Genetics of celiac disease and its diagnostic value
Romanos, Jihane

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Six new celiac disease loci replicated in an Italian population confirm association with celiac disease

Jihane Romanos 1,*, Donatella Barisani 2,*, Gosia Trynka 1, Alexandra Zhernakova 3, Maria Teresa Bardella 4,5, Cisca Wijmenga 1.

1 Department of Genetics, University Medical Center Groningen, University of Groningen, the Netherlands; 2 Department of Experimental Medicine, University of Milano Bicocca, Italy; 3 Complex Genetics Section, Department of Medical Genetics, University Medical Center Utrecht, the Netherlands; 4 Fondazione IRCCS Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan, Italy; 5 Department of Medical Sciences, University of Milan, Italy.

* These authors contributed equally to this work.

Abstract

**Background and aims:** The first genome wide association study on coeliac disease (CD) and its follow-up have identified eight new loci that contribute significantly towards CD risk. Seven of these loci contain genes controlling adaptive immune responses, including *IL2/IL21* (4q27), *RGS1* (1q31), *IL18RAP* (2q11–2q12), *CCR3* (3p21), *IL12A* (3q25–3q26), *TAGAP* (6q25) and *SH2B3* (12q24). **Methods:** We selected the nine most associated single nucleotide polymorphisms to tag the eight new loci in an Italian cohort comprising 538 CD patients and 593 healthy controls. **Results:** Common variation in *IL2/IL21*, *RGS1*, *IL12A/ SCHIP* and *SH2B3* was associated with susceptibility to CD in our Italian cohort. The LPP and TAGAP regions also showed moderate association, whereas there was no association with *CCR3* and *IL18RAP*. **Conclusion:** This is the first replication study of six of the eight new CD loci; it is also the first CD association study in a southern European cohort. Our results may imply there is a genuine population difference across Europe regarding the loci contributing to CD.

**Keywords:** coeliac disease, replication, association study, meta-analysis.
Introduction

Coeliac disease (CD) is a chronic disorder of the small intestine, resulting from an aberrant cellular response to gluten peptides; it affects as much as 1% of the European population. The only treatment is a lifelong gluten-free diet. In the past decade, tremendous progress has been achieved in unravelling the genetic aetiology of CD. Twin and family based studies clearly show a strong genetic component to CD development. The clearly identified genetic risk factors for this disease are the HLA-DQ2 and HLA-DQ8 molecules. These are estimated to explain ~40% of the heritability of CD. The other 60% of the genetic susceptibility to CD is shared between an unknown number of non-HLA genes, each of which is estimated to contribute only a small risk effect. Linkage screens and candidate gene studies have led to the discovery of several susceptibility loci and genes, such as the CELIAC2 locus (5q), MYO9B and CTLA4.

The first genome wide association study (GWAS) in CD was recently performed in 778 CD cases and 1422 population controls from the UK. The only locus other than the HLA region showing genome wide significance was 4q27, a ~500 kb block of linkage disequilibrium (LD) containing the IL2 and IL21 genes. We established independent replication of single nucleotide polymorphisms (SNPs) from the IL2/IL21 region in both Dutch and Irish cohorts of coeliac patients and healthy controls. Moreover, the same region was found to be associated with type 1 diabetes and rheumatoid arthritis, suggesting it is a common autoimmune locus. Both IL2 and IL21 molecules are widely expressed cytokines important for T cell maturation and proliferation, and they are therefore attractive candidates for CD pathogenesis.

In a more extensive follow-up of 1020 top GWAS associated single nucleotide polymorphisms (SNPs) in several independent cohorts from the UK, Dutch and Irish populations, Hunt et al identified seven new risk regions that meet a genome wide significance threshold in 7238 samples (p value overall ,561027). Six of these new CD loci contain genes controlling adaptive immune responses, including RGS1 (1q31), IL18RAP (2q11–2q12), CCR3 (3p21), IL12A (3q25–3q26), TAGAP (6q25) and SH2B3 (12q24). The seventh associated locus is located on 3q28 and harbours the LPP gene, which might play a role in maintaining cell adhesion and motility. Three of these loci have also been associated with other inflammatory and autoimmune disorders: the CCR3 and SH2B3 loci with type 1 diabetes and the IL18RAP locus with Crohn’s disease.

We set out to replicate the associations found with the seven new loci and the IL2/IL21 locus to CD in an Italian CD cohort.
METHODS

Subjects and controls

DNA isolated from whole blood was available from 538 patients diagnosed by a referral centre for CD (Centro per la prevenzione e diagnosi della malattia celiaca, Fondazione IRCCS Ospedale Maggiore Policlinico) and from 593 healthy controls from the north of Italy. The average age of onset was 24.7 years (range 1–78 years). All the affected individuals were diagnosed according to the revised ESPGHAN criteria showing a Marsh III lesion. In addition, patients' serum samples tested positive for both anti-transglutaminase and anti-endomysium antibodies. Only 1.3% of the affected individuals had no HLA-DQ2 and/or HLA-DQ8 risk alleles, which is in accordance with published data. Written informed consent was obtained from all individuals before enrolment in the study. The study was approved by the ethics committee of the Fondazione IRCCS Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan, Italy.

Genotyping

In order to tag the eight loci, we selected the nine most associated SNPs reported by Hunt et al. For the IL2/Il21 locus, the most associated SNP rs13119723 was discarded as this SNP showed bad clustering. Therefore, we selected the second most associated SNP rs6822844 to tag the IL2/IL21 locus. For the IL12A/SCHIP locus, we genotyped two SNPs, rs17810546 and rs9811792, since they were reported to be independently associated. SNPs were genotyped using TaqMan probes and primers, using assays developed by Applied Biosystems, and an ABI 7900HT system (Applied Biosystems, Nieuwerkerk a/d IJssel, the Netherlands). Genotyping was performed following the manufacturer’s specifications. DNA samples were processed in 384 well plates and each plate with patients' and control DNA contained eight negative controls and 16 genotyping controls (four duplicates of four different samples obtained from the Centre d’Etude du Polymorphisme Humain (CEPH)). There was no discordance in the genotypes of any of the CEPH samples. Laboratory staff were blind to the disease status of each sample.

Statistical analysis

The genotype frequencies were tested for Hardy–Weinberg equilibrium (HWE) with a value of p<0.05 considered as not being in HWE. Allele frequencies were determined in patients and controls. Difference in allele distribution between patients and controls and association analysis were performed using two tailed x2 analysis while meta-analysis of our Italian cohort, the UK GWAS of van Heel et al and the Irish, Dutch and UK2 cohorts of Hunt et al was performed using the Maentel–Haenszel method. Odds ratios (OR) and confidence intervals (CI) were calculated using Woolf’s method with Haldane’s correction. Power calculations
were performed using the genetic power calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) assuming allele frequencies of 0.2 and 0.3.

RESULTS

Replicating genetic findings in several populations is an important step in establishing a genetic effect on disease predisposition. We therefore genotyped nine associated SNPs tagging the eight CD susceptibility loci identified by Van Heel et al and Hunt et al. Our study had ~80% power to detect an odds ratio of 1.45, while it had ~28% power to detect an odds ratio of 1.2. We observed association for six of the loci in our Italian cohort. Table 1 summarises the genotyping results and case–control association analysis at the single SNP level. All SNPs were in HWE in our control population (data not shown).

The first GWAS in CD in a UK cohort identified 4q27 region as a susceptibility region for CD. We genotyped SNP rs6822844 which was the most associated SNP identified by meta-analysis in the first GWAS and the second most associated one reported in the follow-up. We saw a decrease in frequency of the rs6822844*A allele in Italian cases (8.1%) compared to controls (10.9%); this association was significant (p = 0.025) and in the same direction as described earlier (OR 0.72, 95% CI 0.54 to 0.96).

In our cohort, the most associated locus was located on chromosome 3q25–3q26. Two SNPs were tested in this block due to independent
Table 2. Differences among five celiac disease populations.

<table>
<thead>
<tr>
<th>Population (number of cases)</th>
<th>UK GWAS (778)</th>
<th>UK2 (719)</th>
<th>Irish (416)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP</td>
<td>MAF cases/controls</td>
<td>p-value</td>
<td>MAF cases/controls</td>
</tr>
<tr>
<td>RGS1 rs2816316</td>
<td>0.136/0.174</td>
<td>1.16E-03</td>
<td>0.129/0.188</td>
</tr>
<tr>
<td>IL18RAP rs917997</td>
<td>0.267/0.215</td>
<td>9.06E-05</td>
<td>0.245/0.208</td>
</tr>
<tr>
<td>CCR3 rs6441961</td>
<td>0.341/0.297</td>
<td>2.84E-03</td>
<td>0.342/0.302</td>
</tr>
<tr>
<td>IL12A, SCHIP1 rs17810546</td>
<td>0.162/0.123</td>
<td>3.39E-04</td>
<td>0.146/0.127</td>
</tr>
<tr>
<td>IL12A rs9811792</td>
<td>0.486/0.439</td>
<td>2.78E-03</td>
<td>0.483/0.440</td>
</tr>
<tr>
<td>LPP rs1464510</td>
<td>0.52/0.457</td>
<td>7.65E-05</td>
<td>0.516/0.446</td>
</tr>
<tr>
<td>IL2/IL21 rs822844</td>
<td>0.126/0.179</td>
<td>4.80E-06</td>
<td>0.138/0.175</td>
</tr>
<tr>
<td>TAGAP rs1738074</td>
<td>0.472/0.422</td>
<td>1.56E-03</td>
<td>0.459/0.428</td>
</tr>
<tr>
<td>SH2B3 rs3184504</td>
<td>0.541/0.489</td>
<td>1.08E-03</td>
<td>0.522/0.472</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Population (number of cases)</th>
<th>Dutch (508)</th>
<th>Italian (538)</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP</td>
<td>MAF cases/controls</td>
<td>p-value</td>
<td>MAF cases/controls</td>
</tr>
<tr>
<td>RGS1 rs2816316</td>
<td>0.141/0.189</td>
<td>1.09E-03</td>
<td>0.128/0.166</td>
</tr>
<tr>
<td>IL18RAP rs917997</td>
<td>0.293/0.215</td>
<td>3.61E-06</td>
<td>0.244/0.227</td>
</tr>
<tr>
<td>CCR3 rs6441961</td>
<td>0.38/0.322</td>
<td>1.95E-03</td>
<td>0.372/0.368</td>
</tr>
<tr>
<td>IL12A, SCHIP1 rs17810546</td>
<td>0.171/0.124</td>
<td>6.37E-04</td>
<td>0.109/0.067</td>
</tr>
<tr>
<td>IL12A rs9811792</td>
<td>0.497/0.439</td>
<td>3.17E-03</td>
<td>0.467/0.421</td>
</tr>
<tr>
<td>LPP rs1464510</td>
<td>0.52/0.457</td>
<td>2.93E-03</td>
<td>0.519/0.474</td>
</tr>
<tr>
<td>IL2/IL21 rs822844</td>
<td>0.13/0.186</td>
<td>1.14E-04</td>
<td>0.081/0.109</td>
</tr>
<tr>
<td>TAGAP rs1738074</td>
<td>0.46/0.395</td>
<td>8.18E-04</td>
<td>0.454/0.412</td>
</tr>
<tr>
<td>SH2B3 rs3184504</td>
<td>0.528/0.479</td>
<td>0.013</td>
<td>0.554/0.494</td>
</tr>
</tbody>
</table>

CHR chromosome; MAF minor allele frequency; OR odds ratio; CI confidence interval; bold indicates p-value < 0.05.
The second SNP in the same block, rs9811792, showed a moderate association ($p = 0.031; \text{OR} = 1.21, 95\% \text{CI} 1.02$ to $1.43$). These two SNPs may also represent an independent association signal ($D' = 0.97; r^2 = 0.113$) in our cohort. This region harbours two potentially interesting genes, $IL12A$ (interleukin-12A) and $SCHIP1$ (schwannomin interacting protein 1). The second most associated SNP rs3184504 ($p = 5.04E-03; \text{OR} = 1.27, 95\% \text{CI} 1.07$ to $1.50$) mapped on chromosome 12q24, in the vicinity of $SH2B3$ and $ATNX2$ genes.

SNP rs2816316 was the most significant SNP outside the HLA and IL2/IL21 loci identified by Hunt et al.\(^8\) It is located on chromosome 1q31, in a ~70 kb LD block containing the $RGS1$ gene (regulator of G-protein signalling 1). We also found association for this SNP in our cohort ($p = 0.012; \text{OR} = 0.74, 95\% \text{CI} 0.58$ to $0.94$). Moderate association ($p = 0.035; \text{OR} = 1.2, 95\% \text{CI} 1.20$ to $1.42$), consistent with previous findings, was found for SNP rs1464510, located on 3q28, in a ~70 kb LD block harbouring the $LPP$ gene. SNP rs1738074, located on chromosome 6q25, showed a trend towards association with CD in our cohort ($p = 0.05$). This SNP is a ~200 kb LD block containing $TAGAP$ (T cell activation GTPase activating protein). This 6q25 region was also found to be linked to CD in a large Dutch family.\(^13\) All associations were observed with the same allele as in the original study.\(^8\)

We saw no association for two SNPs (rs917997 and rs6441961) which were located on chromosome 2q11–2q12 (IL18RAP locus) and 3p21 (CCR3 locus), respectively (table 1).

**DISCUSSION**

In the last decade, our understanding of CD pathogenesis was mainly based on the binding of HLA-DQ2/DQ8 to gluten peptides, the role of tissue transglutaminase, and the identification of immunological dominant T cell epitopes. However, advances in genetics now allow us to identify novel genes involved in the susceptibility to CD, thereby helping us to understand the pathogenesis of this disorder. We have performed the first replication...
of eight new loci identified in the first GWAS performed in CD.\(^6,8\) We found a positive association for six of these loci: IL2/IL21, RGS1, IL12A/SCHIP, LPP, TAGAP and SH2B3 (table 1). These loci harbour candidate genes involved mainly in the Th1 pathway: IL2 and IL21 are important in T cell activation, while RGS1 regulates chemokine receptors’ signalling and is involved in B cell activation and proliferation. Not much is known about the SCHIP1 gene but the IL12A gene, located in the same LD block, encodes the IL12p35 of IL12 subunit, which is important for T cells and natural killer cells, both of which are involved in the Th1 pathway. LPP shows a very high expression in the small intestine and may play a structural role in maintaining cell shape and motility at sites of cell adhesion. The TAGAP gene is interesting since it is expressed in activated T cells and has a Rho-GAP domain similar to MYO9B, another CD associated gene.\(^5\) The SH2B3 gene is a good candidate for CD since it is expressed in the small intestine, mainly in monocytes and dendritic cells and to a lesser extent in resting B, T and natural killer cells. Moreover, the SNP that we have tested could be a causal variant since it is a non-synonymous SNP leading to an amino acid change R262W in an important domain of the protein. Fine mapping and deep sequencing of these regions, and functional studies, are needed to identify the true causal genes and their role in the disease process.

We were unable to detect association between the CCR3 locus or the IL18RAP locus and CD in our Italian cohort. This might be due to clinical heterogeneity although this seems highly unlikely given that our Italian cohort is a mixture of adult and paediatric patients. It might also be due to the power being too low to detect an odds ratio of 1.2, or even genetic heterogeneity among populations. For the SNP in the CCR3 gene, there was no significant difference between the frequencies of the minor allele in patients compared to controls (37.2% and 36.8%, respectively). We looked at the separate results of the four populations analysed for the GWAS and noticed that this region was only associated in the UK GWAS, UK2 and Dutch samples, whereas in the Irish cohort, the frequency between patients and controls was not significantly different (table 2). Similarly, the IL18RAP locus was not associated in the Irish cohort.

These results support the hypothesis of risk genes in complex disorders being population specific. Genetic heterogeneity is recognised for the CD associated HLA risk alleles within Europe. In southern Europe, individuals carrying DQ2.5 in trans are more prevalent than in the north, where individuals carry DQ2.5 in cis more frequently.\(^14\) In addition, risk allele HLA-DQ8 is more frequent in southern Europe and accounts for 6–10% of the CD patients there.\(^15\) Differences within European populations due to regional founder effects were also suggested for the inflammatory bowel disease associated genes, NOD2 and DLG5,\(^16,17\) while allele frequencies for NOD2 risk alleles were reported to vary significantly between European populations.\(^18\) Another reason for discrepancies among populations could be the complex interaction between marker allele frequencies and founder mutations. Since the linkage disequilibrium varies between distinct populations, the causative variant could be in less LD with the tested SNP. Fine mapping with a dense SNP set is necessary to exclude these
genes as disease causing variants in the Italian patients.

In conclusion, our study confirms the association of six of the eight new loci with CD and may point to heterogeneity among European populations.

**Acknowledgements:** We thank all the patients and controls who participated in this study. We thank Agata Szperl, Eleonora AM Festen and Cleo van Diemen for their help in the laboratory and Jackie Senior for critically reading the manuscript.

**Funding:** The study was supported by grants from the Celiac Disease Consortium (an innovative cluster approved by the Netherlands Genomics Initiative and partly funded by the Dutch Government, grant BSIK03009 to CW) and KP6 EU grant 036383 (PREVENTCD).

**Competing Interests:** None. Patient consent: Obtained.
REFERENCES


