



ARTICLE

# Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11

Åsa Torinsson Naluai<sup>\*1</sup>, Staffan Nilsson<sup>3</sup>, Audur H Gudjónsdóttir<sup>2</sup>, Andrew S Louka<sup>4</sup>, Henry Ascher<sup>2</sup>, Johan Ek<sup>5</sup>, Birgitta Hallberg<sup>1</sup>, Lena Samuelsson<sup>1</sup>, Bengt Kristiansson<sup>2</sup>, Tommy Martinsson<sup>1</sup>, Olle Nerman<sup>3</sup>, Ludvig M Sollid<sup>4</sup> and Jan Wahlström<sup>1</sup>

<sup>1</sup>Department of Clinical Genetics, Göteborg University, Sweden; <sup>2</sup>Department of Paediatrics, The Queen Silvia Children's Hospital, Sahlgrenska University Hospital/Östra, Sweden; <sup>3</sup>Chalmers University of Technology, Göteborg, Sweden; <sup>4</sup>Institute of Immunology, Rikshospitalet, University of Oslo, Norway; <sup>5</sup>Department of Paediatrics, Buskerud Central Hospital, Drammen, Norway

Celiac disease (CD) is a common chronic inflammatory disorder of the small intestine with a multifactorial aetiology. HLA is a well-known risk factor, but other genetic factors also influence disease susceptibility. To identify the genes involved in this disorder, we performed a genome-wide scan on 106 well-defined Swedish and Norwegian families with at least two affected siblings. We investigated familial segregation of 398 microsatellite markers, and utilised non-parametric linkage analysis. The strongest linkage with disease was found to the HLA locus (6p) ( $P < 0.000006$ ). There were eight regions besides HLA with a point wise  $P$  value below 0.05. Among these eight regions were 11q and 5q, both of which have been suggested in several linkage studies of independent celiac disease families. We also performed a stratification analysis of families according to their HLA genotypes. This resulted in significant differences on chromosome 2q. These results indicate that 11q, 5q and possibly also 2q are true susceptibility regions in CD.

*European Journal of Human Genetics* (2001) 9, 938–944.

**Keywords:** autoimmune; celiac disease; complex diseases; genome-wide scan; linkage analysis; affected sib-pairs

## Introduction

Celiac disease (CD [MIM 212750]) or gluten sensitive enteropathy is a chronic inflammatory disorder of the small intestine with multifactorial etiology.<sup>1</sup> Gluten is a critical environmental factor, as removal of gluten from the diet leads to ablation of inflammation and disease remission. Although gluten is a necessary factor for disease development, it is not sufficient. Several lines of evidence suggest that genetic factors are important. The risk of a sibling to a CD patient to develop disease is about 20 to 60 times higher

than that of the normal population ( $\lambda_s=20-60$ ).<sup>2,3</sup> Moreover, the twin concordance in CD is high. A recent large population based study in Italy using 59 twin couples, claims a proband-wise concordance of 85–95% for monozygotic twins and about 25% for dizygotic twins (Romino *et al.*, personal communication).

The HLA-region on chromosome 6 is one of the genetic factors. More than 90% of CD patients carry both the HLA DQA1\*05 and the DQB1\*02 alleles, which code for a variant of the HLA-DQ2 molecule.<sup>4</sup> The majority of patients who are DQ2 negative, carry the DR4-DQ8 haplotype. Although the evidence is less clear than for DQ2, it is most likely the DQ8 molecule encoded by DQA1\*0301/DQB1\*0302, which confers the primary predisposition in this subgroup of patients.<sup>1</sup> Accumulating data suggests that the DQ2 and DQ8 molecules confer susceptibility to the disease by presenting gluten epitopes to CD4+ T cells in the small intestinal

\*Correspondence: Åsa T Naluai, Department of Clinical Genetics, Sahlgrenska University Hospital/Östra, 41685 Göteborg, Sweden.

Tel: +46 31 343 5464; Fax: +46 31 84 21 60;

E-mail: asa.torinsson.naluai@clingen.gu.se

Received 15 August 2001; revised 16 October 2001; accepted 30 October 2001

mucosa.<sup>1</sup> Non-HLA genes also contribute to the genetic susceptibility. The combination of DQA1\*05 and DQB1\*02 occur in 20–30% of the healthy Caucasian population, which means that less than 1% of carriers develop CD. Furthermore, the concordance rate of monozygotic twins is significantly higher than that of HLA-identical dizygotic twins (30%), (Romino *et al.*, personal communication).<sup>5,6</sup> Assuming a multiplicative model of disease predisposing genes, the overall importance of non-HLA genes has been calculated to be greater than that of the HLA genes.<sup>3,7</sup> Despite this, knowledge about non-HLA genes that predispose to CD is scarce.

Four genome-wide scans have been reported on CD. Two studied affected sib-pairs<sup>8</sup> (Rioux *et al.*, personal communication), one used extended families,<sup>9</sup> and in another a combination of extended families/sib-pairs was examined.<sup>10</sup> Additionally, there are a few studies where candidate regions of the genome-wide scans have been followed up by analysis of sib-pairs or extended families.<sup>11–13</sup>

We have performed a genome-wide scan in a population where all families have a Scandinavian background. Several of the regions for which we find the strongest linkage have been implicated in earlier linkage studies with affected sib pairs, whereas the results from studies with extended pedigrees is less consensual. This may indicate that there exist multiple susceptibility loci in CD. By studying a larger number of independent families, it may be easier to identify genetic factors which are shared by the majority of CD patients. Studies of a few extended pedigrees may, to a larger extent, uncover rare and almost monogenic variants.

## Materials and methods

### Subjects

Genetic material from families with at least two affected children were collected. In Sweden this was done through an advertisement in the journal of the Swedish Celiac Society, by inquiries to all Swedish paediatric departments and through our own patient registries. The Norwegian families were collected via local registries. There were 12 families with two affected sib-pairs from Norway and the remaining 92 families were from Sweden. All families in which the affected siblings did not strictly fulfil the ESPGHAN diagnostic criteria<sup>14</sup> were excluded. The study included one extended pedigree with three nuclear families and 103 separate families with a minimum of two affected siblings. Family structures are presented in Table 1. The median age at diagnosis of siblings with CD was 2.2 years (6 months–58 years) and 88% were under the age of 16 at diagnosis. The median age at diagnosis of affected parents was 39 years (range 19–78 years). Genomic DNA was extracted from blood using standard procedures.<sup>15</sup> This study was approved by the Ethical Committee of The Medical Faculty of Göteborg University. Informed consent was received from all families involved.

### Analysis strategy

A two-step strategy was used for the analysis. First we selected 70 families which we considered to be the most informative: those with a large number of children and affected individuals (Group A). The remaining 36 families (Group B) were added to group A in a second step and were genotyped over selected regions based on either of three different criteria. First, chromosomes implicated by the previously published CD genome-wide screens: chromosomes 5, 11 and 15.<sup>8,10,12,24</sup> Second, novel chromosomes identified from Group A that showed NPL-values of 2.0 or more and third, chromosome 2 and 20 because of the location of two candidate genes, suggested to be involved in CD: cytotoxic T-lymphocyte associated protein (CTLA4) and tissue transglutaminase (TG2). (Figure 1). Altogether, 137 markers across nine chromosomes: 2, 5, 6, 9, 11, 14, 17, 20 and X were analysed for all 106 nuclear families (Group A and B). The remaining 261 markers on the other chromosomes were analysed only for the initial 70 families (Group A).

### Microsatellite genotyping and linkage analysis

We used the Weber screening set version six from Research Genetics containing 390 microsatellite markers with an average distance between markers of 10 cM (<ftp://ftp.resgen.com/pub/mappairs/humanset>). Additional markers were used in two regions: in the HLA class II region on chromosome 6 and in the CTLA4/CD28 region on chromosome 2q.<sup>16,17</sup> We amplified the microsatellite marker regions using an ABI 877 PCR robot under standard conditions. PCR products were separated by electrophoresis on an ABI 377XL sequencer and a 5% denaturing polyacrylamide gel. Genotyping was performed using GENESCAN ANALYSIS 2.1 and GENOTYPER 2.0 software (<http://www.appliedbiosystems.com>). Genotypes were checked for errors in the expected pattern of Mendelian inheritance using the PEDCHECK program.<sup>18</sup>

A non-parametric linkage (NPL) analysis was performed. The linkage analysis was done with the statistic  $NPL_{all}$  in multipoint mode using the software GENEHUNTER version 2.0 for autosomals and version 1.3 for the X chromosome.<sup>19</sup>

$NPL_{all}$  is based on the allele sharing among all affected family members simultaneously.

### HLA typing and stratification by HLA-DQ genotypes

The HLA-DQA1 and DQB1 genotyping was done by PCR and sequence specific oligo nucleotide probing (PCR-SSOP) as previously described.<sup>20</sup> All samples that could not be unambiguously determined were additionally typed using a PCR-sequence specific primer (PCR-SSP) kit (Dynal, Norway).

Two HLA-DQ genotypes confer substantially higher risk to CD than the other HLA-DQ risk genotypes and genotypes lacking the risk alleles completely.<sup>21</sup> Carrying DQA1\*05-DQB1\*02 on one chromosome, and either the same haplotype (i.e. homozygous) or DQA1\*02-DQB1\*02 on the second chromosome, bestows additional risk. It is conceivable that individuals without these genotypes may require

risk factors outside HLA with a stronger effect, to accumulate sufficient risk to develop CD. The families were therefore divided into a group of 36 families where at least two siblings carried the high risk genotypes (high risk group), and another group of 70 families where at least one of the siblings carried a lower risk genotype (low risk group).

Linkage analysis was performed on each group, for the eight highest peak markers of the genome-wide screen (2q11–13, 3p24, 5q31–33, 9p21, 11p15, 11q23–25, 17q22 and Xp11; see below) and microsatellite marker D2S1391 which gave the highest linkage score in the region of CTLA4, a previously identified risk factor.<sup>16,22</sup> Significance testing for each marker was done by a one-sided two-sample *t*-test. The test was performed on each family and its individual NPL-value. Correction for multiple comparison was done by multiplying the significance values by a factor of nine (nine

markers tested). We also tested the low and high HLA-risk groups for association to the CTLA4 A/G polymorphism in exon 1, codon 17, using the transmission disequilibrium test (TDT) developed by Spielman and Ewens.

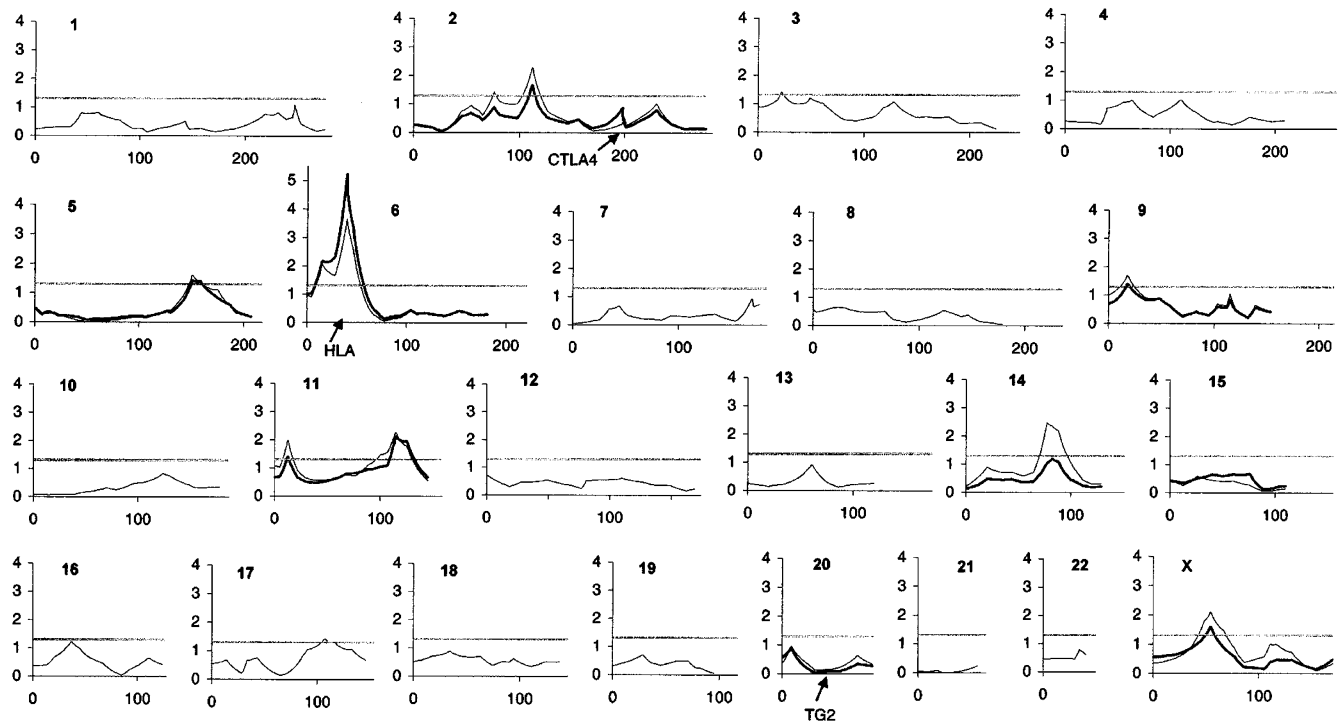
## Results

We have estimated the power to detect at least one non-HLA gene by assuming that the non-HLA component consists of two genes in a heterogeneity model and assuming the HLA and non-HLA component to be multiplicative. By varying the frequencies of the non-HLA genes, the power varied between 55 and 94%<sup>23</sup> for suggestive linkage.<sup>26</sup>

Four new regions on chromosomes 2q, 14q, Xp and 9p were detected in family Group A, (70 families) which had NPL scores above 2.0. These regions were subsequently genotyped in all families (Group A and B). In addition all families were genotyped for chromosomes 5, 11 and 15 detected in previous genome-wide scans,<sup>8,10,12,24</sup> the HLA region on chromosome 6, the region 2q33 containing CTLA-4 and chromosome 20 containing the gene for tissue transglutaminase. Mean spacing of the microsatellite markers was 10 cM except for within the HLA and CTLA4 regions where several markers were typed over a 2–5 cM region. In Figure 1, a thick line indicates regions where all families were genotyped (Group A+B).

**Table 1** Family structures

Number of families	Affected parents	Affected children
76	0	2
6	0	3
16	1	2
5	1	3
1	3	7



**Figure 1** Non Parametric Linkage results (NPL) displayed as the negative tenth logarithm of the *P* value (Y-axis). The X-axis indicates chromosome distance in cM. A thin line indicates results for 70 families (Group A). A thick line indicates the result from all 106 nuclear families (Group A+B). The grey line across chromosomes indicates a nominal *P* value of 0.05.

HLA was the only region reaching Lander and Kruglyaks proposed level for genome-wide significance (nominal  $P$  value of  $2 \times 10^{-5}$ ).<sup>26</sup>

Chromosomal regions 2q11–13, 3p24, 5q31–33, 9p21, 11p15, 11q23–25, 17q22 and Xp11 all had nominal  $P$  values slightly below 0.05 (Table 2), while the peak on 14q dropped markedly after genotyping for all families (Figure 1).

The candidate regions around CTLA4 and tissue transglutaminase did not reach any significance.

The information content from the GENEHUNTER calculations for all chromosomes are posted on our home page: <http://www.clingen.gu.se/cel/index.htm>

As a measure of the familiar contribution of each locus,  $\lambda_{s\text{MARKER}}$  (Table 2) was calculated as  $\lambda_{s\text{MARKER}}=0.25/p_0$ .<sup>7</sup> The probability of no allele sharing in a sib pair,  $p_0$ , was estimated for the markers using GENEHUNTER.

HLA stratification did not lead to a significant difference in linkage between families with low versus high HLA risk, except for the CTLA4 locus (2q33) which gave a nominal  $P$  value below 0.001 (corrected  $P$  value of 0.009) (Figure 2). However, the CTLA4 A/G polymorphism which has shown association to CD in these families<sup>16</sup> did not show stronger association in the group with low risk at HLA. The one-sided TDT results in the high risk group, 27 transmissions of the G-allele and 50 transmissions of the A-allele ( $P=0.006$ ) and in the low risk group 47 transmissions of the G-allele versus 59 transmissions of the A-allele ( $P=0.14$ ) showed the high risk group to be more strongly associated.

## Discussion

CD is unique among the multifactorial, chronic inflammatory diseases with complex genetics for at least two reasons. First, a critical environmental factor triggering the disease (gluten) has been identified. Second, the HLA association in CD is unusually clear.<sup>1</sup> The DQ2 and DQ8 molecules, which confer the primary susceptibility, present gluten peptide antigens to CD4+ T cell in the intestinal lesion. There is evidence that these T cells control the immune response to gluten. Hence, the activation and effector function of these T cells will be a critical checkpoint for development of the

disease. Since a large number of CD patients also suffer from other chronic inflammatory disorders<sup>24</sup> for example type 1 diabetes, it is likely that these diseases have susceptibility genes in common. Chronic inflammatory diseases are generally characterized by great complexity. Compared with other chronic inflammatory diseases, the understanding of the basic molecular mechanisms of CD is good. It should therefore be easier to attain a complete understanding of how genes and environment interact to produce disease for this disorder. A significant step towards this goal would be the mapping of predisposing genes.

We chose to perform a genome-wide screen by sib-pair analysis to map genetic factors in CD. Since genes predisposing to CD most likely are variants common in the general population, we believe a good approach is to use many independent families with affected sib-pairs. The sib-pair method could prove to be more powerful in this situation than a few large multiply affected pedigrees.<sup>25</sup>

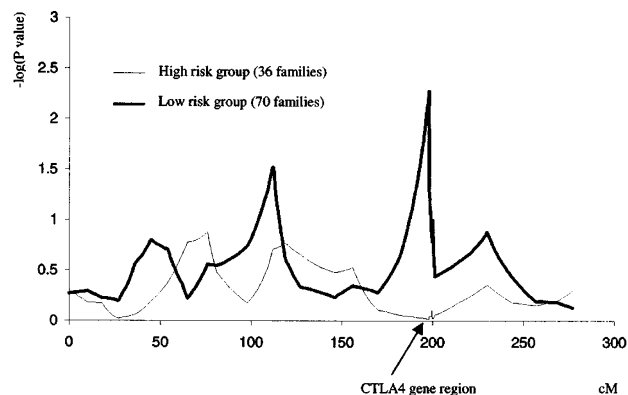
In our genome-wide scan 11q was the only region close to the level of suggestive linkage ( $P=0.0007$ ) and HLA the only region reaching significant linkage ( $P=0.00002$ ).<sup>26</sup> This is not surprising when dealing with complex genetic disorders where significant linkage as proposed by Lander and Kruglyak<sup>26</sup> is difficult to obtain. When setting the threshold lower, i.e.  $P=0.05$ , one must consider the fact that this  $P$  value will often be a chance result in a genome scan, but this threshold can be useful when comparing results with other linkage studies.

Four celiac disease genome-wide scans have been published: one Italian in sib-pairs,<sup>8</sup> one Irish in extended pedigrees and sib-pairs,<sup>10</sup> one with extended pedigrees from the UK<sup>9</sup> and one in Finnish families, mainly with sib-pairs (Rioux *et al.*, personal communication). Some regions were indicated in several studies including follow-up studies, as outlined in Table 3. There appears to be a trend of replication between the different sib-pair family studies (5q and 11q) and a trend of replication in extended pedigrees. Chromosomes 5q and 11q have been implicated in almost all independent sib-pair studies reported so far. Also a few regions have been replicated in two independent extended pedigree studies each.<sup>10,12</sup> (Table 3)

**Table 2** Regions in the genome-wide scan with a point-wise  $P$  value below 0.05

Marker at max peak	Region with $P < 0.05$ (cM from P-telomere)		NPL-value	P value	$\lambda_{s\text{ MARKER}}^*$
HLA	6p	(8–55)	4.40	0.000006	5.0
D11S4464	11q23–25	(110–130)	2.42	0.0078	1.75
D2S436	2q11–13	(109–115)	2.03	0.021	1.25
DX6810	Xp11	(50–57)	1.94	0.026	–
D5S816	5q31–33	(151–159)	1.78	0.038	1.61
D9S925	9p21	(23)	1.78	0.038	1.37
GATA31B11	17q22	(107)	1.77	0.038	1.96
D3S2403	3p24	(30)	1.76	0.040	1.31
D11S1999	11p15	(12.5)	1.75	0.040	1.34

\*Calculated according to Risch.<sup>7</sup>



**Figure 2** Non Parametric Linkage results (NPL) on chromosome 2 for the HLA stratified groups. The Y-axis shows the negative tenth logarithm of the *P* value. The X-axis indicates chromosome distance in cM. Sibling pairs were both are homozygous for high risk alleles in HLA (the thin line), and sibling pairs where one or both siblings are heterozygous for high risk alleles at HLA (the thick line).

Chromosomes 11q and 5q were the major regions of interest in the genome-wide scan of Italian sib-pairs.<sup>8</sup> In the genome-wide screen of Finnish families, 5q contained one of the major peaks (Rioux *et al.*, personal communication). The region 5q also had a point-wise *P* value below 0.05 in a Finnish follow up study<sup>11</sup> and in the genome-wide scan of Irish families.<sup>10</sup> Thus, five (including ours), of six linkage studies in CD implicate the 5q region in disease susceptibility.

In addition to the Italian study, chromosome 11q showed a point-wise *P* value below 0.05 in a Finnish follow-up study.<sup>11</sup> Since 11q gave the highest peak after HLA in our study, there are now three independent replications of the 11q region. Our study clearly implicates chromosomal regions 11q and 5q, two regions which hold a strong position as candidate gene regions for CD.

For studies using an extended family approach, several regions have now been implicated in two independent studies: 6p12-q14, 11p11, 19q13 and 15q. Chromosome 6, 11 and 19 were first implicated by Zhong *et al.*<sup>10</sup> in their genome screen, and more recently in the British genome-wide screen.<sup>9</sup> A follow-up study of extended pedigrees also from the UK<sup>12</sup> replicated (point-wise *P* value of <0.05) the genome-wide study findings by Zhong *et al.*<sup>10</sup> in the region of 15q.

The fact that our results coincide better with the linkage studies that have employed a sib-pair approach and not with those that employed an extended family approach is interesting. One could speculate that there is a tendency towards replication in similar types of families. Results from sib-pair studies are replicated in other sib-pair studies and results from extended pedigree material are replicated in other extended pedigrees (to a lesser degree). This may reflect the differences in disease causing alleles and genes in complex genetic diseases. Common genetic variants adding

to the susceptibility are not likely to be present in all families. This gives very different results depending on the genetic makeup of those particular families in the study. A larger number of independent families are more likely to have the common variant represented in the majority of cases, while a few very large pedigrees by chance could have different genetic backgrounds and susceptibility genes. Multiply affected pedigrees could also, to a larger extent, have rare disease alleles with a high penetrance or even almost monogenic variants (as the MODY genes in diabetes and BRCA1 and 2 in breast cancer), since these types of genes are more likely to be over represented in pedigrees with a large number of affected individuals. This could explain the weaker trend of replication seen in studies on extended pedigrees.

So far the only candidate gene reported to be genetically associated with CD is the cytotoxic T-lymphocyte associated gene (CTLA4).<sup>16,17,22</sup> We therefore included CTLA4 in our HLA stratification analysis. There is a good understanding of the predisposing genes in HLA, and we divided families into a low risk and a high risk group according to their HLA genotypes. This division led to a significantly higher linkage score at the CTLA4 locus in the low risk group, compared to the high risk group (*P*=0.001). Disappointingly, when we applied a similar stratification to an association analysis of the A/G exon 1 polymorphism in CTLA4, we found a stronger association in the group of the high risk HLA genotypes. The effect of HLA on linkage with CTLA4 is hence not mirrored in the association analysis.

There are about twice as many affected females as there are males in CD.<sup>27</sup> Excess affection of females is often seen in chronic inflammatory diseases. This gender skewing may be caused by sex effects on susceptibility loci located on autosomes, but it may also be due to the location of a predisposing gene on the X chromosome which makes the finding on chromosome X particularly interesting. The X chromosome harbors immunologically important gene clusters.<sup>28</sup> Several human X-linked immune deficiencies have been identified. Notably, Xp11 is a region where linkage has been found in type 1 diabetes,<sup>29,30</sup> rheumatoid arthritis<sup>31</sup> and multiple sclerosis.<sup>32,33</sup>

The region on chromosome 5 linked with CD also demonstrates linkage to type 1 diabetes (IDDM18). Recently IL12B was identified as a possible candidate gene for this effect.<sup>34</sup> A single base change in the 3' untranslated region showed strong linkage disequilibrium with the type 1 diabetes susceptibility locus, and cell lines with distinct alleles of the IL12B 3' untranslated region polymorphism showed different levels of expression of the pro-inflammatory cytokine, interleukin-12. Interestingly, this polymorphism is not associated with CD in our material<sup>35</sup> and hence there is probably a polymorphism of another gene nearby, which is relevant for CD.

In conclusion, this full genome-wide scan of CD in a Scandinavian population reveals, beside the HLA region,

**Table 3** Comparison of regions reported by linkage studies in celiac disease. Regions with a point-wise *P* value below 0.05 in different studies. Mainly, regions 5q and 11q show reproducibility in the studies of sib-pairs and 11p, 6p, 19q and 15q in studies of extended pedigrees

Type of families/study:	Chromosome regions:					
	5q	11q	11p11	6p12-q14	19q13	15q
Sib-pair (Greco, 1998) <sup>8</sup>	X	X				
sib-pair (Rioux <i>et al.</i> 1998)	X					
sib-pair (Holopainen 2001) <sup>11</sup>	X	X				
Sib-pair* (Naluai <i>et al.</i> )	X	X				
Extended and sib-pairs (Zhong, 1996) <sup>10</sup>	X		X	X	X	X
Extended King, 2000) <sup>9</sup>			X	X	X	
Extended (Houlston, 1997) <sup>12</sup>						X

\*Our study.

eight regions with a nominal *P* value below 0.05. Several of the regions are reported in earlier linkage studies of CD. Particularly obvious are the regions on chromosome 5 and 11. This leads us to believe that these regions are true disease loci for CD common across different populations.

#### Acknowledgments

We would like to thank Britt-Marie Käck and The Celiac Society in Sweden for help in collecting families and blood samples, Bente Talseth for HLA typing, and all the families participating in the study. The work was supported by grants from the Foundation for Strategic Research, the Swedish Medical Research Council (Project no. K98-13X-12568-01A and K2000-27X-12568-03A), the Gothenburg Medical Society, the European Commission (QLRT-1999-00037), EXTRA funds from the Norwegian Foundation for Health and Rehabilitation, Aktieselskabet Freia Chocolate Fabriks Medicinske Fond, The Swedish Society of Medicine, Frimurare-Barnhusdirektionen, The Foundation of the National Board of Health and Welfare, Sigurd and Elsa Goljes Memory Foundation and the Wilhelm and Martina Lundgren Research Foundation.

#### References

- Sollid LM: Molecular basis of celiac disease. *Annu Rev Immunol* 2000; **18**: 53–81.
- Ascher H, Krantz I, Rydberg L, Nordin P, Kristiansson B: Influence of infant feeding and gluten intake on coeliac disease. *Arch Dis Child* 1997; **76**: 113–117.
- Petronzelli F, Bonamico M, Ferrante P *et al.*: Genetic contribution of the HLA region to the familial clustering of coeliac disease. *Ann Hum Genet* 1997; **61**: 307–317.
- Sollid LM, Thorsby E: HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 1993; **105**: 910–922.
- Hernandez JL, Michalski JP, McCombs CC, McCarthy CF, Stevens FM, Elston RC: Evidence for a dominant gene mechanism underlying coeliac disease in the west of Ireland. *Genet Epidemiol* 1991; **8**: 13–27.
- Hervonen K, Karell K, Holopainen P, Collin P, Partanen J, Reunala T: Concordance of dermatitis herpetiformis and celiac disease in monozygous twins. *J Invest Dermatol* 2000; **115**: 990–993.
- Risch N: Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet* 1987; **40**: 1–14.
- Greco L, Corazza G, Babron MC *et al.*: Genome search in celiac disease. *Am J Hum Genet* 1998; **62**: 669–675.
- King AL, Yiannakou JY, Brett PM *et al.*: A genome-wide family-based linkage study of coeliac disease. *Ann Hum Genet* 2000; **64**: 479–490.
- Zhong F, McCombs CC, Olson JM *et al.*: An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. *Nat Genet* 1996; **14**: 329–333.
- Holopainen P, Mustalahti K, Uimari P, Collin P, Maki M, Partanen J: Candidate gene regions and genetic heterogeneity in gluten sensitivity. *Gut* 2001; **48**: 696–701.
- Houlston RS, Tomlinson IP, Ford D *et al.*: Linkage analysis of candidate regions for coeliac disease genes. *Hum Mol Genet* 1997; **6**: 1335–1339.
- Brett PM, Yiannakou JY, Morris MA *et al.*: A pedigree-based linkage study of coeliac disease: failure to replicate previous positive findings. *Ann Hum Genet* 1998; **62**: 25–32.
- Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child* 1990; **65**: 909–911.
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smoth JA, Struhl K (eds): Current protocols in molecular biology. Wiley Interscience, New York, 1995.
- Naluai AT, Nilsson S, Samuelsson L *et al.*: The CTLA4/CD28 gene region on chromosome 2q33 confers susceptibility to celiac disease in a way possibly distinct from that of type 1 diabetes and other chronic inflammatory disorders. *Tissue Antigens* 2000; **56**: 350–355.
- Holopainen P, Arvas M, Sistonen P *et al.*: CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. *Tissue Antigens* 1999; **53**: 470–475.
- O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
- Daly MJ, Kruglyak L, Pratt S, Houstis N, Reeve MP, Kirby A, Lander ES: GENEHUNTER 2.0-a complete linkage analysis system. *Am J Hum Genet Suppl* 1998; **63**: A286.
- Ronningen KS, Iwe T, Halstensen TS, Spurkland A, Thorsby E: The amino acid at position 57 of the HLA-DQ beta chain and susceptibility to develop insulin-dependent diabetes mellitus. *Hum Immunol* 1989; **26**: 215–225.
- Ploski R, Ascher H, Sollid LM: HLA genotypes and the increased incidence of coeliac disease in Sweden. *Scand J Gastroenterol* 1996; **31**: 1092–1097.
- Djilali-Saiah I, Schmitz J, Harfouch-Hammoud E, Mougnot JF, Bach JF, Caillat-Zucman S: CTLA-4 gene polymorphism is associated with predisposition to coeliac disease. *Gut* 1998; **43**: 187–189.
- Nilsson S: Which genes are involved? -Statistical Planning and Analysis of Human Genetic Samples. Department of Mathematical Statistics, Chalmers University of Technology, Göteborg University, 2001.

- 24 Ventura A, Magazzu G, Greco L: Duration of exposure to gluten and risk for autoimmune disorders in patients with celiac disease. SIGEP Study Group for Autoimmune Disorders in Celiac Disease. *Gastroenterology* 1999; **117**: 297–303.
- 25 Durner M, Vieland VJ, Greenberg DA: Further evidence for the increased power of LOD scores compared with nonparametric methods. *Am J Hum Genet* 1999; **64**: 281–289.
- 26 Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.
- 27 Hawkes ND, Swift GL, Smith PM, Jenkins HR: Incidence and presentation of coeliac disease in South Glamorgan. *Eur J Gastroenterol Hepatol* 2000; **12**: 345–349.
- 28 Litwin SD: Does the X chromosome have a special role in immune biology? *Med Hypotheses* 1982; **9**: 313–323.
- 29 Davies JL, Kawaguchi Y, Bennett ST *et al.*: A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 1994; **371**: 130–136.
- 30 Cucca F, Goy JV, Kawaguchi Y *et al.*: A male-female bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients. *Nat Genet* 1998; **19**: 301–302.
- 31 Furuya T, Salstrom JL, McCall-Vining S *et al.*: Genetic dissection of a rat model for rheumatoid arthritis: significant gender influences on autosomal modifier loci. *Hum Mol Genet* 2000; **9**: 2241–2250.
- 32 Sawcer S, Jones HB, Feakes R *et al.*: A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat Genet* 1996; **13**: 464–468.
- 33 Ebers GC, Kukay K, Bulman DE *et al.*: A full genome search in multiple sclerosis. *Nat Genet* 1996; **13**: 472–476.
- 34 Morahan G, Huang D, Ymer SI *et al.*: Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory IL12B allele. *Nat Genet* 2001; **27**: 218–221.
- 35 Louka AS, Torinsson Naluai Á, Gudjónsdóttir AH *et al.*: IL12B and coeliac disease in Scandinavia. *Tissue Antigens* 2001; **59**: in press.