Gender-stratified analysis of DLG5 R30Q in 4707 patients with Crohn disease and 4973 controls from 12 Caucasian cohorts


ABSTRACT

Background: DLG5 p.R30Q has been reported to be associated with Crohn disease (CD), but this association has not been replicated in most studies. A recent analysis of gender-stratified data from two case–control studies and two population cohorts found an association of DLG5 30Q with increased risk of CD in men but not in women and found differences between 30Q population frequencies for males and females. Male–female differences in population allele frequencies and male-specific risk could explain the difficulty in replicating the association with CD.

Methods: DLG5 R30Q genotype data were collected for patients with CD and controls from 11 studies that did not include gender-stratified allele counts in their published reports and tested for male–female frequency differences in controls and for case–control frequency differences in men and in women.

Results: The data showed no male–female allele frequency differences in controls. An exact conditional test gave marginal evidence that 30Q is associated with decreased risk of CD in women (p = 0.049, OR = 0.87, 95% CI 0.77 to 1.00). There was also a trend towards reduced 30Q frequencies in male patients with CD compared with male controls, but this was not significant at the 0.05 level (p = 0.058, OR = 0.87, 95% CI 0.74 to 1.01). When data from this study were combined with previously published, gender-stratified data, the 30Q allele was found to be associated with decreased risk of CD in women (p = 0.010, OR = 0.86, 95% CI 0.76 to 0.97), but not in men.

Conclusion: DLG5 30Q is associated with a small reduction in risk of CD in women.

Inflammatory bowel disease (IBD; OMIM 266600) comprises Crohn disease (CD; OMIM 266600) and ulcerative colitis (UC; OMIM 191390). Epidemiological and genetic studies have demonstrated that genes play an important role in the pathogenesis of IBD. Genetic variants associated with CD have been identified and replicated in the CARD15/NOD2 gene (OMIM 608956) and in the chromosome 5q31 region (OMIM 606348), and recently, the first generation of whole-genome association studies have identified multiple new susceptibility variants. In 2004, Stoll et al described an association between IBD and DLG5 variants (DLG5: Drosophila Discs Large Homolog 5 (OMIM 604090)). In particular, p.R30Q (c.89G>A, rs1248696, previously reported as c.115G>A) was found to be associated with IBD in a German family cohort. Stoll et al replicated the 30Q association with CD in a European case–control cohort and found a non-significant overtransmission of the 30Q allele in patients with CD in a German/UK family cohort. Subsequently, Daly et al replicated the association of 30Q with CD in one of two case–control cohorts and in a family cohort. However, subsequent studies have failed to replicate the association of DLG5 30Q with CD.

Recently, Friedrichs et al examined gender-stratified data from the case–control cohorts showing association in the initial two DLG5 studies, and found that the DLG5 30Q allele was a risk factor for CD in men but not in women, and that the 30Q allele frequency was significantly lower in male compared with female controls (0.05 vs 0.11). Friedrichs et al typed genotyped R30Q in two additional control samples: a sample of 190 male and 271 female participants from Germany and in a newborn sample of 501 male and 299 female infants from Wisconsin, USA. For all three samples (combined control cohort, the German population sample and the Wisconsin population sample) the DLG5 30Q allele frequency was lower in male than in female subjects and the differences were significant at the 0.05 level. The presence of the male–female allele frequency differences in the newborn sample suggests that the allele frequency differences are due to prenatal processes rather than selection later in life and Friedrichs et al proposed that the allele frequency differences were a consequence of a gender-dependent transmission ratio distortion.

Differences in male and female allele frequencies combined with male–specific risk could explain the difficulties that have been encountered in replicating the association of 30Q with CD, as the estimated effect size would depend on the proportion of male cases and controls in the study. If the minor allele increases risk in men but not in
women and if the frequency in male patients with CD is less than the minor allele frequency in females in a population, then the 30Q allele would be a risk allele in a study of female patients with CD and male controls. Replication of association could be further complicated by differences in effect sizes between populations, owing to interaction of R30Q with other genetic and environmental factors that vary between populations.

Most studies investigating R30Q were published prior to the report of Friedrichs et al., which first described differences in male and female R30Q population frequencies and male–specific risk. Consequently, in most DLG5 studies, genotype data stratified by gender have not been reported. In this study, we analysed male and female R30Q allele counts from 11 previous DLG5 case–control studies that did not report gender-stratified data. We tested R30Q for association with CD in men and in women, and tested for differences in male and female control allele frequencies.

METHODS

Study participants

Published association studies of DLG5 were identified through PubMed using a search with the keyword “DLG5”. All association studies had sampled from Caucasian populations. The limited available evidence indicates that the DLG5 R30Q polymorphism is absent in populations from Asia and Africa. We identified 12 case–control association studies of DLG5 R30Q and IBD in Caucasian populations, which had not reported gender-stratified R30Q allele counts or frequencies. One of these 12 studies had a relatively high proportion of non-Caucasian participants. Gender-stratified data from the remaining 11 DLG5 case–control studies were included in the current study and represent 4707 patients with CD and 4973 controls from 12 case–control cohorts.

Allele counts reported in the current study may be greater or less than allele counts in the original study if gender was not available for some participants or if additional individuals from the same population were genotyped for DLG5 R30Q subsequent to the initial study. Our dataset does not include data from the Ashkenazi Jewish participants in the study of Newman et al., nor the CD and control patients in the study of Cucchiara et al., who had previously been included in the study of Friedrich et al., local ethics committee approval was obtained for each study that contributed gender-stratified data to the current study.

All mutations are numbered at the cDNA level, indicated by a “+” before the number. Position +1 corresponds to the A of the ATG translation initiation codon located at nucleotide 401 in the NM_004747.5 DLG5 mRNA reference sequence.

Statistical analysis

Statistical analysis was performed using R V.2.2.0 software. For comparison of male and female R30Q population frequencies, allelic odds ratios (ORs) < 1.0 correspond to a lower 30Q allele frequency in male than in female subjects. For case–control analysis, allelic ORs < 1.0 correspond to a lower 30Q allele frequency in patients with CD than in controls. Allelic ORs with 95% confidence intervals (CIs) were calculated under the assumption of Hardy–Weinberg equilibrium both in cases and in controls.

The combined evidence for association was assessed with an exact conditional test (conditional upon cohort). The exact test was used in preference to the Cochrane–Mantel–Haenszel test because the exact test does not use an asymptotic approximation. Heterogeneity between studies was assessed using Cochran’s Q statistic, which has a χ² distribution under the null hypothesis of no heterogeneity between studies, and using the I² statistic, which is an estimate of the proportion of the total variation that is due to heterogeneity between studies. When there was evidence for heterogeneity between studies, we performed a meta-analysis under a random effects model to estimate the mean OR. Under a random effects model, the variance of the allelic OR for each study population is assumed to be drawn from a random distribution with variance δ and the variance of the estimated allelic OR is the sum of the sampling variance and δ. Estimates of mean OR under a random effects model are calculated by weighting each study by the reciprocal of the sampling variance of the OR to account for differences in sample sizes.

RESULTS

Male and female DLG5 R30Q allele counts for 12 case–control cohorts are given in table 1. Genotype counts in CD cases and in controls were tested for departures from Hardy–Weinberg equilibrium using an exact test. None of the tests for Hardy–Weinberg equilibrium were significant, indicating that the samples were in equilibrium.

Table 1 Male and female DLG5 R30Q allele frequency data reported in this study

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Male patients with CD</th>
<th>Male controls</th>
<th>OR (95% CI)</th>
<th>Female patients with CD</th>
<th>Female controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n 30Q %</td>
<td>n 30Q %</td>
<td>OR (95% CI)</td>
<td>n 30Q %</td>
<td>OR (95% CI)</td>
<td>n 30Q %</td>
</tr>
<tr>
<td>Browning, 2007*</td>
<td>New Zealand</td>
<td>139 37 13.3</td>
<td>173 32 9.2</td>
<td>1.51 (0.91 to 2.49)</td>
<td>250 60 12.0</td>
<td>237 50 10.5</td>
</tr>
<tr>
<td>Cucchiara, 2007*</td>
<td>Italian</td>
<td>92 5 2.7</td>
<td>153 36 11.8</td>
<td>0.21 (0.08 to 0.54)</td>
<td>65 7 5.4</td>
<td>101 16 7.9</td>
</tr>
<tr>
<td>Pearce, 2007*</td>
<td>British</td>
<td>270 40 7.4</td>
<td>178 37 10.4</td>
<td>0.69 (0.43 to 1.10)</td>
<td>360 82 11.4</td>
<td>278 47 8.5</td>
</tr>
<tr>
<td>Ferraris, 2006*</td>
<td>Italian</td>
<td>69 14 10.1</td>
<td>63 13 10.3</td>
<td>0.98 (0.44 to 2.18)</td>
<td>85 9 6.9</td>
<td>47 8 8.5</td>
</tr>
<tr>
<td>Lakatos, 2006*</td>
<td>Hungarian</td>
<td>309 71 11.5</td>
<td>72 26 18.1</td>
<td>0.59 (0.36 to 0.96)</td>
<td>330 73 11.1</td>
<td>78 19 12.2</td>
</tr>
<tr>
<td>Tremelling, 2006*</td>
<td>British</td>
<td>168 39 11.6</td>
<td>330 75 11.4</td>
<td>1.02 (0.68 to 1.55)</td>
<td>326 64 9.8</td>
<td>426 100 11.7</td>
</tr>
<tr>
<td>Newman, 2006*</td>
<td>Canadian</td>
<td>141 27 9.6</td>
<td>171 33 9.6</td>
<td>0.99 (0.58 to 1.69)</td>
<td>172 26 7.6</td>
<td>512 114 11.1</td>
</tr>
<tr>
<td>Büning, 2006*</td>
<td>German</td>
<td>90 14 7.8</td>
<td>212 42 9.9</td>
<td>0.77 (0.41 to 1.44)</td>
<td>148 20 6.8</td>
<td>208 40 9.6</td>
</tr>
<tr>
<td>Büning, 2006*</td>
<td>Hungarian</td>
<td>65 11 8.5</td>
<td>95 19 10.0</td>
<td>0.83 (0.38 to 1.81)</td>
<td>82 12 7.3</td>
<td>110 27 12.3</td>
</tr>
<tr>
<td>Torók, 2005*</td>
<td>South German</td>
<td>225 45 10.0</td>
<td>594 128 10.8</td>
<td>0.92 (0.64 to 1.32)</td>
<td>252 50 9.9</td>
<td>378 82 10.8</td>
</tr>
<tr>
<td>Noble, 2005*</td>
<td>Scottish</td>
<td>142 28 9.9</td>
<td>128 33 12.9</td>
<td>0.74 (0.43 to 1.26)</td>
<td>214 45 10.5</td>
<td>128 34 13.3</td>
</tr>
<tr>
<td>Vermeiren, 2005*</td>
<td>Flemish</td>
<td>255 62 10.5</td>
<td>137 24 8.8</td>
<td>1.22 (0.75 to 2.01)</td>
<td>438 90 10.3</td>
<td>164 41 12.5</td>
</tr>
<tr>
<td>Cumulative report</td>
<td>— —</td>
<td>2306 —</td>
<td>— —</td>
<td>0.87 (0.74 to 1.01)</td>
<td>2702 — —</td>
<td>2667 — —</td>
</tr>
</tbody>
</table>

Study, population, case and control sample sizes, 30Q allele counts, 30Q allele frequencies and ORs with 95% CIs are given. The OR and 95% CI for the cumulative data were calculated using an exact conditional test.

J Med Genet 2008;45:36–42. doi:10.1136/jmg.2007.050773
The observed difference in allele frequency between male and female controls in our data did not reach statistical significance (exact test OR = 0.99, 95% CI 0.87 to 1.13). There was no evidence of heterogeneity of OR between populations (Q = 8.83, 11 degrees of freedom (df), p = 0.64, I^2 = 0.00). In our female case–control data, there was a lower allele frequency in female patients with CD compared with female controls (exact test p = 0.049, OR = 0.87, 95% CI 0.77 to 1.00) and there was no evidence of heterogeneity of OR between female populations (Q = 12.7, 11 df, p = 0.32, I^2 = 0.13).

In our male case–control data, there was marginal evidence of heterogeneity of OR (Q = 19.9, 11 df, p = 0.047, I^2 = 0.45), so we performed analysis using a random effects model, which allows for heterogeneity of ORs between populations, in addition to analysis using an exact test, which assumes homogeneity of effect between populations. For the male data, the exact test was not significant at the 0.05 level (p = 0.058, OR = 0.87, 95% CI 0.74 to 1.01) and the 95% CI for the mean OR under a random effects model included 1.0 (mean OR = 0.86, 95% CI 0.70 to 1.06).

Combining our male and female data, there was marginal evidence for heterogeneity of ORs (Q = 20.0, 11 df, p = 0.046, I^2 = 0.45) and our combined male and female data supported an association of 30Q with reduced risk of CD (random effects mean OR = 0.85, 95% CI 0.74 to 0.98).

In summary, we found no male–female differences in DLG5 R30Q frequencies among controls, but our data gives marginal evidence that the 30Q was associated with reduced risk of CD in both men and women.

We next compared our findings with previously published gender-stratified R30Q data, which include gender-stratified data from Stoll et al. and Daly et al. Population, sample sizes, ORs and CIs are shown in the figures for male controls and female controls (fig 1), for female patients with CD and female controls (fig 2) and for male patients with CD and male controls (fig 3). In all three figures, the sampled population, the sample size, the sample OR and 95% CIs for the sample OR are given. The horizontal dotted line divides previously published data (above the line) and data in the current study (below the line). The ORs with 95% CI for the pooled data from previous studies, for the pooled data from the current study and for the pooled data from all studies are calculated using a random effects model. Allelic ORs are graphed using a logarithmic scale for the x axis.

Published gender-stratified data from previous studies show no evidence of heterogeneity of ORs in male controls versus female controls, but give strong evidence for reduced 30Q frequency in males compared with females (exact test p = 10^-7).
OR = 0.58, 95% CI 0.47 to 0.71). This contrasts with our data, which showed no male–female 30Q frequency differences (OR = 0.99, see fig 1). When our data were combined with previously published data using a random effects model, the estimated mean OR was 0.84 with 95% CI 0.72 to 0.98. Our large sample of 4973 controls showed no difference in male and female 30Q frequencies, but the magnitude of the effect described in previous studies is sufficiently large that analysis of the combined data shows allele frequency differences. The discordance between our data and previously published data suggests that data from additional studies are needed and that it would be premature to conclude that male–female allele frequency differences exist or are absent for DLG5 R30Q.

Female data from previous studies show a trend toward reduced 30Q allele frequency in female patients with CD compared with female controls (exact test p = 0.088, OR = 0.79, 95% CI 0.61 to 1.04). This is consistent with our data. When our data were combined with previously published female data, there was a significant reduction in 30Q frequency in female patients with CD compared with female controls (exact test p = 0.010, OR = 0.86, 95% CI 0.76 to 0.97). None of the female datasets (our female data, previously published female data, or combined female data), showed evidence for heterogeneity of ORs (combined data Q = 19.0, 16 df, p = 0.027, I² = 0.16).

Published male data from previous studies show a significant increase in 30Q frequency in male patients with CD compared with male controls (exact test p = 0.0020, OR = 1.60, 95% CI 1.18 to 2.17). In contrast, our data show a reduced 30Q frequency in male patients with CD compared with controls (OR = 0.87). Consequently, when our male data are combined with previously published male data, there is evidence for heterogeneity of ORs between studies (Q = 36.5, 16 df, p = 0.0025, I² = 0.56), but no evidence that the mean OR differs from 1.0 under a random effects model (mean OR = 0.99, 95% CI 0.80 to 1.22).

**DISCUSSION**

Evidence for association of the 30Q allele with increased risk for CD is provided by the initial two DLG5 studies but attempts to replicate this association generally have not been successful. The first four published whole-genome association studies for CD, three of which genotyped the R30Q variant, did not report DLG5 variants to be associated with CD in their initial published analyses. It has been suggested that differences in male and female population allele frequencies and different effects in males compared with females could make it difficult to replicate the association of 30Q with CD.

In this study, we report gender-stratified data for 4311 males and 5569 females from 12 case–control cohorts. We have tested our data for male–female control frequency differences and for case–control frequency differences in men and in women and we place our data in the context of previously published gender-stratified R30Q data.

The most striking difference between the data in the current study and data in previous studies is the absence of male and female population allele frequency differences for DLG5 R30Q.
in our data (OR = 0.99), but the presence of strong differences in previous data (OR = 0.58, p = 10^{-7}). It is unlikely that sampling variability alone can account for this difference. When we extend CIs to have 99.5% confidence, the CI for the OR in our data and the CI for the OR in previously published data are disjoint.

We considered whether population differences could account for the absence of male–female allele frequency differences in our study. The use of “healthy” controls versus unphenotyped population controls is unlikely to explain the difference, as CD has 1% prevalence. Different age distributions between control cohorts is also unlikely to explain the difference, as previous studies have found male–female allele frequency differences in both newborn and adult population samples. Undetected population stratification could result in the presence (or absence) of male–female allele frequency differences if male and female controls tended to be sampled from different subpopulations.

We investigated the possibility that discrepancies between our data and previously published data could be due to genotyping artefacts. In particular, we examined the possibility that nonspecific primer binding could lead to different distributions of missing and incorrect genotypes in men and women. We compared the DLG5 cDNA sequence (NM_004747) against the human genome using the NCBI MEGABLAST tool (http://www.ncbi.nlm.nih.gov/BLAST/), but found no homologies other than DLG5. A variety of genotyping platforms has been used in the published DLG5 studies and we were not able to trace the difference between published data and our data to the use of any single genotyping platform.

For the studies that have reported the primers used to genotype the R30Q variant, we tested the specificity of primers using the NCBI Nucleotide BLAST tool (“search for short nearly exact matches”). We did not identify any hits on the X or Y chromosome that could lead to non-specific binding in men or women.

A combination of sampling variability combined with a reporting bias may help explain the difference between the set of DLG5 studies that have reported gender-stratified data and the set of DLG5 studies in the current study. Studies that observe a significant male–female allele frequency difference in controls would presumably be more likely to report gender-stratified control data.

Fortunately, many of the whole-genome association studies will genotype the DLG5 R30Q variant in their control cohorts. The R30Q variant is not highly correlated with any other known variant, and no single-nucleotide polymorphisms genotyped for the HapMap project that are within 200 kb of R30Q have pairwise squared correlation coefficients $r^2 > 0.5$ in Caucasians. Consequently, the DLG5 R30Q polymorphism will generally be genotyped in whole-genome association studies that use tagging markers, such as those using the Illumina genotyping platform. We expect that these whole-genome association studies will provide the additional genotype data necessary to confirm or refute the presence of male–female population allele frequency differences for DLG5 R30Q.
Previously published male data show higher 30Q frequency in male patients with CD compared with male controls, but our male data show a trend toward lower 30Q frequencies in male patients with CD compared with male controls. When the data are combined, the estimated mean OR is close to 1.0 (OR = 0.97), so the association of DLG5 30Q with CD in men appears doubtful. We note that there is evidence for heterogeneity of ORs between studies in the combined data (p = 0.0020). If heterogeneity between ORs in different male populations does exist, then an estimated mean OR of 0.97 indicates that the 30Q allele is associated with a reduced risk in some male populations and increased risk in other male populations, which seems unlikely.

Our female data are consistent with previously published female data, and the combined data indicate that 30Q is associated with reduced risk of CD. Thus, although a significant decrease in female 30Q frequency is seen in only 1 of 17 cohorts studied (fig 2), the trend toward lower ORs seen in figure 2 supports an association of 30Q with reduced risk of CD in women (exact test p = 0.010, OR = 0.86, 95% CI 0.76 to 0.97).

The presence of female-specific effects could arise out of interaction with gender-specific risk factors such as oral contraceptive use.40 41 Interestingly, Purmonen et al42 showed that DLG5 is a hormonally regulated member of the MAGUK gene family and suggested DLG5 to be a primary progesterone target gene in human breast cancer cells. This makes biological sense, as progesterone is uniquely a female hormone, produced at significant levels from puberty. Synthetic forms of progesterone and a major component of the progesterone-only contraceptive pill. These associations may help provide an explanation for observed gender specific differences in incidence of CD.42

The estimated OR of 0.84 for the minor 30Q allele is equivalent to an OR of 1.16 for the major R30 allele. The modest DLG5 30Q allele frequency (~0.10) and the small estimated effect size in women indicates that large collaborative efforts will be needed to confirm the association of 30Q with reduced risk of CD in women.

Acknowledgements: Nutrigenomics New Zealand is a collaboration between AgResearch Ltd., Crop and Food Research, HortResearch and the University of Auckland, with funding through the Foundation for Research Science and Technology.

Competing interests: None declared.

REFERENCES