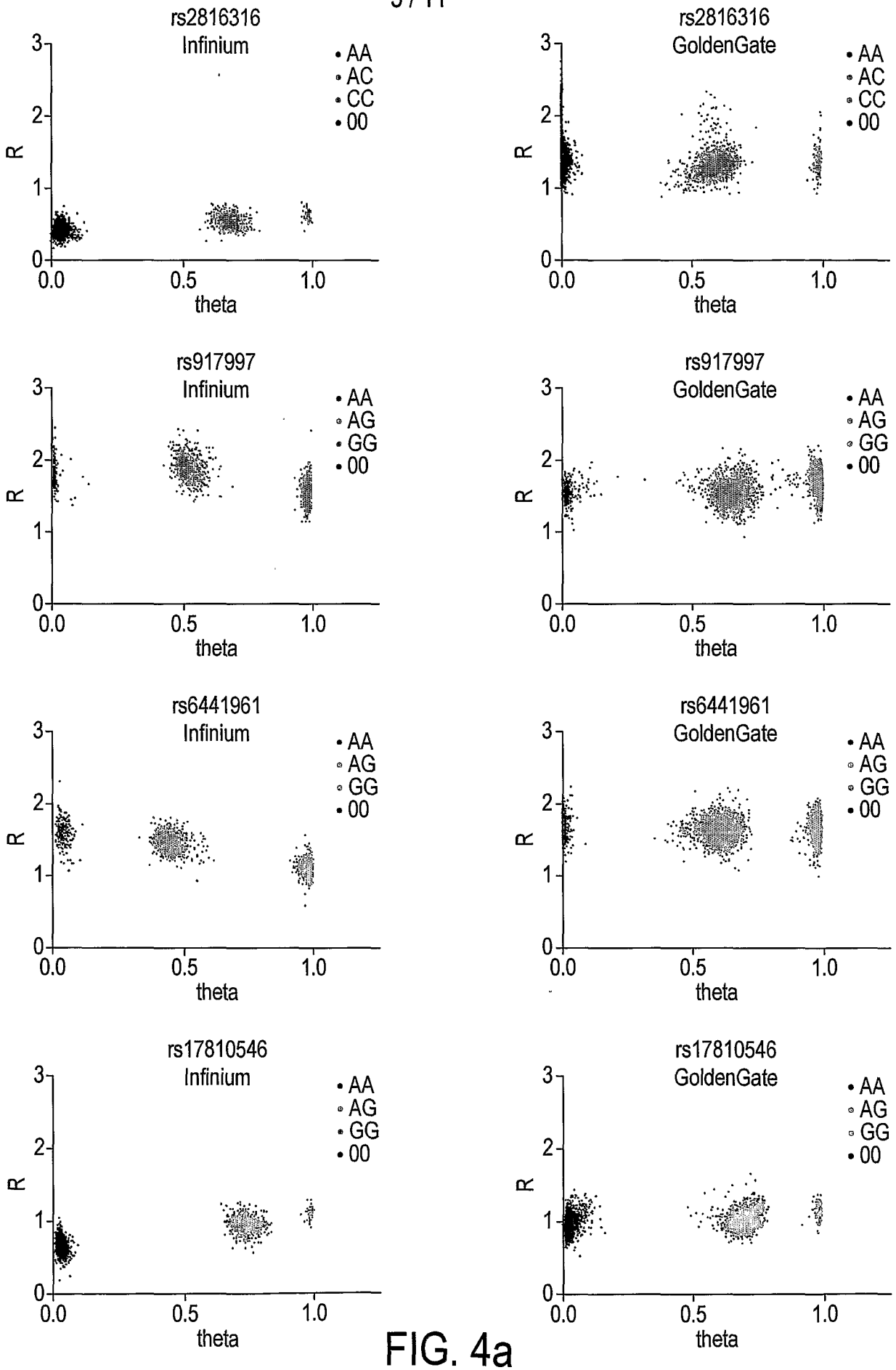
rs2816316
Infinium
rs2816316
GoldenGate
rs917997
Infinium
rs917997
GoldenGate
rs6441961
Infinium
rs6441961
GoldenGate
rs17810546
Infinium
rs17810546
GoldenGate
AA
AC
CC
AG
GG
00
R
theta
0.0
0.5
1.0
0
1
2
3

FIG. 4a

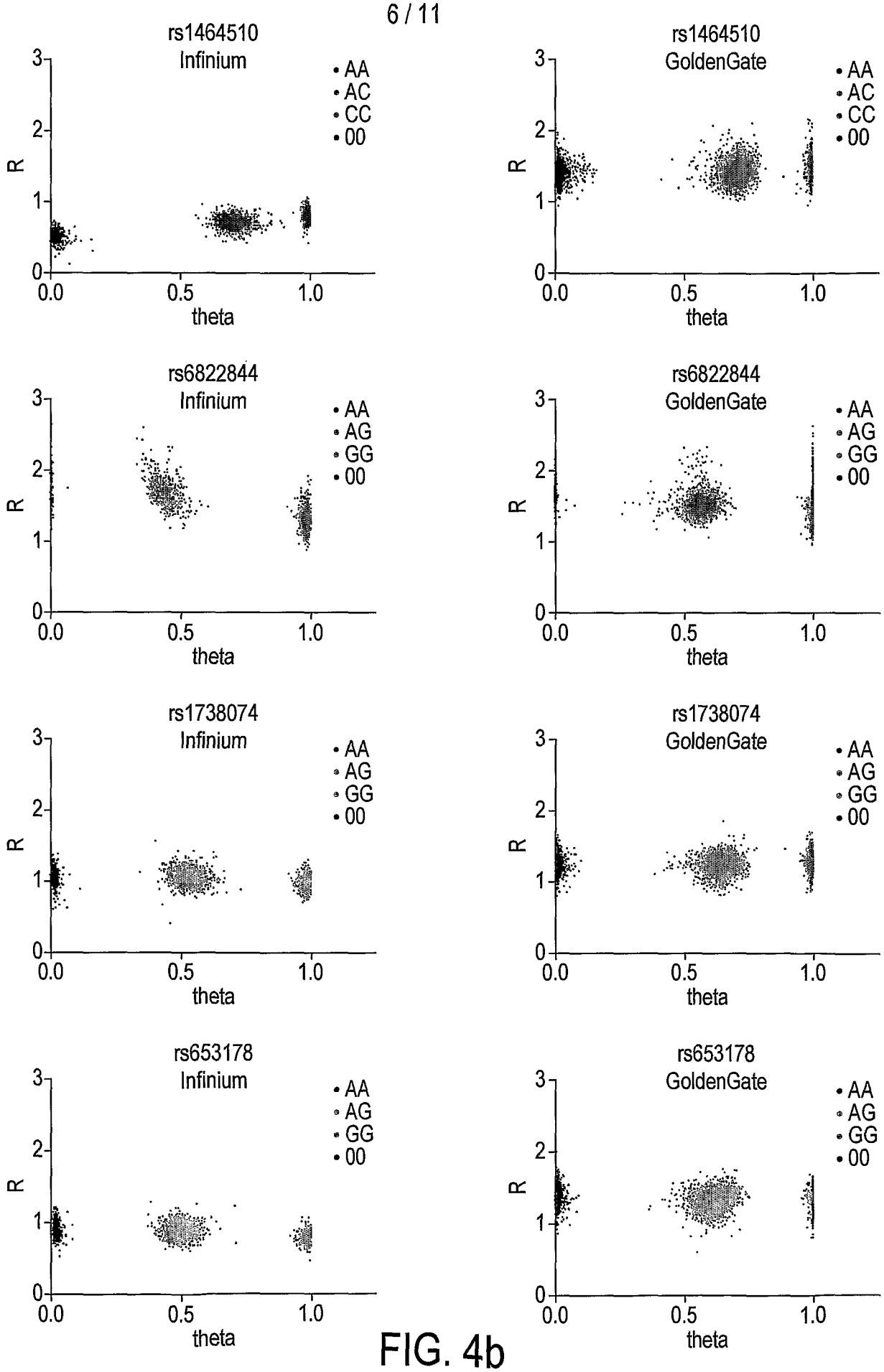
rs1464510
Infinium
• AA
• AC
• CC
• 00
R
theta
rs1464510
GoldenGate
• AA
• AC
• CC
• 00
R
theta
rs6822844
Infinium
• AA
• AG
• GG
• 00
R
theta
rs6822844
GoldenGate
• AA
• AG
• GG
• 00
R
theta
rs1738074
Infinium
• AA
• AG
• GG
• 00
R
theta
rs1738074
GoldenGate
• AA
• AG
• GG
• 00
R
theta
rs653178
Infinium
• AA
• AG
• GG
• 00
R
theta
rs653178
GoldenGate
• AA
• AG
• GG
• 00
R
theta

FIG. 4b

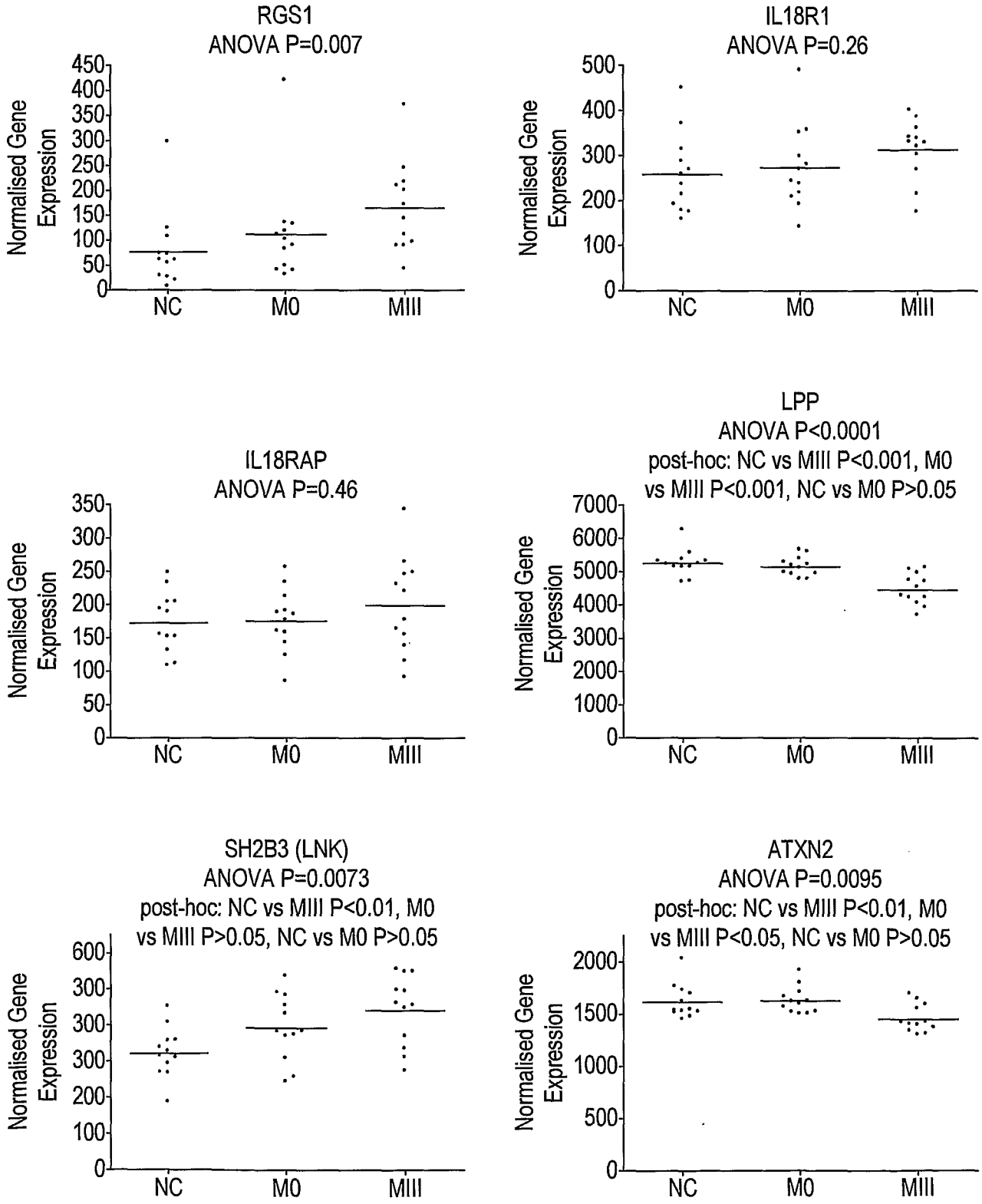
RGS1
ANOVA P=0.007
Normalised Gene Expression
450
400
350
300
250
200
150
100
50
0
NC
M0
MIII
IL18R1
ANOVA P=0.26
Normalised Gene Expression
500
400
300
200
100
0
NC
M0
MIII
IL18RAP
ANOVA P=0.46
Normalised Gene Expression
350
300
250
200
150
100
50
0
NC
M0
MIII
LPP
ANOVA P<0.0001
post-hoc: NC vs MIII P<0.001, M0 vs MIII P<0.001, NC vs M0 P>0.05
Normalised Gene Expression
7000
6000
5000
4000
3000
2000
1000
0
NC
M0
MIII
SH2B3 (LNK)
ANOVA P=0.0073
post-hoc: NC vs MIII P<0.01, M0 vs MIII P>0.05, NC vs M0 P>0.05
Normalised Gene Expression
600
300
300
300
200
100
0
NC
M0
MIII
ATXN2
ANOVA P=0.0095
post-hoc: NC vs MIII P<0.01, M0 vs MIII P<0.05, NC vs M0 P>0.05
Normalised Gene Expression
2000
1500
1000
500
0
NC
M0
MIII

FIG. 5

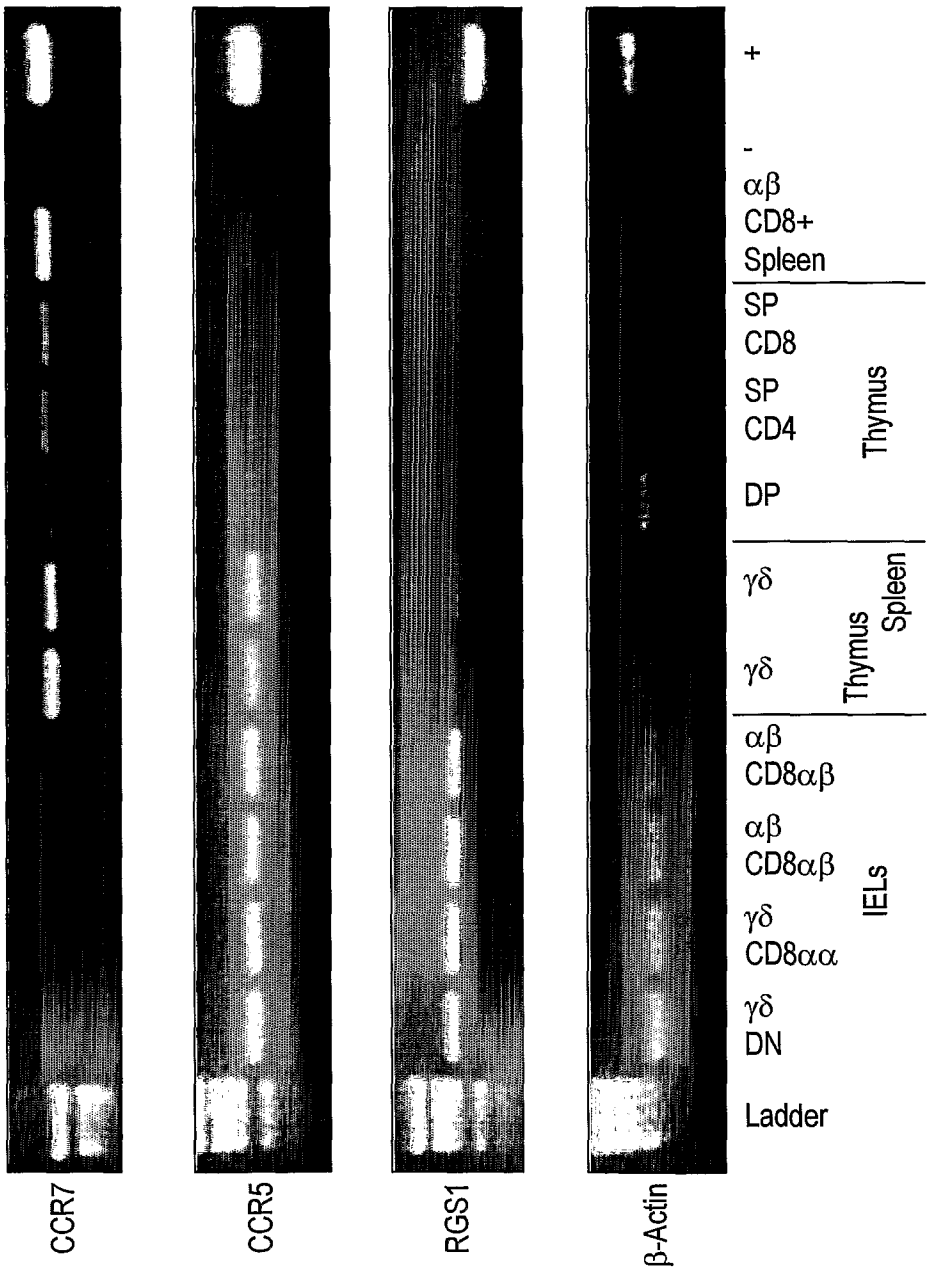
CCR7
CCR5
RGS1
β-Actin
+
-
αβ CD8+ Spleen
SP CD8
SP CD4
DP
Thymus
γδ
γδ
Thymus
Spleen
αβ CD8αβ
αβ CD8αβ
γδ CD8αα
γδ DN
IELs
Ladder

FIG. 6

SEQ ID NO: 1

rs2816316 *[Homo sapiens]*

GCTCCCTCCTGTTGTGGAGGAATATT[G/T]AGTTGTCTGTTGTTTTAGATAGGAT

SEQ ID NO: 2

rs13015714 *[Homo sapiens]*

TCGGCTATGGGTTTCCCTTTTCCTTT[G/T]GTTAAATAACAGTTCTGCCACAAAA

SEQ ID NO: 3

rs917997 *[Homo sapiens]*

GCTAAGGTCAAAACGCTGGACATCTG[A/G]ATAGCTTGGTTCTAGCATTATCTAT

SEQ ID NO: 4

rs6441961 *[Homo sapiens]*

AGCAATTATTTCCACTTGGTTATAGG[C/T]AGCCTTGATAACATCTCCAGCAGCT

SEQ ID NO: 5

rs17810546 *[Homo sapiens]*

AAAAAATTGTGTCCTGTTTAGACATC[A/G]TACCACAGAAAGCCATACAGAAAAC

SEQ ID NO: 6

rs9811792 *[Homo sapiens]*

ATAAAAACAGGAAAAGGTCTGATGTG[C/T]AATTAGAGCTTCAGAAGGGTAGGAA

SEQ ID NO: 7

rs9851967 *[Homo sapiens]*

TTTACTCTTTGTAAAAGAATCACAGA[C/T]TCAGGTTGGATGGAAGGATTCTTCT

FIG. 7

SEQ ID NO: 8

rs13076312 *[Homo sapiens]*

CTACCATGACCAGGTATCAGACATGC[C/T]ACTGGAACTGTGGACATCTGGCAGG

SEQ ID NO: 9

rs1464510 *[Homo sapiens]*

GCAACACAGTAAAAATGAACCAGGCT[G/T]TTTCAAATCAGTTTCAGAATCACAT

SEQ ID NO: 10

rs1559810 *[Homo sapiens]*

AAAGGCATTGAGCTGTGACTAAAGAG[G/T]CCAGGTTTCTAGTTTGAGCTCTAAC

SEQ ID NO: 11

rs1738074 *[Homo sapiens]*

CAGTGGACTAGAAGGAGCAGAGAGTT[A/G]TGCTGTTTCTCCCATTCTTTACAGC

SEQ ID NO: 12

rs3184504 *[Homo sapiens]*

CTTGCTCCAGCATCCAGGAGGTCCGG[C/T]GGTGCACACGGCTTGAGATGCCTGA

SEQ ID NO: 13

rs653178 *[Homo sapiens]*

GTGCCTAATGGCTGCAATATTGGACA[A/G]CATGACATAGGACATCTTCATCATT

SEQ ID NO: 14

rs11938795 *[Homo sapiens]*

ACAAATAAATATCTGCGACATTCCTC[C/T]AATTCCAAGAAGTCGTTTGCCTTGG

SEQ ID NO: 15

rs13151961 *[Homo sapiens]*

CATTATGCTAAAACAATACTTCATTA[A/G]GTTCTGCATTAGGTCATTTGTTAAA

FIG. 7 CONT'D

SEQ ID NO: 16

rs13119723 *[Homo sapiens]*

AGACAGAATGGGGAAAATTTTTTTAA[A/G]TCTACCCATCTGACAGATGTAATAT

SEQ ID NO: 17

rs11734090 *[Homo sapiens]*

TTTTATTTTGCTGAATCCCATGTACA[C/T]GTTTTTCAGTTCTCATTTTACTGAC

SEQ ID NO: 18

rs7684187 *[Homo sapiens]*

TTATTTACTGTTATTGCTGTTACTCA[A/G]CATTTTACTGGAGGCATTAATACAT

SEQ ID NO: 19

rs12642902 *[Homo sapiens]*

AGTCTGGAAAGAGCAGTGTTTGGGTG[A/G]ATCTGATTTGCAACTTCTCACTGCT

SEQ ID NO: 20

rs6822844 *[Homo sapiens]*

CCCTGTCTCGCTCTCCATAGCAAAAA[G/T]AGAGGACTCTTTTCATGTTGCCACT

SEQ ID NO: 21

rs6840978 *[Homo sapiens]*

ACCAATATGCATTTGAGATGATTTTT[C/T]AGTTACATGGAGTTTTCCTTCCATA

FIG. 7 CONT'D

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2009/000475

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, EMBASE, MEDLINE, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	RUEDA B ET AL: "CTLA4/CT60 polymorphism is not relevant in susceptibility to autoimmune inflammatory intestinal disorders" HUMAN IMMUNOLOGY, NEW YORK, NY, US, vol. 66, no. 3, 1 March 2005 (2005-03-01), pages 321-325, XP025379671 ISSN: 0198-8859 [retrieved on 2005-03-01] page 321 - page 324 ----- -/--	1-6

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

[X] See patent family annex.

* Special categories of cited documents :

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

"E" earlier document 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
30 July 2009	06/08/2009

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

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2009/000475

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	AMUNDSEN S S ET AL: "Association Analysis of MYO9B Gene Polymorphisms with Celiac Disease in a Swedish/Norwegian Cohort" HUMAN IMMUNOLOGY, NEW YORK, NY, US, vol. 67, no. 4-5, 1 April 2006 (2006-04-01), pages 341-345, XP024993263 ISSN: 0198-8859 [retrieved on 2006-04-01] page 341 - page 344	1-6
A	CICLITIRA P J ET AL: "The pathogenesis of coeliac disease" MOLECULAR ASPECTS OF MEDICINE, PERGAMON PRESS, OXFORD, GB, vol. 26, no. 6, 1 December 2005 (2005-12-01), pages 421-458, XP025272604 ISSN: 0098-2997 [retrieved on 2005-12-01] the whole document	1-6
A	WO 2007/028795 A (INNOGENETICS NV [BE]; NUYTINCK LIEVE [BE]) 15 March 2007 (2007-03-15) the whole document	1-6
A	WO 2007/023148 A (INNOGENETICS NV [BE]; NUYTINCK LIEVE [BE]) 1 March 2007 (2007-03-01) the whole document	1-6

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2009/000475

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007028795 A	15-03-2007	NONE	
WO 2007023148 A	01-03-2007	NONE	