

ARTICLE

Meta and pooled analysis of European coeliac disease data

Marie-Claude Babron^{*1}, Staffan Nilsson^{2,3}, Svetlana Adamovic³, Åsa Torinsson Naluai³, Jan Wahlström³, Henry Ascher⁴, Paul J Ciclitira⁵, Ludvig M Sollid⁶, Jukka Partanen⁷, Luigi Greco⁸, Françoise Clerget-Darpoux¹ and the European Genetics Cluster on Coeliac Disease⁹

¹INSERM U535, Villejuif, France; ²Chalmers University of Technology, Göteborg, Sweden; ³Department of Clinical Genetics, Göteborg University, Göteborg Sweden; ⁴Department of Pediatrics, The Queen Silvia Children's Hospital, Göteborg University, Göteborg, Sweden; ⁵Department of Gastroenterology, St Thomas' Hospital, London, UK; ⁶Institute of Immunology, Rikshospitalet, University of Oslo, Oslo, Norway; ⁷Department of Tissue Typing, Finnish Red Cross Transfusion Service, Helsinki, Finland; ⁸Dipartimento di Pediatria, Università Federico II and European Laboratory of Food-Induced Disease (ELFID), Naples, Italy

Four full genome scans have been carried out by the partners of the European cluster on coeliac disease as well as follow-up studies of candidate regions. No region outside HLA showed significant linkage to the disease in any single study. We first applied a meta-analysis based on a modification of Genome Screen Meta-Analysis to take into account the different linkage statistics, the arbitrariness of bin cutoff points, as well as the sample size of each study. We then performed a pooled linkage analysis of all families and raw genotypes. Besides the HLA region, already known to harbour a risk factor for coeliac disease, both approaches leave very little doubt on the presence of a genetic risk factor in the 5q31–33 region. This region was suggested by several individual studies, but did not reach statistical values high enough to be conclusive when data sets were analysed separately.

European Journal of Human Genetics (2003) 11, 828–834. doi:10.1038/sj.ejhg.5201051

Keywords: coeliac disease; meta-analysis; mega-analysis; GSMA; genome scan; pooled analysis

Introduction

Coeliac disease (CD) is a permanent immune-mediated intolerance to the gluten contained in cereals such as wheat, barley and rye. CD has a multigenic and multifactorial origin: the existence of a strong genetic predisposition is supported by a 10% prevalence in first-degree relatives and a 75% concordance in monozygotic twins, compared to 11% in dizygotic twins.¹ A well-known

predisposing factor lies in the HLA class II region. More than 85% of coeliac patients carry the same DQ heterodimer DQA1*05 DQB1*02 encoded in *cis* or in *trans*.² Almost all the remaining cases carry either the DQA1*05 or the DQB1*02 part of the DQ heterodimer, or DRB1*04-DQ8. However, only part of the familial aggregation is explained by the HLA component, suggesting the presence of additional non-HLA genetic risk factors, as shown by the concordance rate of 20% in dizygotic twins sharing two HLA haplotypes identical by descent compared to 75% in monozygotic twins.¹

The European Cluster on the Genetics of Coeliac Disease has been set up in 2000 to investigate the genetic component of CD. The collaboration involves six partners (Italy, Sweden, Norway, Finland, France and the UK). Within this cluster, four independent genome scans^{3–6} have been performed. While each study highlighted

*Correspondence: Dr M-C Babron, INSERM U535, Hôpital Paul Brousse, Bâtiment Leriche, Secteur Jaune-Porte 18, B.P. 1000, 94817 Villejuif Cedex, France. Tel: +33 1 45 59 53 64; Fax: +33 1 45 59 53 31; E-mail: babron@vjf.inserm.fr

The first two authors contributed equally to this work

⁹See the Appendix for the list of participants of the European genetics cluster of coeliac disease.

Received 16 September 2002; revised 16 May 2003; accepted 21 May 2003

regions of interest, none of them achieved the genome-wide level of significance, except for the HLA region. Regions of interest were further investigated without definite conclusions,⁷⁻¹⁰ even if region 5q, originally found in the Italian genome scan⁷ gave repeatedly moderate evidence for linkage. The interpretation of regions that coincide is, however, not so easy to do. This study addresses the question by analysing the previous genome screens and follow-up studies within a common framework through two approaches: a meta-analysis and a mega-analysis, in which the individual data sets are pooled prior to analysis.

Many research groups across the world are studying the genetic component of the same human disease. The need for globally interpreting results has promoted the development of meta-analysis methods.¹¹⁻¹⁹

The meta-analysis method used here is the Genome Screen Meta-Analysis (GSMA).¹³ The method is based on ranking genetic regions or bins according to the value of the statistic or of the *P*-value achieved in each scan. The first application of GSMA was on four published genome scans on multiple sclerosis. An extension of the method²⁰ has been proposed to take into account studies performed on more restricted regions of the genome.

Since we had access to the raw data of each partner, it was possible to perform a pooled linkage analysis of the cluster data, or mega-analysis, in order to compare the results obtained by the two methods.

Family data

Families with a minimum of two affected siblings were collected in all participating countries as described in the individual papers. The affected individuals were almost exclusively diagnosed according to the ESPGHAN criteria.²¹ The Italian sample consists of 103 families from the genome-wide scan³ and 83 families from a replication study.⁷ The Scandinavian sample includes 60 families.⁵ Between 15 and 40 additional families, depending on the region, were used in the densification analysis.^{5,8} The Finnish sample totals 60 families for the full genome screen,⁶ and a second set of 38 families.⁹ The UK sample has 16 families in the genome-wide scan⁴ and an additional 34 families in the replication study.¹⁰ Some additional families and/or markers, not included in the original publications, were also made available to us.

All families are basically affected sibship families except for six of the initial 16 British families,⁴ which are more extended and one Finnish family with two affected sibships on three generations. When all the data are pooled together in a single data set, there are altogether 442 families with CD and raw genotypes of 2025 individuals of whom 1056 are affected. Parents were generally available except in the Finnish and British families. In total, 80% of all founders were available for genotyping.

Marker data

The marker and map panel used in the different screens were different from centre to centre, but the average intermarker distances were around 10–15 cM. All centres performed additional studies of more restricted regions. The number of regions and markers followed up also varied from centre to centre, with an intermarker spacing of the order of 5 cM.

Altogether 1092 markers, mostly microsatellites, were genotyped in at least one population. A common map of all markers was constructed using genetic distances from Marshfield (<http://research.marshfieldclinic.org/genetics/>). Markers that had the same position in Marshfield were ordered according to the human genome database (<http://gdbwww.gdb.org>) and arbitrarily separated by 0.01 cM. More than 97% of the markers were mapped in this way. When a marker was not found in the Marshfield database, other map databases were used and the genetic position interpolated using flanking markers available in both databases. Many markers were used in more than one population. Two problems had to be faced in these cases. The allele coding of the length polymorphisms was different and homogenous allele frequencies between populations could not *a priori* be assumed. Given the fact that 20% of founders are untyped, information on marker allele frequencies is needed to avoid potential bias in linkage analysis results. Both these problems can be nicely overcome by the same approach used in a meta-analysis of multiple sclerosis genome screens,²² that is, a prefix indicating the population was added to each marker name. So, for example, marker D1S2141 was called FD1S2141 when genotyped in Finnish families and SDS2141 when genotyped in Scandinavian families, but both synonyms have the same genetic position. The allele frequencies were separately estimated from the genotype data in each population.

Principle of meta-analysis by GSMA

To apply the GSMA method,¹³ the genome needs to be fragmented into bins of equal lengths. Based on the common map described above, each chromosome is divided into bins of approximately 30 cM, making sure that at least one marker from each country was present in any bin. This bin length was suggested by Wise *et al* as a good compromise to avoid having peaks corresponding to the same genetic risk factor falling into different bins and having peaks corresponding to two different genetic risk factors falling into the same bin. A total of 115 bins were thus available for ranking.

For each genome scan, the maximum value of the linkage statistic obtained in each bin scan is recorded. The bins are then ranked, with rank 1 and 115 representing the lowest and highest value of the statistics, respectively. The ranks are then summed within each bin across the

genome scans. The summed rank across studies is compared to its probability distribution under the hypothesis of no linkage assuming ranks are randomly assigned. This procedure provides the nominal *P*-value of each bin.

Follow-up studies do not, by definition, cover the whole genome. The problem is then to assign a rank between 1 and 115 to each reinvestigated bin. We applied the extension proposed by Wise.²⁰ First, the observed maximum value of the statistics in the follow-up bin is recorded. Simulations under the hypothesis of no linkage, using the family structures, affected status, and marker information of the follow-up studies, are carried out to obtain the distribution of the maximum statistic values. The ranks the observed statistics would have are obtained from this distribution. The analysis then proceeds as above, with calculation of the summed rank across studies and of the *P*-value.

Of the total 115 bins, 69 (60%) were followed up: 55, 11 and three bins were available in five, six and seven studies, respectively.

Application of GSMA to Coeliac Disease data

Here, GSMA was modified for the application to the Coeliac Disease data, to account for some difficulties of the method, in particular the different linkage statistics which can be used in a genome scan, the arbitrariness of bin cutoff points and the different information content of each sample. Sensitivity of the results to those three points was evaluated.

Different linkage statistics

As originally proposed, GSMA is based on published results of studies that are not necessarily based on the same linkage statistics. Here, all four CD genome scans were carried out using different statistics. Recording the *P*-value of the maximum statistics for each bin in some studies was therefore difficult. Here, since the raw data were available, all samples were reanalysed with the same statistics. The statistics chosen was the Zlr of Kong and Cox²³ implemented in the Allegro software.²⁴ This model-free statistic is based on the allele-sharing among the affected individuals in the pedigree and then transformed to get an asymptotically standard normal distribution when there is no linkage. It is similar to the perhaps better known Non Parametric Linkage (NPL) statistic in Genehunter,²⁵ but not as conservative when the meiotic information, as is the case in most real studies, is far from complete. The marker map and marker allele frequencies were those specified in the original studies. The bins were ranked according to the observed maximum Zlr value. Note that there were no tied ranks, as for each genome scan, all values of Zlr were different.

To evaluate the sensitivity of the method to the linkage statistics, we also performed the meta-analysis using the

ranks obtained with the initial statistics for the Italian and Scandinavian data.

Arbitrariness of bin cutoff points

Even when there is a risk factor on the genome, the maximum statistics will often not be obtained on the exact position of the functional factor. It has been shown²⁶ that, for a genetic risk factor with a weak effect, only one third of the maximum Maximum Lod Score (MLS)²⁷ values are at the correct location, while as many as 14% are observed at more than 10 cM from the true position. This result also holds for the Zlr statistic. Thus, the maximum in two different studies could be observed in two contiguous bins. This is true whatever the length of the bins, if the true position is close to a cut-off point. GSMA would then fail to detect the risk factor. To minimise this problem, we propose to use an average rank (A1) over three bins, where the new rank of a bin is the average of its own rank and of the two flanking bins. Denote by *R* the rank of the bin under consideration and by *R*₋₁ and *R*₊₁ the ranks of the two flanking bins. Then the new rank *R* is computed as $(R_{-1} + R + R_{+1})/3$. If the bin has only one flanking bin—the two most extreme bins in a chromosome, or chromosomes 21 and 22 that consist of two bins only—the average is calculated over those two bins (ie either $(R + R_{+1})/2$ or $(R_{-1} + R)/2$). We investigated the sensitivity of the meta-analysis to the choice of the averaging procedure by comparing the results under another averaging (A2). In A2, the new rank of a bin *R* equals $(R_{-1} + 2R + R_{+1})/4$. In other words, score A1 gives equal weight to three consecutive bins, while score A2 gives more weight to the bin under consideration.

Information content

GSMA does not take into account the different information content of each study. This can be performed by weighting each rank. For the sake of simplicity, we chose to weight the individual ranks by the number of pedigrees in the data set, where extended families, as well as nuclear families with several affected sibs, contributed one unit.

We studied the sensitivity of the meta-analysis to the weighting procedure by applying another weighting score W2, where the rank of a bin in a sample is weighted by the value of the statistics observed for HLA in that sample, since HLA is known to be a genetic risk factor for CD. The comparison of the results, however, can only be made on the full genome scans since the HLA information is not available in the follow-up studies. The Zlr for HLA were equal to 5.15, 2.69, 3.99 and 5.58 in the Italian, British, Scandinavian and Finnish data sets, respectively.

Determination of *P*-value of each bin

For a given weighting (W1 or W2) and averaging (A1 or A2) scheme, the null distribution of the summed score across

studies for a given bin is assessed as follows: for each study a simulated rank value is obtained by a random draw from the 115 observed average rank values in that study. The weighted sum across studies is then calculated. This process is repeated 50 000 times. The *P*-value of an observed weighted summed rank is then obtained by the number of times the observed value was exceeded by a corresponding simulated value divided by 50 000.

This procedure provides the nominal *P*-value for a given bin, and not the overall significance of the meta-analysis, which could only be assessed by heavy simulations. Using a Bonferroni correction, a nominal *P*-value of 4.3×10^{-4} would correspond to a global type I error of 5% if the bin ranks were independent. However, the observations are not independent since the ranks of adjacent bins are strongly correlated, even more so with the averaging procedure.

Mega-analysis

Mega-analysis was performed on all the pooled raw genotypes, genome scan and follow-up, using all the markers of the common map. Linkage scores were calculated as the statistic Zlr²³ using the Allegro²⁴ linkage software.

For the pooled analysis, a nominal *P*-value of 2×10^{-5} (as suggested by Lander and Kruglyak²⁸) will correspond to a global *P*-value smaller than 5%, first because the genome scans were not performed with a dense map (average intermarker spacing about 10 cM). Moreover, the follow-up studies were performed only on a restricted number of regions chosen conditionally to the genome scans results.

Results

Figure 1 shows the result of GSMA with average score A1, weighting scheme W1 and the Zlr statistics, on the full genome scans and follow-up regions. Two regions are significant at a nominal 1% level.

The most significant region ($P \sim 10^{-6}$) corresponds to the HLA region on chromosome 6. HLA is located in bin 42, but the effect of the HLA risk factor is so large that the two neighbouring bins (41 and 43) are also very significant. In fact, bin 42 was ranked highest in three genome scans (Italy, Finland, UK) and third highest in the Norway/Sweden one.

The second most significant region is chromosome 5qter. Overall, two bins, 39 and 40, achieve a nominal *P*-value of 0.003 and 0.006 respectively. Note that bin 39 and 40 were studied six and seven times, respectively. This region, found originally in the Italian study³ was also found in the Scandinavian⁵ and Finnish⁶ screens, although it did not reach statistical significance in the individual genome scans. It is of interest that the meta-analysis supports these results.

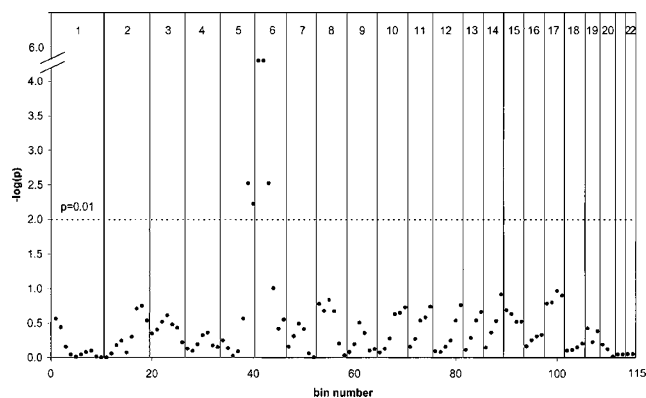


Figure 1 Meta-analysis of the coeliac disease genome scans and follow-up studies. Each dot represents $-\log(P\text{-value})$ of a given bin. The vertical lines indicate the chromosome.

We then assessed the robustness of results to our modifications of GSMA. The three HLA bins (41, 42 and 43) and bin 39 and/or 40 on 5q were always the most significant ones, whatever the averaging and ranking procedures. Note, however, that only HLA is detected when GSMA is applied without averaging ranks. Indeed, although 5q was found in most studies, the peak statistic shifted from study to study. This is what is expected for a risk factor with moderate effect.²⁶ Averaging ranks across adjacent bins minimises this effect and thus allows detection of 5q. Finally, the ranks obtained using the MLS for the Italian genome scan and the NPL for the Scandinavian one, did not have an impact on the results of the meta-analysis, since the rank order was the same as that of the Zlr for HLA and 5q.

The two regions found by meta-analysis are coherent with those detected by the mega-analysis. The results of the pooling method for the full genome scans and the follow-up studies are shown graphically for each chromosome in Figure 2. The best marker on each chromosome is shown in Table 1. Apart from the HLA region, only 5q31–33, with a maximum of Zlr = 4.39 ($P = 6 \times 10^{-6}$), showed genome-wide significant linkage according to standard thresholds.²⁸ Two other regions, 11q and 14q, achieved a nominal $P < 0.01$. The 11q region corresponds to that highlighted in the Italian³ and Scandinavian⁵ genome scans, while the 14q exhibited a moderate NPL in the Finnish scan.⁶

Discussion

The most important result of this study is that it leaves very little doubt on the presence of a genetic risk factor involved in CD in the 5q31–33 region. Several genome scans had already pinpointed this region, without however individually achieving the recommended genome-wide

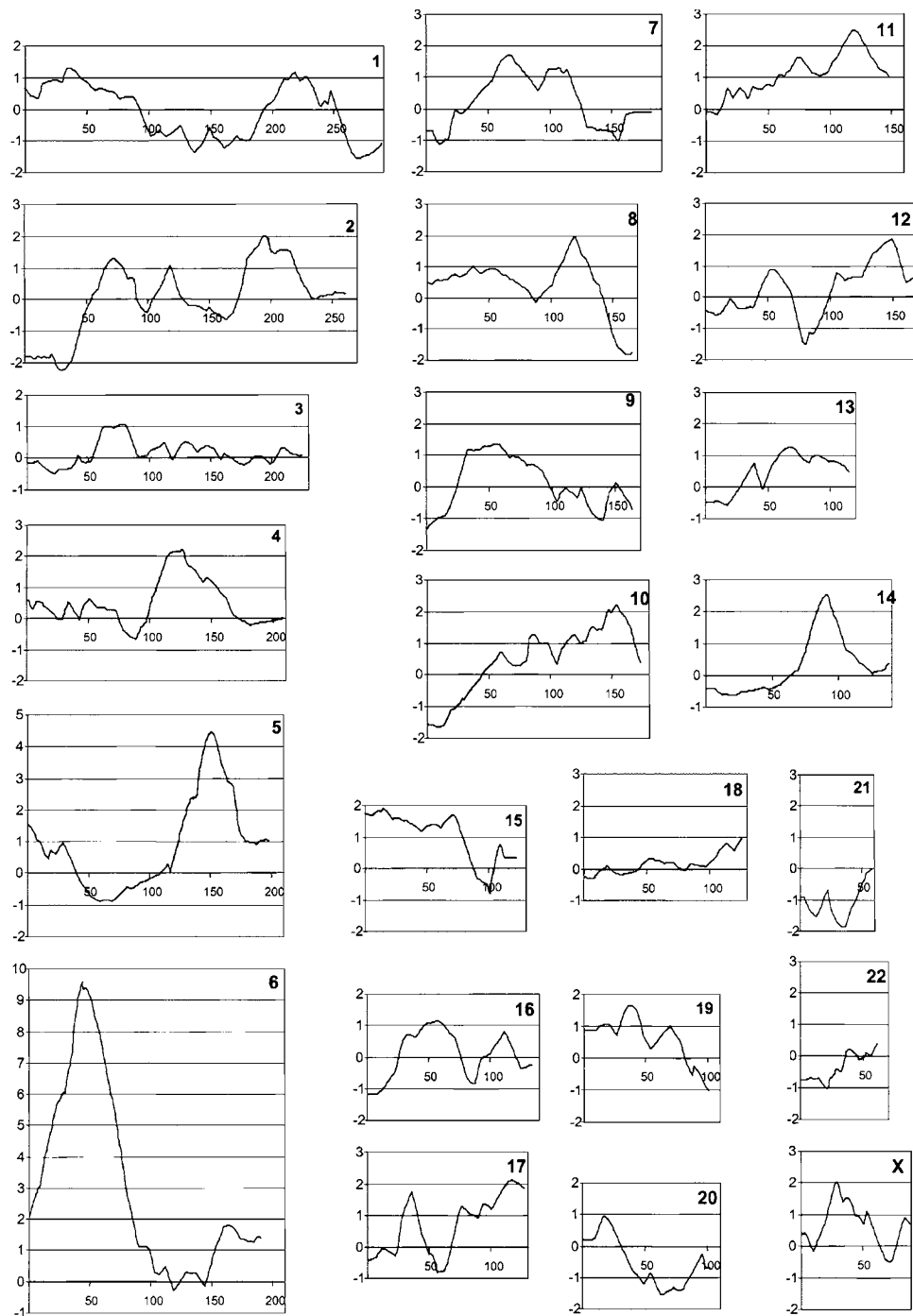


Figure 2 Zlr curves of pooled analysis for each chromosome.

level of significance. This result shows the interest of international collaboration and of joining efforts for global interpretation.

The meta-analysis by GSMA reflects the importance of taking into account the repeatability of linkage findings.

One expects such a meta-analysis based on the ranks of the maximum value of a statistics to be not as powerful as an analysis, where all data are pooled and analysed as one data set. However, the two approaches gave consistent results, by giving evidence for the same two regions, HLA and 5q31–33.

Table 1 Highest linkage scores per chromosome in the pooled analysis

Chr.	cM	Marker	Linkage score	
			Zlr	P-value
6	44.41	D6S1281	9.58	<10 ⁻²¹
5	152.62	D5S640	4.39	0.000006
14	91.62	D14S606	2.52	0.006
11	118.47	D11S925	2.47	0.007
4	126.70	D4S3250	2.22	0.013
10	153.78	D10S575	2.19	0.014
17	116.86	D17S784	2.12	0.017
X	29.25	DXS6810	2.03	0.021
2	194.45	D2S117	2.00	0.023
8	119.23	D8S1132	1.96	0.025
15	14.58	D15S1002	1.88	0.030
12	149.60	D12S2078	1.85	0.032
7	69.03	D7S519	1.69	0.046
19	36.22	D19S221	1.65	0.050
9	58.26	D9S1118	1.35	0.089
1	33.75	D1S507	1.29	0.098
13	68.73	D13S265	1.26	0.103
16	57.79	D16S753	1.15	0.126
3	78.64	D3S1766	1.07	0.142
18	126.00	D18S70	0.97	0.166
20	16.65	D20S95	0.95	0.172
22	60.61	D22S1169	0.37	0.356
21	57.77	D21S1446	-0.03	0.512

The modifications we proposed in GSMA-AW attempt to address the three main difficulties of the method: the use of different linkage statistics, the bin construction and the incorporation of the information content provided by each sample.

The bins are ranked according to the scores achieved in each scan for a particular statistic. Exactly which statistic was used can be difficult to gather from the literature. Here, the situation was ideal, since the same statistic was used in all studies. However, we showed here that the most significant results were found, even with different statistics.

Similarly, HLA and 5q31–33 are detected, if the bins are attributed ranks that take into account those of adjacent bins. If no such averaging score is introduced, the significance of region 5q is considerably decreased. Indeed, in the two Italian studies, the maximum MLS values were obtained for markers in two adjacent bins, bin 40 in the genome scan (MLS = 2.0 for D5S2006) and in bin 39 in the replication study (MLS = 1.06 for D5S640). However, we had shown⁷ that under the hypothesis of a genetic risk factor with moderate effect, the probability of observing two peaks at a distance of 35 cM was not negligible and that the 95% confidence interval of the maximum MLS spanned the entire interval tested.

Weighting each rank to take into account the linkage information provided by each sample is fairly intuitive. Indeed, there is a substantial variation of the information

content between different regions due to the number of families analysed, the availability of founders, marker heterozygosity, marker density and genotyping success rate. Here, each rank was weighted by the number of families, whatever their structure (nuclear or extended), and whatever their number of affecteds. We investigated the impact of another weighting scheme, based on the value of the statistic for the already known risk factor in HLA. This could only be done on the full genome scans. We showed that the results for the two most significant regions were the same with both weighting procedures.

Note that other weighting schemes could be devised. However, the optimal weights depend on the genetic model of the risk factor to detect, which is unknown.

The 5q31–qter region is interesting as it has been repeatedly highlighted in studies of several multifactorial diseases, either inflammatory, such as Crohn disease,²⁹ abnormal immune response diseases, such as type I diabetes,³⁰ asthma and allergy,³¹ as well as infectious diseases such as schistosomiasis.³² This chromosome 5 region contains many candidate genes of interest for Coeliac Disease. A candidate gene strategy is now underway by the partners of the EU cluster.

Acknowledgements

We thank all the patients and families who participated in this study. We thank all those who contributed to the genome scans, in particular MC Fulchignoni-Lataud (Paris, France), R Sica, and P Zavattari (Napoli, Italia), J Liu and C Gilliam (Columbia Genome Center, New York), and M Guilloud-Bataille (Le Kremlin-Bicêtre, France) for technical assistance.

This study was funded by the Commission of the European Communities, RTD programme 'Quality of Life and Management of Living Resources' (QLKT-1999-00037), 'Evaluation of the prevalence of coeliac disease and its genetic components in the European population'. This paper does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area. Grants for the study were also given by The Swedish Medical Research Council, Wilhelm and Martina Lundgren's Research Foundation, The Swedish Medical Society, the Swedish Coeliac Society, the Sigrid Juselius Foundation, the Medical Research Funds of Finnish Red Cross Blood Transfusion Service, the University of Helsinki, Emil Aaltonen Foundation, Maud Kuistila Memorial Foundation, Päivikki and Sakari Sohlberg Foundation, Foundation of the Friends of the University Children's Hospitals in Finland, Academy of Finland Research Council for Health (decision 73489), and the Medical Research Fund of Tampere University Hospital.

References

- Greco L, Romino R, Coto I *et al*: The first large population based twin study of coeliac disease. *Gut* 2002; 50: 624–628.
- Sollid LM: Molecular basis of celiac disease. *Annu Rev Immunol* 2000; 18: 53–81.
- 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.

- 5 Naluai AT, Nilsson S, Gudjonsdottir AH *et al*: Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11. *Eur J Hum Genet* 2001; **9**: 938–944.
- 6 Liu J, Juo SH, Holopainen P *et al*: Genomewide linkage analysis of celiac disease in Finnish families. *Am J Hum Genet* 2002; **70**: 51–59.
- 7 Greco L, Babron MC, Corazza GR *et al*: Existence of a genetic risk factor on chromosome 5q in Italian coeliac disease families. *Ann Hum Genet* 2001; **65**: 35–41.
- 8 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.
- 9 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.
- 10 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**: 377–386.
- 11 Allison DB, Heo M: Meta-analysis of linkage data under worst-case conditions: a demonstration using the human OB region. *Genetics* 1998; **148**: 859–865.
- 12 Gu C, Province M, Todorov A, Rao DC: Meta-analysis methodology for combining non-parametric sibpair linkage results: genetic homogeneity and identical markers. *Genet Epidemiol* 1998; **15**: 609–626.
- 13 Wise LH, Lanchbury JS, Lewis CM: Meta-analysis of genome searches. *Ann Hum Genet* 1999; **63** (Part 3): 263–272.
- 14 Wang K, Vieland V, Huang J: A Bayesian approach to replication of linkage findings. *Genet Epidemiol* 1999; **17** (Suppl. 1): S749–S754.
- 15 Xu J, Wiesch DG, Taylor EW, Meyers DA: Evaluation of replication studies, combined data analysis, and analytical methods in complex diseases. *Genet Epidemiol* 1999; **17** (Suppl. 1): S773–S778.
- 16 Guerra R, Etzel CJ, Goldstein DR, Sain SR: Meta-analysis by combining *P*-values: simulated linkage studies. *Genet Epidemiol* 1999; **17** (Suppl. 1): S605–S609.
- 17 Heo M, Leibel RL, Boyer BB *et al*: Pooling analysis of genetic data: the association of leptin receptor (LEPR) polymorphisms with variables related to human adiposity. *Genetics* 2001; **159**: 1163–1178.
- 18 Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M: Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 2001; **69**: 936–950.
- 19 Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS: Truncated product method for combining *P*-values. *Genet Epidemiol* 2002; **22**: 170–185.
- 20 Wise LH: Inclusion of candidate region studies in meta-analysis using the genome screen meta-analysis method: application to asthma data. *Genet Epidemiol* 2001; **21** (Suppl. 1): S160–S165.
- 21 European Society of Paediatric Gastroenterology, Nutrition: 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.
- 22 Transatlantic Multiple Sclerosis Genetics Cooperative: A meta-analysis of genomic screens in multiple sclerosis. *The Transatlantic Multiple Sclerosis Genetics Cooperative. Mult Scler* 2001; **7**: 3–11.
- 23 Kong A, Cox NJ: Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997; **61**: 1179–1188.
- 24 Gudbjartsson DF, Jonasson K, Frigge ML, Kong A: Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000; **25**: 12–13.
- 25 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
- 26 Clerget-Darpoux F, Selinger-Leneman H, Babron MC: Why do complex traits resist DNA analysis? *Int J Hum Genet* 2001; **1**: 55–63.
- 27 Risch N: Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet* 1990; **46**: 242–253.
- 28 Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.
- 29 Rioux JD, Daly MJ, Silverberg MS *et al*: Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001; **29**: 223–228.
- 30 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.
- 31 Cookson WO, Moffatt MF: Genetics of asthma and allergic disease. *Hum Mol Genet* 2000; **9**: 2359–2364.
- 32 Marquet S, Abel L, Hillaire D, Dessein A: Full results of the genome-wide scan which localises a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31–q33. *Eur J Hum Genet* 1999; **7**: 88–97.

Appendix

Members of the European Genetics Cluster on Coeliac Disease

Participants in the European Genetics Cluster on Coeliac Disease (project partner group leaders are underlined):

Finland—P Holopainen, K Karell, J Partanen (*Department of Tissue Typing, Finnish Red Cross Blood Transfusion Service, Helsinki*), P Collin, K Mustalahti, M Mäki (*Institute of Medical Technology and Department of Pediatrics, University of Tampere, Tampere*). **France**—F Clerget-Darpoux, M-C Babron (*INSERM U535, Villejuif*), F Clot (*INSERM U535, Villejuif, Fondation Jean Dausset, CEPH, Paris*), J-P Hugot (*Fondation Jean Dausset, CEPH, Paris; Hôpital Robert Debré, Paris*). **Italy**—S D'Alfonso, E Bolognesi, M Giordano, M Mellai, P Momigliano-Richiardi (*Department Medical Sciences, Eastern Piedmont University and I.R.C.A.D. (Interdisciplinary Research Center on Autoimmune Diseases), Novara*), I Coto, L Greco, S Percopo (*Department of Paediatrics and International Laboratory for Food Induced Diseases, University of Naples Federico II, Naples*), R Tosi (*Istituto Biologia Cellulare, CNR, Monterotondo, Rome*). **Norway**—AS Louka, LM Sollid and B Talseth (*Institute of Immunology, Rikshospitalet University Hospital, Oslo*), J. Ek (*Buskerud Hospital Trust, Drammen*). **Sweden**—H Ascher, AH Gudjonsdottir, B Kristiansson (*Department of Paediatrics, Göteborg University, The Queen Silvia Children's Hospital, Göteborg*), S Adamovic, T Martinsson, L Samuelsson, Å Torinsson Naluai, J Wahlström (*Department of Clinical Genetics, Göteborg University, Sahlgrenska University Hospital/Östra, Göteborg*), S. Nilsson, Olle Nerman (*Department of Mathematics, Chalmers University of Technology, Göteborg*). **United Kingdom**—PJ Ciclitira, J Fraser, A King, SJ Moodie (*Gastroenterology Unit, GKT, The Rayne Institute, St Thomas's Hospital, London*).