Linkage analysis of HLA and candidate genes for celiac disease in a North American family-based study

Susan L Neuhausen*, Michael Feolo¹,², James Farnham¹, Linda Book³ and John J Zone⁴

Address: ¹Medical Informatics, University of Utah, Salt Lake City, UT, USA, ²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA, ³Pediatric Gastroenterology and Nutrition, Department Pediatrics, University of Utah, Salt Lake City, UT, USA and ⁴Dermatology, University of Utah, Salt Lake City, UT, USA

E-mail: Susan L Neuhausen* - susan@genepi.med.utah.edu; Michael Feolo - feolo@ncbi.nlm.nih.gov; James Farnham - jfarnham@genepi.med.utah.edu; Linda Book - pclbook@ihc.com; John J Zone - zone@ultraderm.med.utah.edu

*Corresponding author

Background

Celiac disease (CD) is a common, familial, autoimmune gastrointestinal disease. It is caused by sensitivity to the dietary protein gluten, which is present in wheat, rye and barley. Symptoms include growth failure, abdominal pain, and diarrhea. Dermatitis herpetiformis is a cutaneous manifestation of CD. Complications of CD include lymphoma, osteoporosis, anemia, and seizures. The prevalence of CD in the US is 1:250 [1] and the ratio of symptomatic to asymptomatic cases is between 1:5 and 1:7 [2]. Before the advent of serological testing for diagnosing CD, it was considered a rare disease in the US.

The clinical standard for diagnosis of CD is a small intestinal biopsy showing villus atrophy and resolution of symptoms on a gluten-free diet. However, small intestinal biopsy is expensive, invasive, and often rejected by the US patient population. The serological IgA endomy-

Abstract

Background: Celiac disease has a strong genetic association with HLA. However, this association only explains approximately half of the sibling risk for celiac disease. Therefore, other genes must be involved in susceptibility to celiac disease. We tested for linkage to genes or loci that could play a role in pathogenesis of celiac disease.

Methods: DNA samples, from members of 62 families with a minimum of two cases of celiac disease, were genotyped at HLA and at 13 candidate gene regions, including CD4, CTLA4, four T-cell receptor regions, and 7 insulin-dependent diabetes regions. Two-point and multipoint heterogeneity LOD (HLOD) scores were examined.

Results: The highest two-point and multipoint HLOD scores were obtained in the HLA region, with a two-point HLOD of 3.1 and a multipoint HLOD of 5.0. For the candidate genes, we found no evidence for linkage.

Conclusions: Our significant evidence of linkage to HLA replicates the known linkage and association of HLA with CD. In our families, likely candidate genes did not explain the susceptibility to celiac disease.
sial antibody (EMA) test is a screening tool that has greatly facilitated evaluation for CD in people with suggestive symptoms and in high-risk populations. IgA EMA testing has proven to be greater than 95% sensitive for adults and children with classic symptomatic CD [3–10] and greater than 98% specific in controls without known clinical disease [11,12]. It is therefore an inexpensive and specific method of screening family members for genetic studies. Moreover, a recent study has identified symptomatic EMA positive individuals who have CD in whom intestinal biopsies were normal with only minor mucosal lesions. All the patients showed clinical and serological recovery on a gluten-free diet. They propose that serologic criteria may be more definitive in the diagnostic process than traditional biopsy criteria [13].

CD has a strong genetic association with the HLA class II DQ2 genotype composed of the DQA1*05 and DQB1*02 alleles [14]. However, the HLA association alone is insufficient to explain the hereditary nature of the disease, and is estimated to explain less than half the sibling risk [15–18]. There appears to be genetic heterogeneity, implying that more than one additional gene is involved in the disease. With current analysis software, it is possible to map complex traits like CD, where several genetic loci are probably involved and the mode of inheritance is unclear.

One first step to identifying genes predisposing to CD is to investigate candidate genes. Likely candidates include the classes of genes involved in immune function, e.g., T-cell receptor (TCR) genes and immune-modulating genes. Other candidate genes are those from associated, independent diseases in which there is a higher rate of CD than in the general population, e.g., other autoimmune diseases such as insulin dependent diabetes mellitus (IDDM). These associations may be explained by common gene(s) responsible for both diseases or the diseases may share a similar autoimmune pathogenic mechanism [19]. There have been several European studies to localize genes for CD, but no significant evidence for linkage has been reported other than at HLA [20–29].

In this first study of families with CD from North America, we investigated linkage to several candidate genes that could play a role in the pathogenesis of CD using 62 families with at least two cases of CD.

Methods

Ascertainment of families with CD

Families with at least two cases of CD or dermatitis herpetiformis were ascertained through local gastroenterologists, gluten intolerance support groups, and advertising at local and national celiac disease support meetings. There was no selection of cases based on sex or race, although all individuals were Caucasian. None of the families appear to be related. The research study was approved by the University of Utah Health Sciences Center Institutional Review Board. Participants ranged in age from 2 years to 100+ years. Blood samples were collected from affected individuals and their first-degree relatives. For more distantly related cases, we also collected blood from individuals that are connections between the cases. For example, for two affected grandchildren (with different parents) and an affected grandparent, we would collect samples from the grandchildren, their parents, and the grandparent. The breakdown of the affected individuals is shown in Table 1.

### Table 1: Characteristics of CD Cases in the Study Population

|                          | N (%) |
|--------------------------|-------|
| Number of cases          | 175   |
| Number of pedigrees      | 62    |
| Number of males          | 55 (31%) |
| Number of females        | 120 (69%) |
| Diagnostic group:        |       |
| D1: Intestinal biopsy    | 115 (66%) |
| D2: Skin biopsy          | 12 (7%)  |
| D3: Positive serum test  | 48 (27%) |

*Of the 175 affecteds, 90% carried the HLA DQ2 genotype, with the majority of the remainder being DQ8. For those affected diagnosed by serology with no biopsy, 94% had the DQ2 genotype (45/48) and the remaining 3 affecteds were DQ8.

Diagnostic criteria

Medical records were obtained to confirm previous biopsy-proven CD or dermatitis herpetiformis. IgA EMA testing was performed for participants who did not have a biopsy proven diagnosis of CD or dermatitis herpetiformis. Since IgA EMA is highly sensitive and specific for CD, we did not require biopsy confirmation for phenotype assignment.

IgA EMA was measured by indirect immunofluorescence using primate smooth muscle (IMCO Diagnostics, Buffalo, New York) as substrate [30]. IgA EMA titers greater than or equal to 1:5 were considered positive. Limiting dilution was performed on the positive sera.

Genotyping at short tandem repeat markers (STRs)

DNA was extracted from lymphocytes using PureGene DNA isolation kits (Genta Systems Inc.). HLA DQA1 and DQB1 genotypes were determined as described in Feolo et al. [31]. Genotyping of DNA samples from 175 affected individuals, their parents, and any connecting rel-
atives from 62 families was performed with 25 markers at 13 candidate gene regions and 4 markers at HLA. However, all families were not genotyped with all markers, because some families were collected after genotyping had been done for some of the STRs. The candidate gene regions, markers, and chromosomal locations are listed in Table 2. For all markers, amplification of 20 ng genomic DNA in a total reaction mix of 10 µl was performed according to standard PCR procedures, with minor modifications to optimize product clarity. Genotyping was performed either using an ABI373 or radioactively using polyacrylamide gels. Genotypic data were stored in the same database as all kindred and phenotype information.

### Linkage analysis methods

Analyses were performed using dominant and recessive genetic models, each with 2 liability classes of either affected or unknown/unknown based on diagnostic criteria (Table 3). For each model, unaffected individuals and individuals with serology or biopsy based diagnosis were given a penetrance function based on disease prevalence. For linkage analysis, we used the FASTLINK [32] implementation of the LINKAGE program [33,34] for two-point analysis, and the GENEHUNTER program [35] for both parametric and non-parametric (NPL) multi-point analyses. Two-point linkage in the presence of locus heterogeneity was assessed by the admixture test of Ott, using HOMOG [36]. We used a heterogeneity LOD (HLOD) of > 1.3 to indicate nominal evidence for linkage for all linkage analyses [37].

### Results

Candidate genes were selected based on function of those genes (i.e., T-cell receptors, CTLA4, and CD4) or from loci of associated diseases (i.e., IDDM). Although associated diseases were not considered in the selection of families, in several families, members had IDDM. In one

| Candidate gene | Chromosome | Position (cM) | Marker | Number families | Model | Best two-point HLOD (α/θ)+ | Positive multi-point HLOD(α) |
|----------------|------------|--------------|--------|-----------------|-------|---------------------------|-----------------------------|
| CD4 | 12pter-pl2 | 16.4 | CD4 | 44 | Rec | 0.0 |
| CTLA-4 | 2q33 | 198.7 | CTLA4A/G | 36 | Dom/Rec | 0.0 |
| IDDM3 | 15q26 | 109.3 | D15S107 | 47 | Dom | 0.14 (0.15/0.0) |
| IDDM4 | 11q13 | 85.5 | GATA30G01 | 47 | Dom | 0.29 (1.00/0.40) |
| IDDM5 | 6q25 | 146.1 | GATA184A08 | 35 | Dom | 0.0 |
| IDDM6 | 18q21 | 84.8 | D18S64 | 41 | Dom | 0.0 |
| IDDM7 | 6q27 | 173.3 | D6S1277 | 39 | Dom | 0.0 |
| IDDM9 | 3q22-q25 | 126.8 | D3S3044 | 46 | Rec | 0.84 (0.05/0.0) |
| IDDM10 | 10p11-q11 | 70.2 | D10S220 | 45 | Dom | 0.69 (1.00/0.20) |
| TCRα | 14q11.2 | 6.3 | D14S261 | 43 | Dom | 0.34 (1.00/0.30) |
| TCRβ | 7q35 | 144.7 | D7S495 | 45 | Rec | 0.25 (1.00/0.30) |
| HLA | 6p21.3 | 44.4 | D6S276 | 34 | Dom | 1.48 (0.95/0.10) |

There were a total of 62 families, however not all families were genotyped for all markers or missing genotypes precluded analysis in the family. + α = proportion of families linked; θ = recombination fraction; Best HLOD from the analyses of dominant and recessive models.
family, a CD case, his sibling and 3 extended relatives had IDDM; in a second family, the CD case had IDDM; in a third family, the mother, 2 siblings, a daughter, and a cousin of a CD case had IDDM; and in a fourth family, the sister of a CD case had IDDM.

The highest 2-point HLOD scores obtained with either model are shown in Table 3. The multipoint HLODs were obtained using the same model as the 2-point HLOD shown. The largest two-point and multipoint LOD scores were obtained in the HLA region. Under the dominant model, the two-point HLOD was 3.1 (α = 1.0) at D6S426 (position 60.4 cM), the multi-point HLOD was 5.02 (α = 0.66) (position 54.6 cM), and the NPL was 4.38 (p < .0001) (position 50.2 cM). The estimate for the proportion of families linked was 0.66 for the multipoint HLOD, suggesting that approximately half of the families are linked to an HLA susceptibility locus for CD. Of the 13 candidate gene regions investigated, none of the regions had even nominal evidence for linkage (HLOD > 1.3) or an NPL score with p < 0.05.

Discussion

In this study, we examined linkage to a set of candidate genes for CD. This subset of genes was selected based on genes that could be related to CD through function or an associated disease. For statistical and linkage analysis of complex diseases, we used general recessive and dominant models. Several biostatisticians have suggested that general models provide power to distinguish linkage signals independent of the true underlying disease mode of inheritance, provided both dominant and recessive models are used [38–40]. As expected, the highest two-point and multipoint LOD scores were obtained in the HLA region, with a two-point HLOD of 3.1 and a multipoint HLOD of 5.0. This result replicates the known association and linkage of HLA to CD [22,25,29] and demonstrates the power of the family resource to detect linkage in the set of candidate gene markers.

We were interested in identifying non-HLA loci for celiac disease. We were unable to detect even nominal evidence for linkage at any of the loci investigated. For those regions where we examined only 1 marker, it may be that one marker was insufficient in order to detect linkage even if it existed. A number of candidate genes investigated in this study were examined previously in European populations. Our results are in agreement with previous linkage and/or association studies of CD and T-cell receptor genes (TCRα, TCRγ, TCRβ, and TCRδ), where they saw no evidence for linkage or association, although sample sizes were small [28,41]. CD28 and CTLA-4, two genes encoding receptors that regulate T-lymphocyte activation, are located at 2q33. Holopainen et al [24] reported linkage and association to this region in a study of 100 Finnish families with CD, which may suggest a possible founder effect in these families. In a case-control study, the CTLA-4 polymorphism, 49A>G, was significantly associated with CD [p = 0.002 with an odds ratio of 2.36 (95% confidence interval 1.37-4.06)] [21]. We did not find evidence for linkage with the CTLA-4 polymorphism.

Genomic searches for CD have been conducted in several European populations. In 1996, Zhong et al. [29] studied 40 affected sib pairs from 11 families, and reported significant linkage at 6p23 and weak evidence at 11p11, 7q31.3, 22cen, 15q26, 5q33.3, 19p13.1 and 19q13.2. Houlston et al [42], studying 28 families, found significant evidence for linkage to HLA, but no evidence for linkage to the regions suggested by Zhong, except at 15q26, where IDDM3 is localized. Greco et al. conducted a genome-wide search with 39 sib pairs, and an additional 71 pairs in regions of interest [22]. They found significant evi-
ence for linkage at HLA and nominal evidence for linkage on 5qter and 11qter. Using an independent set of 89 sibpairs, they reported additional linkage evidence at 5q [23]. King et al. [26] performed a genome-wide search with 16 CD families and reported nominal evidence for linkage at 10q23.1 and 16q23.3. In a follow-up study with 50 families, King et al. [27] reported heterogeneity LOD scores > 2.0 at 5 regions, including 11p11 previously reported by Zhong et al. [29]. From these studies, the only region with at least nominal evidence for linkage, which overlapped with the candidate regions studied here, was at IDDM2 at 15q26. One study reported possible evidence for linkage [28], one reported weak evidence [42], and two reported no linkage [20,22]. We were unable to detect linkage.

Conclusions
Our significant evidence of linkage to HLA replicates the known linkage and association of HLA with CD. In our families, likely candidate genes/loci did not explain the known linkage and association of HLA with CD. In our 50 families, King et al. [27] reported heterogeneity LOD scores > 2.0 at 5 regions, including 11p11 previously reported by Zhong et al. [29]. From these studies, the only region with at least nominal evidence for linkage, which overlapped with the candidate regions studied here, was at IDDM2 at 15q26. One study reported possible evidence for linkage [28], one reported weak evidence [42], and two reported no linkage [20,22]. We were unable to detect linkage.

Abbreviations
CD, celiac disease; HLOD, Heterogeneity LOD: NPL, non-parametric linkage; TCR, T cell receptor

Competing interests
None declared.

Acknowledgements
We would like to thank the families for participating in our study. We would like to thank Thao Tran, Kim Nguyen, Michael Hoffman, and Ted Taylor for technical assistance in the laboratory and Jeff Black for family ascertainment. This work was funded by grant R01-DK50678 from the National Institutes of Health. We gratefully acknowledge the support of the NHLBI Mammalian Genotyping Service for providing some of the genotyping.

References
1. Not T, Horvath K, Hill I, Partanen J, Hammed A, Magazuz G, Fasano A: Celiac disease risk in the USA: high prevalence of antiecdinomysiosis antibodies in healthy blood donors. Scand J of Gastroenterol 1998, 33:494-498
2. Greco L: Epidemiology of celiac disease. In Coeliac Disease, Proceedings of the International Symposium on Coeliac Disease (M. Maki, P. Collin, and J.K. Visakorpi, Editors). Tampere, Finland, Institute of Medical Technology, University of Tampere 1999:7-14
3. Ferreira M, Davies SL, Butler M, Scott D, Clark M, Kumar P: Endomysiosis antibody: is it the best screening test for celiac disease? Gut 1992, 33(12):1633-1637
4. Grodzinsky E, Hed J, Skogh T: IgA antiendomysiosis antibodies have a high predictive value for celiac disease in asymptomatic patients. Allergy 1994, 49:593-597
5. Russo PA, Chartrand LJ, Seidman E: Comparative analysis of serologic screening tests for the initial diagnosis of celiac disease. Pediatrics 1999, 104(1 Pt 1):75-78
6. Sategna Guidetti C, Pulitano R, Grosso S, Ferfoglia G: Serum IgA antidiomysiosis antibody titers as a marker of intestinal involvement and diet compliance in adult celiac sprue. J Cln Gastroenterol 1993, 17:123-127
7. Unsworth F: Serologic diagnosis of gluten sensitive enteropathy. J Clin Path 1996, 49:704-711
8. Vaidimantron T, Fransen L, Grodzinsky E, Skogh T, Strom M: Is small bowel biopsy necessary in adults with suspected celiac disease and IgA anti-endomysiosis antibodies? 100% positive predictive value for celiac disease in adults. Digestive Disease and Science 1996, 41:83-87
9. Volta U, Molinario M, Fusconio M, Cassini F, Bianchi FB: IgA antidiomysiosis antibody test: A step forward in celiac disease screening. Digestive Diseases and Science 1991, 36:752-756
10. Volta U, Molinario N, De Franchis R, Forzenigo L, Landoni M, Fratangelo D, Bianchi FB: Correlation between IgA antidiomysiosis antibodies and subtotal villous atrophy in dermatitis herpetiformis. J Cln Gastroenterol 1992, 14(4):298-301
11. Dieterich W, Laag E, Shopper H, Volta U, Fergusson A, Gillett H, Riecken E, Schuppans D: Autoantibodies to tissue transglutaminase as predictors of celiac disease. Gastroenterol 1998, 115(1317-1321
12. Sulkavan S, Halttunen T, Laurila K, Kolho K, Korporsaz-Sabori I, Sarnesto A, Savihaft E, Collin P, Maki M: Tissue transglutaminase antitbody enzyme-linked immunosorbant assay in detecting celiac disease. Gastroenterol 1998, 115(3):122-128
13. Kaukinen K, Maki M, Partanen J, Sievanen H, Collin P: Celiac disease without villous atrophy: revision of criteria called for. Dig Dis Sci 2001, 46(4):879-887
14. Solidi LM: HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. Gastroenterol 1993, 105:910-922
15. Bevan S, Papas S, Braegger CP, Busch A, O’Donoghue D, Falth-Magnusson K, Fergusson A, Godkin A, Hegberg L, Holmes G, et al: Contribution of the MHC region to the familial risk of celiac disease. J Med Genet. 1997, 34:687-690
16. Lewis C, Book L, Black J, Sawitzke A, Savilahti E, Collin P, Maki M: Correlation between IgA antiendomysiosis antibodies and IgA antiendomysiosis antibodies? 100% positive predictive value for celiac disease in adults. Digestive Disease and Science 1996, 41:83-87
17. Petronzelli F, Bonamico M, Ferrante P, Grillo R, Morb A, Mariani P, Apollonio I, Gemme G, Mazzilli MC: Genetic contribution of the HLA region to the familial clustering of celiac disease. Ann Hum Genet 1997, 61:307-317
18. Risch N: Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 1987, 40:1-14
19. Strober W: Gluten-Sensitive Enteropathy. In Genetic Basis of Common Diseases (R. King, Rotter, JI, Motulsky, AG, Editor). New York, New York Oxford Univer Press 1992:279-304
20. Brett PM, Yaniokuto J, Morris MA, Rosen Bronson S, Mathew C, Curtis D, Ciclitira PJ: A pedigree-based linkage study of celiac disease: failure to replicate previous positive findings. Ann Hum Genet 1998, 62:25-32
21. Dyllali-Saah S, Schmitz J, Harfouche-Hemoud E, Mougenot JF, Bach JF, Caillier-Zucman S: CTLA-4 gene polymorphism is associated with predisposition to celiac disease. Gut 1998, 43(2):187-189
22. Greco L, Corazza G, Babron MC, Clot F, Fulghignoni-Lataud MC, Per- copo S, Zavattari P, Bouguerra F, Dib C, Tosi R, et al: Genome search in celiac disease. Am J Hum Genet 1998, 62(3):669-675
23. Greco L, Babron MC, Corazza GR, Percopo S, Sica R, Clot F, Ful- ghignoni-Lataud MC, Zavattari P, Momigliano-Richiardi P, Casari G, et al: Existence of a genetic risk factor on chromosome 5q in Italian coeliac disease families. Ann Hum Genet 2001, 65:35-41
24. Holopainen P, Arvas M, Sixtonen P, Mustalhti K, Collin P, Maki M, Partanen J: CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. *Tissue Antigens* 1999, 53(5):470-475.

25. Houlston RS, Ford D: Genetics of coeliac disease. *Q J Med* 1996, 89:737-743.

26. King AL, Yiannakou JY, Brett PM, Curtis D, Morris MA, Dearlove AM, Rhodes M, Rosen-Bronson S, Mathew C, Ellis HJ, et al: A genome-wide family-based linkage study of coeliac disease. *Ann Hum Genet* 2000, 64:479-490.

27. King AL, Fraser JS, Moodie SJ, Curtis D, Dearlove AM, Ellis HJ, Rosen-Bronson S, Ciclitira PJ: 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.

28. Roschmann E, Wienker TF, Gerok W, Volk BA: T-cell receptor variable genes and genetic susceptibility to celiac disease: an association and linkage study. *Gastroenterology* 1993, 105(6):1790-1796.

29. Zhong F, McCombs CC, Olson JM, Elston RC, Stevens FM, McCarthy CF, Michalski JP: An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. *Nat Genet* 1996, 14(3):329-333.

30. Lerner A, Kumar V, Iancu T: Immunological diagnosis of childhood coeliac disease: comparison between antigliadin, antireticulin and antiendomysial antibodies. *Clin Exp Immunol* 1994, 95:78-82.

31. Feolo M, Fuller TC, Taylor M, Zone JJ, Neuhausen SL: A strategy for high throughpt HLA-DQ typing. *J Immunol Methods* 2001, 258:65-71.

32. Cottingham RW Jr, Idury RM, Schaffer AA: Faster sequential genetic linkage computations. *Am J Hum Genet* 1993, 53(1):252-263.

33. Lathrop GM, Lalouel JM, Julier C, Ott J: Multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984, 81(11):3443-3446.

34. Lathrop GM, Lalouel JM, Julier C, Ott J: Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. *Am J Hum Genet* 1985, 37(3):482-498.

35. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996, 58(6):1347-1363.

36. Ott J: Linkage probability and its approximate confidence interval under possible heterogeneity. *Genet Epidemiol Suppl* 1986, 1:251-257.

37. Ott J: *Analysis of Human Genetic Linkage*. 3rd ed. Baltimore, Johns Hopkins University Press 1999.

38. Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J: Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* 1986, 42(2):393-399.

39. Greenberg DA, Hodge SE, Rotter JI: Evidence for recessive and against dominant inheritance at the HLA-"linked" locus in celiac disease. *Am J Hum Genet* 1982, 34(2):263-277.

40. Risch N, Claus E, Guiffra L: Linkage and mode of inheritance in complex traits. In *Multipoint mapping and linkage based upon affected pedigree members*. Genetic Analysis Workshop 6. Progress in Clinical and Biological Research. (R.C. Elston, M.A. Spence, S.E. Hodge, and J.W. MacCluer, Editors). New York, Alan R. Liss 1989:183-189.

41. Yiannakou JY, Brett PM, Morris MA, Curtis D, Mathew C, Vaughan R, Rosen-Bronson S, Ciclitira PJ: Family linkage study of the T-cell receptor genes in coeliac disease. *Ital J Gastroenterol Hepatol* 1999, 31(3):198-201.

42. Houlston R, Tomlinson I, Ford D, Seal S, Marosy A, Ferguson A, Holmes G, Hosie K, Howdle P, Jewell D, et al: Linkage analysis of candidate regions for coeliac disease genes. *Hum Mol Genet* 1997, 6:1335-1339.