

Variants within protectin (CD59) and CD44 genes linked to an inherited haplotype in a family with coeliac disease

C. Vidal^{1,2}, J. Borg¹, A. Xuereb-Anastasi^{2,3} & C. A. Scerri^{1,2}

¹ Laboratory of Molecular Genetics, Department of Physiology and Biochemistry, University of Malta, Msida, Malta

² Institute of Healthcare, University of Malta, Msida, Malta

³ DNA Laboratory, Department of Pathology, Medical School, University of Malta, Msida, Malta

Key words

CD59; coeliac disease; human leukocyte antigen linkage; protectin; RNA splicing

Correspondence

Dr Christian A. Scerri, MD, PhD
Laboratory of Molecular Genetics
Department of Physiology and Biochemistry
University of Malta
Msida
Malta
Tel: +356 2340 2774
Fax: +356 2134 3535
e-mail: cas@biotech.um.edu.mt

Received 26 May 2008; revised 17 September 2008, 15 November 2008, 21 November 2008; accepted 24 November 2008

doi: 10.1111/j.1399-0039.2008.01193.x

Abstract

Coeliac disease (CD) is an autoimmune disorder characterised by inflammation, villous atrophy and hyperplasia of the small intestinal mucosa that affects genetically susceptible individuals. A genome-wide scan was performed in 17 family members with high incidence of CD. Highest nonparametric linkage (NPL) and logarithm of odds (LOD) scores were of 6.21 ($P = 0.0107$) and 2.57, respectively, to a region on chromosome 11p13-12. Following fine mapping, NPL and LOD scores did not change, but the linkage interval on chromosome 11 was narrowed to a region that is approximately 50.94 cM from pTer. Two inherited haplotypes on chromosomes 11p13-12 and 9q21 were observed in all affected members but not in the majority of clinically normal individuals. Sequencing of genes at region 11p13-12 showed a number of sequence variants, two of which were linked with the inherited haplotype. One of these variants in the CD59 gene was found at a very low frequency in the population and could possibly affect pre-messenger RNA splicing. This study is of particular importance for the identification of novel genes that might be responsible for CD other than human leukocyte antigen.

Introduction

Gluten sensitive enteropathy, or coeliac disease (CD) (OMIM 212750), is a complex disorder with both genetic and environmental factors playing a role in its pathogenesis. The disease is characterised by an inflammatory process triggered in genetically susceptible individuals after the ingestion of gluten found in wheat, rye and barley, which leads to complications mostly associated with mucosal injury and malabsorption (1).

A number of linkage and association studies have been carried out to try to identify genes that might increase the individual's susceptibility for CD. The human leukocyte antigen (HLA) locus on chromosome 6 was the most commonly reported locus (2–4), although, recently, linkage to this region was not confirmed in a large Bedouin family (5). Other chromosomal loci were identified by genome-wide linkage scans including 5q31-q33 (6), 10q23, 15q26, 16q23, 19p13.1, 6p12, 2q and 11p11 (7–9), with the latter

being further confirmed in a later study (10). A number of single nucleotide polymorphisms (SNPs) within candidate genes were studied for an association with the disease. A variant within the promoter region of the type II transmembrane protein (CD209) gene expressed in macrophages and dendritic cells was significantly associated with CD in HLA-DQ2-negative coeliac individuals (11). Polymorphisms, within the cytotoxic T lymphocyte (CTL)-associated-4 (CTLA4) gene, were extensively studied from which conflicting results were obtained (12, 13). An extended haplotype within the genes CD28/CTLA4 and within the gene coding for inducible costimulator (ICOS) was significantly associated with the disease (12, 14). Other genes associated with the disease included those coding for myosin IXB (MYO9B) and CD14, which, although having different mechanisms, both can affect intestinal permeability (15, 16). A genome-wide association study performed in the UK did not confirm regions where previously

associated genes are found (CTLA4 and MYO9B) but identified a locus on chromosome 4q27 very close to interleukin 2 (IL2) and IL21 and another predicted gene of unknown function (KIAA1109) (17). These findings were later confirmed in Swedish/Norwegian families where it was concluded that the KIAA1109/Tenr/IL2/IL21 region is a true susceptibility region for CD (18). A series of novel regions including IL2-IL21, CCR3 and SH2B3 were recently identified after further follow-up studies of the initial genome-wide association (19). Other genes were indicated by gene expression studies performed in normal and affected individuals with and without a gluten-free diet (20).

In this study, a genome-wide scan was performed in an extended family with a high incidence of CD to identify possible chromosomal loci that might contain genes responsible for the disease. Sequencing of genes at the indicated locus showed an unreported variant within the gene coding for protectin (CD59) (MIM 107271), which was further studied for its possible effects on pre-messenger RNA (mRNA) splicing.

Materials and methods

Subjects

Seventeen members (eight males and nine females) from an extended Maltese family with a high incidence of CD were

recruited for this study (Figure 1). Recruitment was achieved through the Malta Coeliac Association that referred interested individuals to the Molecular Genetics Clinic, Medical School, St Luke's Hospital, G'Mangia, Malta. All participants were asked to answer a questionnaire about their medical history, use of medications, family history of CD and adherence to a gluten-free diet. This study was approved by the Research Ethics Committee, University of Malta, and was funded by the Malta Council for Science and Technology as part of the Research and Development (2004) programme.

As shown in Figure 1, the proband of this family was a 58-year-old male (II: 3) who participated in a previous population study (13) and presented at the clinic with classical symptoms of CD. Serological tests showed elevated antitransglutaminase (tTg), antiendomysium (EmA) and antigliadin (AGA) immunoglobulin (IgA) antibodies. Diagnosis was confirmed by biopsy, which showed histological features including villous atrophy (Marsh 3) consistent with CD. Further follow-ups of this patient over a period of 4 years showed an improvement in serological results, after he followed a gluten-free diet. Another 16 family members from three generations were recruited, 3 of whom were diagnosed with CD by biopsy. His 60-year-old sibling (II: 6) was diagnosed as coeliac 4 years ago when she had a slightly elevated AGA IgA and IgG, a positive EmA and a positive anti-tTg (87 U/ml). Gastric and duodenal

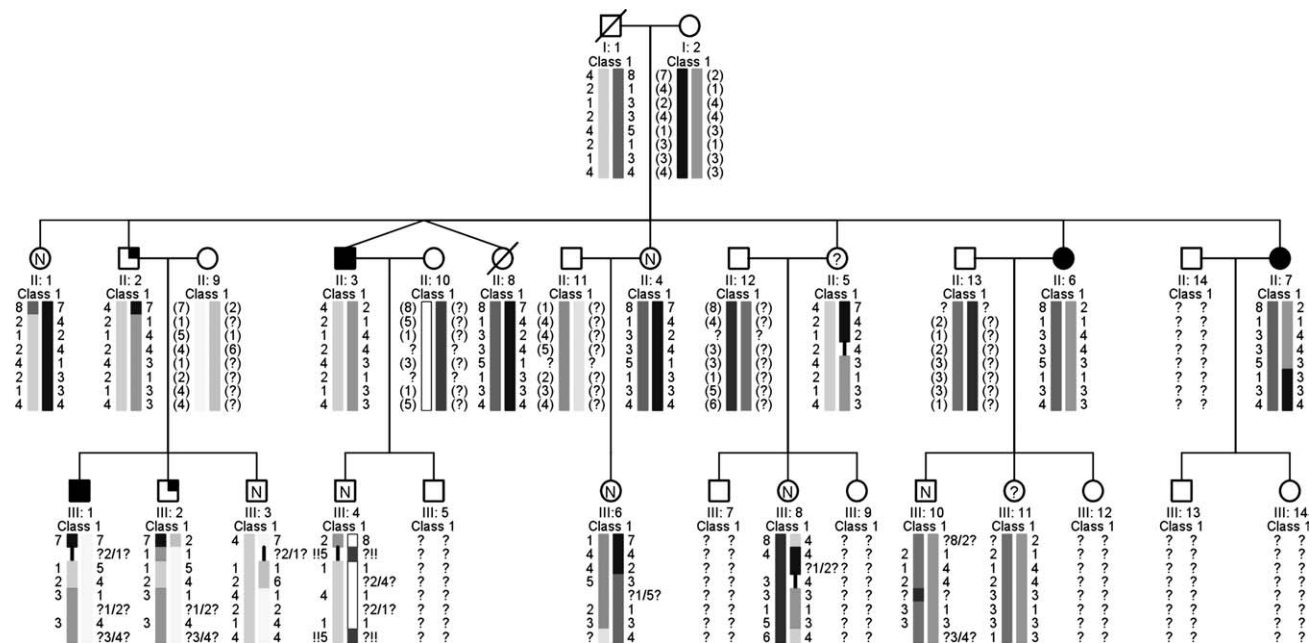


Figure 1 Inheritance of STRs on chromosome 11p13. Affected males and females are shown in black shading; females labelled (?) were reported to have coeliac-like symptoms. Males and females labelled (N) tested individuals showing no evidence of celiac. One-quarter black-shaded individuals only had high antigliadin (AGA). Markers shown top to bottom are D11S4190, D11S904, D11S914, D11S4101, D11S1392, D11S935, D11S4102 and D11S1911. Inferred genotypes by GENEHUNTER are in parentheses.

biopsies showed a chronic gastritis together with villous atrophy, increased cellularity of the lamina propria and increased intraepithelial lymphocyte population. This individual was also instructed to follow a gluten-free diet after which her serological tests improved. Besides being coeliac, this individual was also osteoporotic and asthmatic. She also had a daughter (III: 11) who was possibly affected, but she was following a gluten-free diet and therefore her serological tests appeared to be normal. She had symptoms related to CD and was hospitalised on one occasion after being admitted with severe epigastric pain and haematemesis. Her other aunt (II: 7) had elevated anti-tTg IgA/IgG and was positive for EmA, while her duodenal biopsy confirmed villous atrophy that was consistent with coeliac. A number of follow-ups performed later after she started a gluten-free diet showed that her anti-tTg and AGA went down to normal. Another sibling of the proband was also possibly affected by the condition (individual II: 5) as she was symptomatic before starting a gluten-free diet. Biopsy did not show any cellular characteristics related to CD in this individual, although a weak positive EmA and borderline AGA IgA were reported.

The proband's nephew (III: 1) was diagnosed with CD at the age of 11 years and his intestinal biopsy showed villous atrophy consistent with coeliac. His serological investigations showed elevated anti-tTg (217 U/ml), strong positive EmA and both AGA IgA and IgG were above normal. The patient was instructed to follow a gluten-free diet and a follow-up performed 2 years later showed normal biopsy and serology with only a slightly elevated anti-tTg IgA. His father (II: 2) and one brother (III: 2) did not have any symptoms of CD and were not on a gluten-free diet, but both had slightly elevated AGA IgA or IgG with normal anti-tTg. Serological tests performed following gluten challenge, resulted in an increase in AGA IgG, a positive EmA and positive antireticulin.

Peripheral blood was collected by venipuncture from all participants for DNA analysis and immunological tests. All serology was performed at the Immunology Laboratory, Department of Pathology, St Luke's Hospital, Malta.

DNA analysis

DNA was extracted from peripheral blood leucocytes using a standard laboratory protocol (21). For the initial genome-wide scan, 400 microsatellite markers spread across the 22 autosomes and x-chromosome with an average spacing of 8.63 cM and heterozygosity of 0.77 were used. Fine mapping was performed by increasing the markers at indicated loci from the initial scan. Genotyping was performed by polymerase chain reaction (PCR) followed by fragment analysis using a 3730xl ABI genetic analyser (Applied Biosystems, Foster City, CA). The average performance of the markers was of 96.02%. Genotyping

was performed commercially at the McGill University and Genome Quebec Innovation Centre, Quebec, Canada.

Linkage analysis

Parametric and nonparametric linkage analyses (two point and multipoint) were performed using the software SUPERLINK v1.6 (22) and GENEHUNTER v1.2, respectively, using the EASYLINKAGE v5.05 (23) platform. The Kong and Cox algorithm (24) was used to calculate Zlr scores that are less conservative where inheritance data are incomplete. Parametric analysis was performed using variable penetrances for both dominant and recessive models. For the dominant model, initial penetrances used were 0.01, 0.90 and 0.90 for the wild-type homozygote, mutant heterozygote and mutant homozygote, respectively. Penetrances of 0.01, 0.01 and 0.80 were used for the recessive model. The initial disease allele frequency used for the analysis was of 0.001, equivalent to a population prevalence of 0.2%. A codominant allele frequency algorithm was used for the analysis, as suggested in the EASYLINKAGE manual to be the best option for extended pedigrees.

Because of the extensive phenotypical heterogeneity within this family, the analysis was performed using three different models. Model 1 affected status was defined as those individuals diagnosed as coeliac by biopsy, only. In model 2, two other individuals with coeliac-like symptoms (II: 5 and III: 11) were considered affected together with those diagnosed by biopsy. In model 3, another two individuals having an elevated AGA (II: 2 and III: 2) were also defined as affected. Individuals that were screened and tested negative were assumed to be normal, while those not tested were assumed to have an unknown phenotype. The sex-averaged deCode genetic map was used throughout the study.

HLA typing

HLA typing to detect HLA-DQ2 and/or HLA-DQ8 heterodimers was performed in all family members using a commercial kit for the detection of DQ CD alleles (DiaGene S.r.l., Palermo, Italy). A total of 12 PCRs for each individual were performed as instructed by supplier in a 15- μ l reaction followed by agarose gel electrophoresis.

Gene sequencing

All exons, including intron-exon boundaries, and up to 1 kb upstream from the transcriptional start sites of three genes found within the indicated locus by linkage, were sequenced. Sequencing was performed in two affected and two unaffected members (with and without the linked haplotype). The genes sequenced were those coding for protectin (CD59) (MIM 107271), CD44 (MIM 107269) and Apaf-1 interacting protein (APIP) (GeneID 51074), using

oligonucleotides designed for transcripts ENST00000351554, ENST00000278385 and ENST00000278359, respectively. Sequencing reactions were performed using BigDye[®]Terminators v3.1 and analysed on an ABI 3130 genetic analyser (Applied Biosystems).

Sequence variants, identified and observed to be inherited together with the linked haplotype, were further analysed in all family members either by direct DNA sequencing or by restriction fragment length polymorphism (RFLP). Nomenclature used refers to actual nucleotide changes in the respective polymorphisms.

RNA splicing analysis

A 482- and an 812-bp fragment containing mutated and wild-type exon 3 and adjacent introns of CD59 and CD44, respectively, were amplified by PCR using DNA from family members carrying normal and mutated alleles. Forward and reverse oligonucleotide primers were designed harbouring consensus sequences for restriction enzymes *Xho I* and *Pst I* at the 5' end, together with a noncomplementary tail (CD59_F: 5'-CAATGCGGAAGTTCGAG-TGTGGTAGTACACACTGGGGCTTAT-3'; CD59_R: 5'-CAATGCGGAAGTGCAGCACTGAGGACAAGG-TCATAACATAGG-3'; CD44_F: 5'-CAATGCGGAAGTTCGAGACTCAACACTGCTGAACACA-3'; CD44_R: 5'-CAATGCGGAAGTGCAGTCCAACACAAACCCGTAAAC-3'). The PCR products were initially cloned into a pGEM-T Easy (Promega Corporation, Madison, WI) vector and transformed into DH5- α (Invitrogen Ltd, Paisley, UK). Plasmid DNA was extracted from selected clones using Wizard[®]Plus SV Minipreps DNA purification system (Promega Corporation), digested and subcloned into the exon-trapping plasmid pSPL3 (generously donated by Dr Irene Bottillo, IRCC-CSS, San Giovanni Rotondo and CSS-Mendel Institute, Rome, Italy). The selected clones were checked by sequencing for their integrity and orientation followed by selection of mutated and normal alleles. Highly purified plasmid DNA for transfections was prepared using PureYield[™] Plasmid Midiprep System (Promega Corporation). The plasmid DNA concentration was measured by spectrophotometry and adjusted to 200 μ g/ml.

The purified minigene constructs of both wild-type and mutated alleles were transfected into HeLa and COS-7 cells (DSMZ, Braunschweig, Germany). Both cell lines were cultured in Dulbecco Modified Eagle's Medium (Sigma, St Louis, MO) containing 4 mM glutamine, 4.5 g/l glucose, supplemented with 10% foetal calf serum and incubated at 37°C in 5% CO₂. Twenty-four hours before transfection, 5×10^4 cells were seeded into each of a 24-well plate and incubated overnight in the same conditions described above. Three replicates for each construct were prepared together with positive (pSPL3 only) and negative (without

DNA) controls in two separate plates. Transient transfection was performed using Tfx[™]-20 (Promega Corporation), as recommended, using 1 μ g of each construct and a charge ratio of DNA to transfection reagent of 1:2 (COS-7) and 1:3 (HeLa). Incubation was carried out at 37°C (5% CO₂) for 48 h.

Total cellular RNA was extracted from HeLa and COS-7 cells, after 48 h incubation, using SV Total RNA isolation system (Promega Corporation). First-strand complementary DNA was synthesised by reverse transcription PCR using oligo-(dT)₁₆, followed by PCR using specific primers for the splice donor (SD6) and acceptor (SA2) sites of the pSPL3 plasmid. The products were analysed using 2% agarose gel electrophoresis, visualised under UV irradiation and analysed using IMAGEJ 1.38 \times software (<http://rsb.info.nih.gov/ij/>). All observed transcripts were extracted from gel and sequenced.

Results

Linkage study

Table 1 shows highest NPL and LOD scores observed following the initial genome-wide scan in this family, for all models tested. Chromosome 11p13-12 showed the highest LOD scores when using both two-point and multipoint parametric analysis at a slightly reduced penetrance (80%). The region on chromosome 9q21 showed a lower LOD score than that of chromosome 11, but similar NPL. A region of more than 50 cM on chromosome 10 failed to give consistent results across all models tested. Analysis was performed using a dominant model with penetrances ranging from 90 to 50% and 1% phenocopy rate. No evidence of linkage was observed to other regions including the HLA locus on chromosome 6 when using both dominant and recessive models.

High-resolution scan

Sixteen additional markers were added at the indicated loci on chromosomes 9 (D9S1780, D9S1680), 10 (D10S1709, D10S200, D10S1786, D10S187, D10S1792, D10S168, D10S597 and D10S1671) and 11 (D11S871, D11S914, D11S4101, D11S935, D11S4102 and D11S1911). The spacing between the markers was approximately of 2–5 cM. NPL and LOD scores obtained after increasing the markers remained the same as for the initial scan with a slight increase in marker information (up to 0.99 to marker D11S1392). The linkage interval on chromosome 11 was narrowed to a region between 48.21 and 52.95 cM with marker D11S1392 at 50.94 cM from pTer, showing the highest NPL/LOD (deCode genetic map). The linkage interval on both chromosomes 9 and 10 remained the same even after fine mapping.

Table 1 Two point and multipoint LOD scores and nonparametric analyses

Marker	Chromosome	Position (cM) ^a	LOD (two point)	Theta	LOD (multipoint) ^b	NPL (Zlr) ^c	P	Penetrances ^d
Model 1								
D11S1392	11	50.64	1.29	0.0000	1.66	2.08 (1.65)	0.1250	0.01, 0.80 and 0.80
D10S1731	10	134.96	0.66	0.0000	0.87	3.09 (2.63)	0.0313	0.01, 0.80 and 0.80
D9S167	9	81.01	0.37	0.0500	0.67	2.08 (1.65)	0.1250	0.01, 0.07 and 0.07
Model 2								
D11S1392	11	50.64	1.99	0.0000	2.25	4.12 (2.39)	0.0313	0.01, 0.80 and 0.80
D10S1731	10	134.96	1.53	0.0000	0.26	5.27 (3.31)	0.0039	0.01, 0.60 and 0.60
D9S167	9	81.01	1.18	0.0000	1.25	4.12 (2.39)	0.0313	0.01, 0.60 and 0.60
Model 3								
D11S1392	11	50.64	2.33	0.0000	2.57	6.21 (2.20)	0.0107	0.01, 0.80 and 0.80
D9S167	9	81.01	1.36	0.0000	1.58	6.21 (2.20)	0.0107	0.01, 0.70 and 0.70

NPL, nonparametric linkage; LOD, logarithm of odds.

^a Chromosomal position from pTer using deCode genetic map.^b Highest LOD scores obtained using a dominant model with 90% penetrance (0.01, 0.90 and 0.90).^c Zlr score using Kong and Cox algorithm (21).^d Penetrance frequencies for two-point parametric analysis: phenocopy, mutant heterozygote and mutant homozygote.

Two inherited haplotypes were observed on chromosomes 11 and 9, with a number of recombinations very close to the indicated markers. As shown in Figure 1, allele 3 at marker D11S1392 was inherited identical by descent in all affected individuals within this family but was not found in most unaffected members with the exception of individual III: 8. A haplotype on chromosome 9 (Figure 2) at markers D9S167, D9S1680 and D9S278 (alleles 2-1-3), showed a similar mode of inheritance to that of chromosome 11, with the exception of individual III: 4.

HLA typing

Because no evidence of linkage was observed to the HLA locus on chromosome 6, HLA typing of all family members was performed to determine whether they were HLA-DQ2 or HLA-DQ8. Table 2 shows the HLA types of all family members. Thirteen of 17 members carried the HLA-DQ2 heterodimer (76.5%). Also 88.2% of family members carried the allele DQA1*05 and 70.5% had the DQA1*0201 allele. The youngest affected individual (III: 1) was heterozygous for HLA-DQ2 and DQ8 also carrying

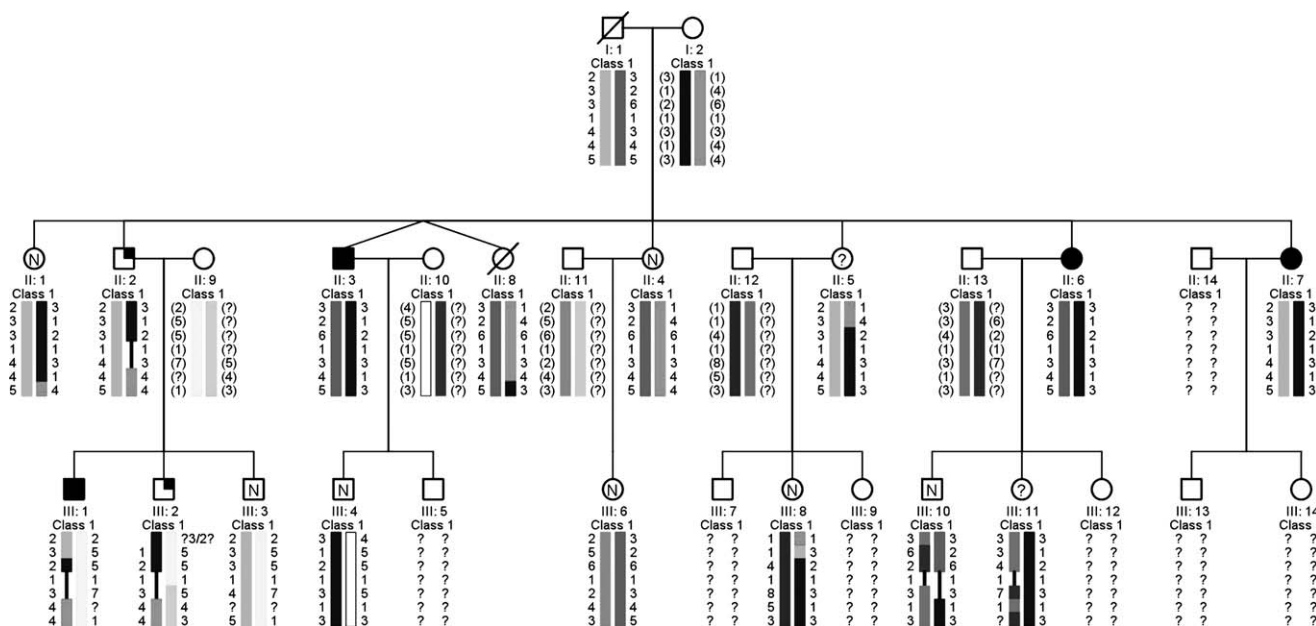


Figure 2 Inheritance of STRs on chromosome 9. Affected males and females are shown in black shading; females labelled (?) were reported to have coeliac-like symptoms. Males and females labelled (N) tested individuals showing no evidence of celiac. One-quarter black-shaded individuals only had high antigliadin (AGA). Markers shown top to bottom are D9S1122, D9S1780, D9S167, D9S1680, D9S278, D9S1689 and D9S938. Inferred genotypes by GENEHUNTER are in parentheses.

Table 2 HLA typing in family members and the GA variant identified in CD59^a

Family member	Status	Risk	DQA1		DQB1		DRB1	HLA type	CD59 G/A variant
I: 1	Not known	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07/11	DR7/DQ2 DR11/DQ7	GG
II: 1	Normal	No risk	05	ND	0301/04	ND	11/12	DR11/DR12/DQ7	GG
II: 2	High AGA	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07/12	DR7/DQ2 DR12/DQ7	GA
II: 3	Affected	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07/11	DR7/DQ2 DR11/DQ7	GA
II: 8	Normal	No risk	05	ND	0301/04	ND	11/12	DR11/DR12/DQ7	GG
II: 4	Normal	No risk	05	ND	0301/04	ND	11/12	DR11/DR12/DQ7	GG
II: 5	Symptomatic	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07	DR7/DQ2 DQ7	GA
II: 6	Affected	Moderate (DQ2.2)	0201	ND	02	ND	07	DR7/DQ2	GA
II: 7	Affected	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07/12	DR7/DQ2 DR12/DQ7	GA
III: 1	Affected	Moderate (DQ8)	0201	<u>03</u>	<u>02</u>	<u>0302</u>	07/04	DR7/DQ2 DR4/DQ8	GA
III: 2	High AGA	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07/12	DR7/DQ2 DR12/DQ7	GA
III: 3	Normal	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07/12	DR7/DQ2 DR12/DQ7	GG
III: 4	Normal	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07/11	DR7/DQ2 DR11/DQ7	GG
III: 6	Normal	High (DQ2.5 in <i>cis</i>)	<u>05</u>	ND	<u>02</u>	0301/04	07/11/12	DR7/DQ2 DR11/12/DQ7	GG
III: 8	Normal	No risk	05	ND	0301/04	ND	11/12	DR11/DR12/DQ7	GA
III: 10	Normal	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07/11	DR7/DQ2 DR11/DQ7	GA
III: 11	Symptomatic	High (DQ2.5 in <i>trans</i>)	0201	<u>05</u>	<u>02</u>	0301/04	07/11	DR7/DQ2 DR11/DQ7	GA

AGA, anti gliadin; ND, not determined; HLA, human leukocyte antigen.

^a Underlined and bold alleles are those increasing risk. DR11/12-DQ7 haplotype (DRB1*11/12, DQA1*05 and DQB1*0301/04); DR7-DQ2 haplotype (low-risk DQ2.2) – (DRB1*07, DQA1*0201 and DQB1*02). High-risk DQ2.5 (DQA1*05 and DQB1*02) (observed in *trans*).

the DR4 and DR7 alleles. None of the family members were carrying the DR3 allele. As shown in Table 2, most individuals were heterozygous DQ2.2 (DQA1*0201 and DQB1*02) and DQ7 (DQA1*05 and DQB1*0301/04), being also DQ2.5 in *trans* (DQA1*05 and DQB1*02). There were no differences in the pattern of inheritance of HLA alleles between affected and nonaffected family members.

Gene sequencing

A total of 28 genes and hypothetical proteins are found within the linkage interval between 48 and 53 cM on chromosome 11p13-12 according to the NCBI map viewer build 36.2 (*Homo sapiens*) (<http://www.ncbi.nlm.nih.gov>).

A number of sequence variants were identified by direct DNA sequencing in three genes sequenced at this region (CD59, APIP and CD44) when compared with reference sequences on NCBI and Ensembl databases. Six variants were found within the CD44 gene, all of which were already reported in public databases (rs11033019, rs1071695,

rs3736812, rs1103025, rs9666607 and rs1467558). Two of these variants (rs9666607 and rs1467558) are found in exons and result in amino acid changes but were not linked to the inherited haplotype at this region. On the other hand, a T to C transition within exon 3 (rs1071695) was found in all family members carrying the inherited haplotype. This variant does not result in an amino acid change. *Nsp I* was used to determine the frequency of this variant in a random sample from the general population ($n = 248$) and in a group of coeliac patients (all diagnosed by biopsy) ($n = 92$). The genotype frequencies observed were 66.5% CC, 31.5% CT and 2% TT: in the general population and 69.6% CC, 26.1% CT and 4.3% TT: in the group of coeliac patients. Both genotype frequencies were in Hardy-Weinberg equilibrium.

Eight variants were identified within the APIP gene (rs16926448, rs2016814, rs2016821, rs1326942, rs1326941, rs2303959, rs2303958 and rs1571133). The A to C transversion, found in exon 5 (rs1571133), does not result in an amino acid change. Another variant within this gene (rs2016814) was found in almost complete linkage with

the inherited haplotype, with the exception of individuals III: 4 and III: 8.

A novel variant (G/A) was identified within exon 3 of the protectin (CD59) gene. The G to A transition was observed to be in linkage with the inherited haplotype in our family (Table 2). This variant is the third nucleotide of the sixth codon for the signal peptide but does not result in a change of the amino acid, glycine (GGG > GGA). PCR-RFLP using restriction enzyme *Bst* *YI* was performed in a random sample from the population and a group of coeliac patients as above. This variant was completely absent in the group of unrelated coeliac patients and only one mutated allele was identified from 442 chromosomes analysed (0.2%), with a population frequency of 0.45%. This suggests that this variant is a rare mutation found only within this family. To assess the possibilities that this variant might have a deleterious effect, the on-line application SNPs3D (<http://www.snps3d.org>) (25) was used to analyse entropy. An entropy of 1.25 was obtained giving a 15% chance that this variant might cause disease against 7.5% to be nondeleterious.

Analysis of RNA splicing

An on-line bioinformatics tool that calculates free energy to predict RNA secondary structure was used (http://www.genebee.msu.su/services/rna2_reduced.html) (26). The results suggested that both variants within CD44 and CD59 might have an effect on RNA secondary structure. Using the on-line application FAS-ESS and RESCUE-ESE (<http://genes.mit.edu/fas-ess/>) (27), the two regions (CD44 and CD59) were tested for possible exonic splicing silencers (ESS) and enhancers (ESE). The region in CD44, where the linked variant was identified, did not show any ESS or ESE both in normal and in mutated alleles. A hexamer (FAS-Hex3) consensus sequence (GGAGGG) for an ESS was only observed in the wild-type allele of CD59 and was absent in the mutated allele, while two ESEs were observed (GGAGGA and GAGGAT) in the mutated allele. To test the hypothesis whether the variants found within both CD44 and CD59 exons might affect the splicing mechanism, a minigene system was used.

As shown in Figure 3a,b, amplified products of transcripts from pSPL3 minigene system were consistent between HeLa and COS-7 cell lines. For the variant within CD44 gene, only one transcript of 395 bp was observed in both normal and mutated alleles, showing that the C/T variant does not have any effect on exon skipping. As for the CD59 gene, two transcripts were observed from both normal and mutated constructs. The longer transcript consisted of 346 bp, while the smaller was 209 bp in size when compared with size marker and the control transcript from pSPL3 (261 bp). When sequencing both fragments, the 346 bp product was found to be made up of the full exon

3 of CD59 with flanking pSPL3 exons on both sides. Sequencing of the shorter product (209 bp) showed that exon 3 was spliced 6 bp upstream from the position of the identified variant, and only a small portion of this exon formed part of this transcript, while a small part (81 bp) of the pSPL3 exon was also missing. Sequencing was performed on both bigger and smaller fragments from wild-type and mutated constructs and the same results were obtained. Only a slight difference in band intensities was observed between the wild-type and mutated alternatively spliced fragments (209-bp fragments), where the latter had a slightly higher intensity. Using IMAGEJ 1.38x software, a slightly higher peak was observed for the smaller PCR fragment (indicated by an arrow) when compared with that from the bigger fragment of those transcripts amplified from mutated constructs (Figure 3d). Measuring peak height of both fragments, the ratio of the alternatively spliced fragment to normal transcript (higher peak) was of 0.43 against 0.26 from mutated and wild-type constructs, respectively, for HeLa cells. This difference was less (0.43 vs 0.36) in mutated and wild-type constructs from COS-7.

Discussion

In this study, suggestive linkage to chromosomes 9q21 and 11p13-12 was observed in a family with high incidence of CD. Two haplotypes on chromosomes 9 and 11 were observed to be inherited identical by descent in all affected members but not in most clinically nonaffected individuals. A rare variant within exon 3 of the CD59 gene was in linkage with the inherited haplotype.

Although no evidence of linkage was observed to the HLA locus on chromosome 6, HLA typing showed that most family members, including clinically normal individuals, carried the haplotype DRB1*07-DQA1*0201, DQB1*02 (DQ2.2), which does not increase the risk for CD (28). Risk for disease was high in the majority of affected family members in the presence of HLA-DQ2.5 (DQA1*05 and DQB1*02) in *trans*, although this was also found in clinically normal family members. This shows that although HLA alleles play a major role in disease, they were not the only causative factors of disease in this family. From this family, it is evident that susceptibility for disease was increased by the presence of both HLA-DQ2.5 and the A allele in CD59. This observation was supported by three normal individuals that were HLA-DQ2.5 in *trans* but that did not carry the CD59 A allele. On the other hand, individual II: 6 was affected even though she was typed as HLA-DQ2.2 in *cis*, which is considered to be of a moderate risk, but then she was also carrying the CD59 A allele. Conversely, individual III: 10 carried both the high-risk HLA-DQ2.5 and the CD59 variant, but still he was phenotypically normal. This suggests that other genetic and/or environmental factors, besides HLA and the CD59

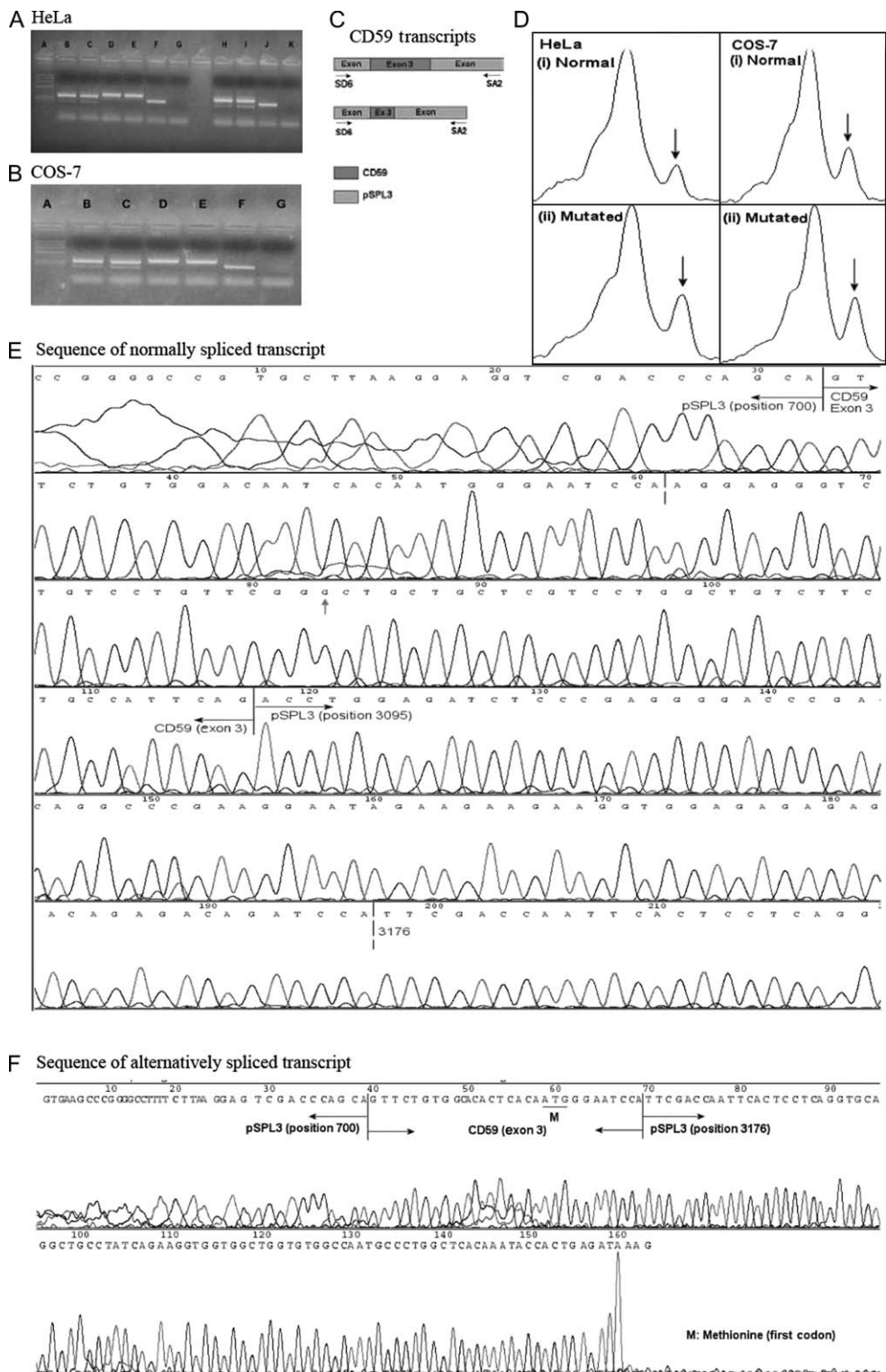


Figure 3 RNA splicing analysis; (a) and (b) agarose gel of reverse transcription PCR: column (A) phiX; (B and H) CD59 normal allele; (C and I) CD59 mutated allele; (D) CD44 normal allele; (E) CD44 mutated allele; (F and J) pSPL3-positive control and (G and K) negative control; (c) schematic diagram of amplified transcripts observed in agarose gel electrophoresis; (d) scanned images of gels using IMAGEJ 1.38x software for both normal and mutated constructs. Arrow indicates peak representing the shorter transcript; (e) electropherogram of normally spliced transcript from normal allele. Red arrow shows position of variant; (f) electropherogram of transcript from mutated allele, showing exon skipping.

variant, might have an effect on the penetrance of disease in this family.

A recent study performed in Bedouin kindred also reported absence of linkage to the HLA locus, where a higher frequency of high-risk HLA alleles was observed in affected members, thus negative linkage was probably because of a lack of distinction between alleles that were identical by descent from those that were identical by state (5). Besides this, there could be various reasons why we failed to detect linkage at the HLA locus including low marker density (closest marker D6S1051 being approximately 4 cM away), no difference in the inheritance pattern of HLA alleles between affected/nonaffected members and because in this family risk for disease was increased by HLA alleles being in *trans* and not in *cis*. This could be a weak point when performing linkage analysis in a single pedigree and where risk for disease is increased by contribution of different alleles found in *trans* rather than in *cis*.

The locus on chromosome 9q21 was not reported in previous linkage studies for CD. This region was only linked to inflammatory diseases such as ankylosing spondylitis and multiple sclerosis (29, 30). In our study, no linkage was found to other previously associated loci with CD as 11q, 4q27, 5q31 and the CTLA4-ICOS region on chromosome 2 (2, 12, 17, 31, 32).

The region on chromosome 11 confirmed reported results from linkage studies performed in 50 British families (10) and of an earlier study performed by Zhong *et al.* (33), but no genes were further studied at this region. On the other hand, linkage to this region was not replicated in other studies performed later (3, 8, 34). Linkage at 11p12 was observed with other inflammatory disorders such as inflammatory bowel disease, rheumatoid arthritis and osteoporosis (35–37).

A number of variants were identified within the CD44 and APIP genes, found within the linkage interval on chromosome 11p13–12. CD44 is the main hyaluronate cell surface receptor, which plays an important role in the selective migration of lymphocytes from the blood stream into the lymphatic system (38). Different isoforms are found in different tissues as a result of the extensive alternative splicing of its gene (39, 40). Expression of CD44 was observed to be increased in the intestinal mucosa of coeliac patients especially those with active disease (41). We failed to show any effect of one of the identified variants on pre-mRNA splicing, although this does not exclude the possible involvement of this gene in disease because this variant might be in linkage disequilibrium with another functional variant found in regions that were not sequenced. Similarly, none of the variants identified within the regions sequenced of the APIP gene were linked with the inherited haplotype. APIP suppresses apoptosis by binding to Apaf-1 and thus preventing further activation through caspase-9 (42). Because apoptosis is known to be responsible for the

flattening of the intestinal epithelium and depletion of peripheral blood lymphocytes in coeliac individuals (43, 44), possible mutations within the APIP gene that render it inactive might lead to an increased intestinal damage because of increased apoptosis.

The identification of the G/A variant within exon 3 of the CD59 gene is important for the identification of novel mechanisms that are involved in the pathogenesis of CD, in this case the complement system. Protectin (CD59) is a membrane glycoprotein widely expressed in a number of tissues, which protects cells from lysis after the formation of the C5b-9 membrane attack complex (MAC) of complement. Its gene is similar to that of the murine lymphocyte antigen Ly-6 protein (45), while a conserved segment of the protein between residues 42–58 of human CD59 is responsible for the species restricted inhibitory function of MAC (46). The role played by protectin in the human intestinal mucosa as well as in other tissues was reported by a number of investigators (47–49). Expression of CD59 together with CD55 was upregulated in epithelial cells but not in parietal cells of patients with chronic gastritis, CD and inflammatory bowel disease (48). These two molecules were also highly expressed in normal colonic epithelium, probably to protect it from the constant complement activation caused by the presence of the normal flora in this region. Conversely, CD59 expression was reduced in affected epithelium of patients with Crohn's disease and ulcerative colitis but was normal in nonaffected areas, thus increasing the vulnerability of the affected areas to complement lysis (47). These observations show the importance of CD59 in protecting the intestinal mucosa from complement and suggest that mutations within its gene might lead to an increased intestinal damage in patients with CD. A higher degree of intestinal damage increases the chance for an individual carrying a high-risk HLA group to be phenotypically affected and to seek medical attention, following activation of the immune response on exposure and antigen presentation of gluten.

Because the variant identified in this study did not result in an amino acid change, we hypothesised that it could affect the splicing mechanism by either altering the secondary structure of pre-mRNA or else by affecting binding of *trans*-acting splicing factors (27). The loss of an ESS and introduction of two possible ESEs, as described above, also indicate that splicing might be affected by this novel variant in CD59. Aberrant splicing was reported to be caused by both missense or silent mutations and by variants found within introns of a number of genes responsible for other human diseases including cystic fibrosis, neurofibromatosis type I and frontotemporal dementia (50–52). A common problem encountered when studying RNA is to obtain samples from the affected tissue itself to avoid obtaining misleading results because of tissue-specific gene expression. Besides affecting RNA splicing, synonymous variants might

affect timing of cotranslational folding when a frequently used codon is replaced by a rare codon possibly because of the abundance of transfer RNA present in cells that correlates to codon usage in a tissue-specific manner (53, 54).

In conclusion, our results suggest the involvement of other inflammatory genes present on chromosome 11p13-12 in the pathogenesis of coeliac in a family with high incidence of CD. These findings have the potential to add to our understanding of the complex physiology of CD and can lead to further identification of potential gene candidates and innovative individualised therapeutic approaches.

Acknowledgments

We thank the family members that participated in this study for their collaboration. Also we thank Dr Pierre Lapage and the staff at the McGill University and Genome Quebec Innovation Centre, Quebec, Canada, for the short tandem repeat (STR) genotyping. We are grateful to Dr Irene Bottillo, CSS-Mendel Institute, Rome, Italy, that generously donated the pSPL3 plasmid. We also thank Dr Anthony Fenech for the use of facilities at the Department of Clinical Pharmacology and Therapeutics, University of Malta, Malta. This work was funded by the Malta Council of Science and Technology Research and Development Programme 2004. The Web resources used were Online Mendelian Inheritance in Man (OMIM) (<http://www.ncbi.nlm.nih.gov/Omim/>) and Ensembl database (<http://www.ensembl.org>). The software CYRILLIC v2.1.3 (Cherwell Scientific Publishing Ltd, Oxford, UK) was used to design pedigree in Figures 1 and 2.

References

- Kagnoff MF. Celiac disease: pathogenesis of a model immunogenetic disease. *J Clin Invest* 2007; **117**: 41–49.
- Nalwai AT, Nilsson S, Gudjonsdottir AH *et al.* Genome-wide linkage analysis of Scandinavian affected sib-pairs support presence of susceptibility loci for coeliac disease on chromosomes 5 and 11. *Eur J Hum Genet* 2001; **9**: 938–44.
- Neuhausen SL, Feolo M, Farnham J, Book L, Zone JJ. Linkage analysis of HLA and candidate genes for celiac disease in a North American family-based study. *BMC Genet* 2001; **2**: 12.
- Garner CP, Ding YC, Steele L *et al.* Genome-wide linkage analysis of 160 North American families with celiac disease. *Genes Immun* 2007; **8**: 108–14.
- Eller E, Vardi P, Babu SR *et al.* Celiac disease and HLA in a Bedouin kindred. *Hum Immunol* 2006; **67**: 940–50.
- Percopo S, Babron MC, Whalen M *et al.* Saturation of 5q31-q33 candidate region for coeliac disease. *Ann Hum Genet* 2003; **67**: 265–68.
- King AL, Yiannakou JY, Brett PM *et al.* A genome-wide family-based linkage study of coeliac disease. *Ann Hum Genet* 2000; **64**: 479–90.
- Houlston RS, Tomlinson IPM, Ford D *et al.* Linkage analysis of candidate regions for coeliac disease genes. *Hum Mol Genet* 1997; **6**: 1335–39.
- Van Belzen MJ, Meijer JWR, Sandkuijl LA *et al.* A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterol* 2003; **125**: 1216–17.
- King AL, Fraser JS, Moodie SJ *et al.* Coeliac disease: follow-up linkage study provides further support for existence of a susceptibility locus on chromosome 11p11. *Ann Hum Genet* 2001; **65**: 377–86.
- Nunez C, Rueda B, Martinez A *et al.* A functional variant in the CD209 promoter is associated with DQ2-negative coeliac disease in the Spanish population. *World J Gastroenterol* 2006; **12**: 4397–400.
- Hunt KA, McGovern DPB, Kumar PJ *et al.* A common CTLA4 haplotype associated with coeliac disease. *Eur J Hum Genet* 2005; **13**: 440–44.
- Borg J, Scerri CA, Vidal C, Xuereb-Anastasi A. There is no association between -318 (C/T) and +49 (A/G) CTLA4 gene polymorphisms and coeliac condition in the Maltese population. *Balk J Med Genet* 2007; **9**: 49–51.
- Brophy K, Ryan AW, Thornton JM *et al.* Haplotypes within the CTLA4 region are associated with coeliac disease in the Irish population. *Genes Immun* 2006; **7**: 19–26.
- Monsuur AJ, de Bakker PI, Alizadeh BZ *et al.* Myosin IXB variant increases the risk of celiac disease and points towards a primary intestinal barrier defect. *Nat Genet* 2005; **37**: 1341–42.
- Boniotto M, Braida L, Ventura A, Percopo S, Amoroso A, Crovella S. Promoter polymorphisms of the CD14 gene in Italian patients with coeliac disease. *J Med Genet* 2003; **40**: 108.
- Van Heel DA, Franke L, Hunt KA *et al.* A genome-wide association study for coeliac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007; **39**: 827–29.
- Adamovic S, Amundsen SS, Lie BA *et al.* Association study IL2/IL21 and FcγRIIa: significant association with the IL2/IL21 region in Scandinavian coeliac families. *Genes Immun* 2008; **9**: 364–67.
- Hunt KA, Zhernakova A, Turner G *et al.* Newly identified genetic risk variants for coeliac disease related to the immune response. *Nat Genet* 2008; **40**: 395–402.
- Diosdado B, Wapenaar MC, Franke L *et al.* A microarray screen for novel candidate genes in coeliac disease pathogenesis. *Gut* 2004; **53**: 944–51.
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1251.
- Fishelson M, Dovgolevsky N, Geirger G. Maximum likelihood haplotyping for general pedigrees. *Hum Hered* 2005; **59**: 41–60.
- Lindner TM, Hoffman K. EasyLinkage: a PERL script for easy and automated two/multi-point linkage analyses. *Bioinformatics* 2005; **21**: 205–7.
- Kong A, Cox NJ. Allele sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997; **61**: 1179–88.

25. Yue P, Melamud E, Moulton J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics* 2006; **22**: 166.
26. Brodsky LI, Ivanov VV, Kalai dzidis YL et al. GeneBee-NET: Internet-based server for analyzing biopolymers structure. *Biochemistry* 1995; **60**: 923–8.
27. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. *Cell* 2004; **119**: 831–45.
28. Vader W, Stepniak D, Kooy Y et al. The HLA-DQ2 gene dose effect in coeliac disease is directly related to the magnitude and breadth of gluten-specific T-cell responses. *Proc Natl Acad Sci U S A* 2003; **100**: 12390–95.
29. Lee YH, Rho YH, Choi SJ, Ji JD, Song GG. Ankylosing spondylitis susceptibility loci defined by genome-search meta-analysis. *J Hum Genet* 2005; **50**: 453–9.
30. Dyment DA, Sadovnick AD, Willer CJ et al. An extended genome scan in 442 Canadian multiple sclerosis-affected sibships: a report from the Canadian Collaborative Study Group. *Hum Mol Genet* 2004; **13**: 1005–15.
31. Rioux JD, Karinen H, Kocher K et al. Genomewide search and association studies in a Finnish celiac disease population: identification of a novel locus and replication of the HLA and CTLA4 loci. *Am J Med Genet* 2004; **130**: 345–50.
32. Babron MC, Nilsson S, Adamovic S et al. Meta and pooled analysis of European coeliac disease data. *Eur J Hum Genet* 2003; **11**: 828–34.
33. Zhong F, McCombs CC, Olson JM et al. An autosomal screen for genes that predispose to coeliac disease in western countries of Ireland. *Nat Genet* 1996; **14**: 329–33.
34. Greco L, Corazza G, Babron MC et al. Genome search in coeliac disease. *Am J Hum Genet* 1998; **62**: 669–75.
35. Paavola-Sakki P, Ollikainen V, Helio T et al. Genome-wide search in Finnish families with inflammatory bowel disease provides evidence for novel susceptibility loci. *Eur J Hum Genet* 2003; **11**: 112–20.
36. Amos CI, Chen WV, Lee A et al. High-density SNP analysis of 642 Caucasian families with rheumatoid arthritis identifies two new linkage regions on 11p12 and 2q33. *Gene Immun* 2006; **7**: 277–86.
37. Vidal C, Galea R, Brincat M, Xuereb-Anastasi A. Linkage to chromosome 11p12 in two Maltese families with a highly penetrant form of osteoporosis. *Eur J Hum Genet* 2007; **15**: 800–9.
38. Aruffo A, Stamenkovic I, Melnick M, Underhill CB, Seed B. CD44 is the principal cell surface receptor for hyaluronate. *Cell* 1990; **61**: 1303–13.
39. Screaton GR, Bell MV, Jackson DG, Cornelis FB, Gerth U, Bell JI. Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. *Proc Natl Acad Sci U S A* 1992; **89**: 12160–64.
40. Vela E, Roca X, Isamat M. Identification of novel splice variants of the human CD44 gene. *Biochem Biophys Res Commun* 2006; **343**: 167–70.
41. Kempainen T, Tammi R, Tammi M et al. Elevated expression of hyaluronan and its CD44 receptor in the duodenal mucosa of coeliac patients. *Histopathology* 2005; **46**: 64–72.
42. Cho DH, Hong YM, Lee HJ et al. Induced inhibition of ischemic/hypoxic injury by APIP, a novel Apaf-1-interacting protein. *J Biol Chem* 2004; **17**: 39942–50.
43. Moss SF, Attia L, Scholes JV, Walters JRF, Holt PR. Increased small intestinal apoptosis in coeliac disease. *Gut* 1996; **39**: 811–17.
44. Di Sabatino A, D'Alo S, Millimaggi D et al. Apoptosis and peripheral blood lymphocyte depletion in coeliac disease. *Immunology* 2001; **103**: 435–44.
45. Petranks JG, Fleenor DE, Sykes K, Kaufman RE, Rosse WF. Structure of the CD59-encoding gene: further evidence of a relationship to murine lymphocyte antigen Ly-6 protein. *Proc Natl Acad Sci U S A* 1992; **89**: 7876–9.
46. Zhao XJ, Zhao J, Zhao Q, Sims PJ. Identity of the residues responsible for the species restricted complement inhibitory function of human CD59. *J Biol Chem* 1998; **273**: 10665–71.
47. Scheinin T, Böhling T, Halme L, Kontiainen S, Björge L, Meri S. Decreased expression of protectin (CD59) in gut epithelium in ulcerative colitis and Crohn's disease. *Hum Pathol* 1999; **30**: 1427–30.
48. Berstad AE, Brandtzaeg P. Expression of cell membrane complement regulatory glycoproteins along the normal and diseased human gastrointestinal tract. *Gut* 1998; **42**: 522–9.
49. Niehans GA, Cherwitz DL, Staley NA, Knapp DJ, Dalmaso AP. Human carcinomas variably express the complement inhibitory proteins CD46 (membrane cofactor protein), CD55 (decay-accelerating factor), and CD59 (protectin). *Am J Pathol* 1996; **149**: 129–42.
50. Hefferon TW, Groman JD, Yurk CE, Cutting GR. A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing. *Proc Natl Acad Sci U S A* 2004; **101**: 3504–9.
51. Bottillo I, DeLuca A, Schirinzi A et al. Functional analysis of splicing mutations in exon 7 of NF1 gene. *BMC Med Genet* 2007; **8**: 4.
52. D'Souza I, Poorkaj P, Hong M et al. Missense and silent tau gene mutations cause frontotemporal dementia with Parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. *Proc Natl Acad Sci U S A* 1999; **76**: 5598–603.
53. Kimchi-Sarfaty C, Oh JM, Kim IW et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 2007; **315**: 525–28.
54. Dittmar KA, Goodenbour JM, Pan T. Tissue-specific differences in human transfer RNA expression. *PLOS Genet* 2006; **2**: 2107–15.