

## ORIGINAL ARTICLE

## Heritability of non-HLA genetics in coeliac disease: a population-based study in 107 000 twins

Ralf Kuja-Halkola,<sup>1</sup> Benjamin Lebwohl,<sup>1,2</sup> Jonas Halfvarson,<sup>3</sup> Cisca Wijmenga,<sup>4</sup> Patrik K E Magnusson,<sup>1</sup> Jonas F Ludvigsson<sup>1,5,6</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2016-311713>).

<sup>1</sup>Department Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

<sup>2</sup>Department of Medicine, Celiac Disease Center, Columbia University Medical Center, Columbia University, New York, USA

<sup>3</sup>Faculty of Medicine and Health, Department of Gastroenterology, Örebro University, Örebro, Sweden

<sup>4</sup>Department of Genetics, University of Groningen, University Medical Center, Groningen, The Netherlands

<sup>5</sup>Department of Pediatrics, Örebro University Hospital, Örebro, Sweden

<sup>6</sup>Division of Epidemiology and Public Health, School of Medicine, University of Nottingham, City Hospital, Nottingham, UK

## Correspondence to

Dr Jonas F Ludvigsson, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 171 77, Sweden; [jonasludvigsson@yahoo.com](mailto:jonasludvigsson@yahoo.com)

Received 23 February 2016

Revised 25 April 2016

Accepted 28 April 2016

Published Online First

20 May 2016



► <http://dx.doi.org/10.1136/gutjnl-2016-312185>



**To cite:** Kuja-Halkola R, Lebwohl B, Halfvarson J, et al. *Gut* 2016;**65**:1793–1798.

## ABSTRACT

**Background and objective** Almost 100% individuals with coeliac disease (CD) are carriers of the human leucocyte antigen (HLA) DQ2/DQ8 alleles. Earlier studies have, however, failed to consider the HLA system when estimating heritability in CD, thus violating an underlying assumption of heritability analysis. We examined the heritability of CD in a large population-based sample of twins, considering HLA.

**Design** In a population-representative sample of 107 912 twins, we identified individuals with CD (equal to villous atrophy) through biopsy reports from all Swedish pathology departments. We calculated concordance rates and tetrachoric correlations for monozygotic (MZ) and dizygotic (DZ) twin pairs. Further, we estimated heritability of CD, first strictly from observed data, and then the non-HLA heritability, representing the heritability of all genetic factors except the HLA locus, using an approach that circumvent the violation of underlying assumptions.

**Results** We identified 513 twins with a diagnosis of CD (prevalence 0.48%). Concordance rates were higher in MZ pairs (0.49) than in DZ pairs (0.10), as were tetrachoric correlations (0.89 in MZ vs 0.51 in DZ pairs). The heritability of CD was 75% (95% CI 55% to 96%). The non-HLA heritability was slightly attenuated, 68% (95% CI 40% to 96%), with shared (17%) and non-shared (15%) environmental factors explaining the remaining variability of CD.

**Conclusions** CD is characterised by a high heritability, but our study also suggests that non-shared environmental factors may be of importance to CD development. HLA seems to have only moderate impact on heritability estimates.

## INTRODUCTION

Coeliac disease (CD) is a chronic immune-mediated disorder that is characterised by small intestinal villous atrophy (VA) and mucosal inflammation.<sup>1</sup> CD occurs in approximately 1% of the Western population<sup>2</sup> and has been linked to an increase in both mortality and morbidity.<sup>3–4</sup> CD is triggered by gluten ingestion in genetically predisposed individuals.<sup>3</sup> Almost all individuals with CD are human leucocyte antigen (HLA)—DQ2/DQ8-positive.<sup>5</sup> While gluten exposure is a prerequisite for the development of CD, recent data indicate that genetic set-up may be more important than timing of gluten introduction.<sup>6–8</sup>

Monozygotic twin (MZ) pairs share 100% of their genetics, while the dizygotic twin (DZ) pairs

## Significance of this study

## What is already known on this subject?

- Genetics play an important role in the aetiology of coeliac disease.
- Human leucocyte antigen (HLA) DQ2/DQ8 is a prerequisite for the coeliac diagnosis.

## What are the new findings?

- Coeliac disease is characterised by a high heritability.
- HLA seems to have only moderate impact on heritability estimates.
- Environmental factors may be of importance to coeliac disease development.

## How might it impact on clinical practice in the foreseeable future?

- Despite a high heritability, researchers should explore environmental risk factors with the aim to decrease the risk of coeliac disease.

have, on average, 50% of their segregating alleles in common. These facts allow researchers to disentangle the relative contribution of genetics and environmental factors in the aetiology of complex diseases like CD. The traditional twin method of estimating heritability of complex diseases relies on the assumption of a large number of non-necessary loci contributing to the disease;<sup>9</sup> in CD, this assumption is violated, since having the DQ2/DQ8 allele is an (almost) necessary cause. To our knowledge, only one independent cohort of twins with CD has been reported in the literature,<sup>10</sup> later followed up by Nisticò *et al.*<sup>11</sup> Nisticò *et al.*<sup>11</sup> observed a concordance rate for CD of 0.83 in MZ twins and 0.17 in DZ twins, but their cohort was only based on 23 MZ and 50 DZ pairs.

The heritability estimates reported by Nisticò *et al.*<sup>11</sup> are limited by the lack of population-based controls and absence of information on the prevalence of CD. Therefore, the reported heritability varied between 57% (95% CI 32% to 93%) and 87% (95% CI 49% to 100%), based on assumed CD prevalence of 1/1000 and 1/91, respectively.<sup>11</sup>

We aimed to investigate the concordance and heritability of biopsy-verified CD in a large population-based cohort of 107 912 twins (53 956 pairs, of which 30 224 were MZ pairs).

## Coeliac disease

Importantly, we wanted to investigate the effect of violation of the underlying assumption of no necessary cause alleles.

### MATERIALS AND METHODS

#### Coeliac disease

All 28 pathology departments in Sweden were contacted, and data on small intestinal biopsies were collected in 2006–2008. The histopathological examinations had been carried out prospectively between 1969 and 2008. This procedure was then updated in 2013 to cover individuals undergoing biopsy after 2008. Data included date of biopsy, biopsy site (duodenum and jejunum), VA (Marsh grade III) and personal identity number.<sup>12</sup> In short, CD was defined as having VA at histopathology examination of mucosal biopsies from the duodenum or jejunum or on a surgical specimen of the duodenum or jejunum. Earlier validation has shown that VA has a positive predictive value of 95% for CD in a Swedish setting,<sup>13</sup> and on average the biopsy involved three small intestinal tissue specimen.<sup>14</sup> The accuracy of VA was thereby higher than when having an inpatient diagnosis of CD (86%).<sup>15</sup> Additional details on the collection of biopsy data have been published previously.<sup>13</sup> In all, we identified 39 935 individuals with CD. These individuals were subsequently matched with the Swedish Twin Registry (STR).

#### The Swedish Twin Registry

The STR began in the late 1950s. The STR contains information about zygosity, sex and dates of birth and death of the twins. In the STR, zygosity determination has been based on questionnaire data about intrapair similarities in childhood, being of opposite sex or DNA analyses. Earlier validation compared with DNA testing has shown that questionnaire data on intrapair similarities have a  $\geq 98\%$  accuracy.<sup>16</sup> STR data were obtained independently of the coeliac data.

We identified all twins alive in 2007 ( $N=146\,830$ ), born since 1906. In this study, we only included twins with identified zygosity ( $N=119\,074$ ), that is, where the twin pairs had been assigned MZ or DZ status based on the above-mentioned procedure, where we had information on birth year and sex of both twins in a pair, resulting in 107 912 twins from 53 956 twin pairs.

We divided twins according to zygosity and gender (male and female MZ, male and female DZ and opposite-sexed DZ). Concordant pairs were those where both twins either had biopsy-verified CD or had no biopsy with VA. Discordant twins were those where one twin had biopsy-verified CD and the other one did not.

#### Statistical analyses

##### Descriptive

We calculated the prevalence of CD in the whole twin cohort as well as in the different zygosity–gender categories. We calculated the median and 1st and 99th percentiles in birth years for the full sample as well as for the zygosity–gender categories. We summarised age of first biopsy-verified CD diagnosis in a histogram, where age was presented in 5-year intervals.

##### Concordance rates and tetrachoric correlations

We calculated the concordance rate as an estimate for the probability of a co-twin having CD if a twin has CD (this was calculated for all zygosity–sex combinations). For this analysis, we used a unity link function and adjusted SEs for dependencies within pairs of twins, using a cluster–robust sandwich estimator. Further, we calculated tetrachoric correlations for all zygosity–sex combinations. A higher tetrachoric correlation in MZ than

in DZ twin pairs signals genetic influence. To test for differences between zygosity–sex subgroups, we (1) fitted a model where each zygosity–sex subgroup was allowed to have different correlations; (2) fitted a model where MZ correlations were constrained to be equal across sexes, and similarly DZ correlations were constrained to be equal across sexes. In both models, the prevalence of CD was assumed equal within sex. The second model was tested against the first, using a likelihood ratio test to see whether assuming equal correlations within zygosity, across sex, made the model fit the data significantly worse. We allowed this test to guide whether sex differences would be considered in any subsequent analyses.

#### Heritability

We proceeded to perform heritability analyses using classic twin methodology.<sup>9</sup> Since we considered a binary outcome, having CD or not having CD, we used the liability–threshold approach, where an underlying normal distribution of the liability of disease is assumed. If the disease is observed in an individual we assume their liability to be above an estimated threshold, and if no disease is observed we assume it is below the threshold. The similarity between twins in pairs is estimated as the correlation between two assumed underlying normal distributions of liability of the two twins; this is equivalent to the tetrachoric correlation.

Based on the assumptions that MZ twins are genetically identical, that DZ twins share on average 50% of their segregating alleles and that MZ and DZ twins share environment to similar degree, we proceeded to partition the variation in the liability of disease into additive genetic (A), shared environmental (C) and non-shared (unique) environmental (E) sources in a model called the ACE model. In this model, we allowed for different prevalence in males and females. We further performed the same analysis with additional adjustment of prevalence for birth year, where we, after an initial inspection of data, allowed a linear effect separately estimated in males and females.

#### Non-HLA heritability

Almost all individuals with CD are DQ2/DQ8-positive; thereby, this trait almost constitutes a prerequisite for diagnosis, just like female sex is almost a condition for developing breast cancer. This implies that the inheritance of at least one of these HLA risk haplotypes can be regarded as an almost necessary causal factor in the aetiology of CD. For a complex trait, this is an unusual situation. Generally, in twin-based modelling of heritability, an assumption of a very large number of factors, each of which has a small and not necessary role, is made. CD violates this assumption, which may bias the result. For these reasons, we performed an analysis where we estimated the heritability, taking the crucial importance of HLA into consideration. Further, since there is little evidence for heterogeneity in effect between the alleles, we considered only two classes of individuals: carriers (with at least one of the risk alleles DQ2/DQ8) and non-carriers (with no risk allele). We analysed the correlation in twin pairs; while accounting for that we only could observe the disease in those who were DQ2/DQ8 carriers; thus, we estimated the non-HLA additive genetic effects on the liability of CD, representing all genetic effects, except that of the HLA locus. We assumed this DQ2/DQ8 carrier frequency to be 25% in the population,<sup>17</sup> equivalent to a combined allele frequency of 13.4% under the assumption of Hardy–Weinberg equilibrium (see online supplementary appendix A). Under these assumptions, we estimated the heritability and environmental sources of variance of the liability for CD, while

incorporating that only carriers of DQ2/DQ8 alleles are at risk for developing the disease. Although we do not use information on molecular measurement of the genotype, the knowledge of allele frequency in the population, plus that MZ twins are genetically identical and that we know how DZ twins inherit parental alleles, allows us to probabilistically include information about individuals being DQ2/DQ8 carriers; see online supplementary appendix A for a technical description. We followed the same steps as for the previous analysis in assumption testing and model fitting.

### Sensitivity analyses

Since our cohort of twins spanned a broad range of birth years, and thus have different coverage ages in the registries, we performed an analysis on a restricted sample. We included all twins who were born between 1970 and 2002 and alive at least until the age of 10, thus taking into account birth years, deaths and, importantly, coverage of registry.

Additionally, considering that Sweden experienced a coeliac 'epidemic' in 1984–1997 among children who were 0–2-year-old,<sup>18</sup> we carried out a sensitivity analysis on the full sample adjusting for prevalence differences in those born in the years 1982–1997.

To test the stability of the DQ2/DQ8-adjusted analyses, with regard to our assumption of carrier frequency, we performed the analysis, assuming prevalence of the necessary cause to vary between 4% and 80% and estimated the heritability accordingly.

### Statistical software

All analyses were performed in R (R: A language and environment for statistical computing. R Foundation for Statistical Computing [program]. Vienna, Austria, 2013. ISBN 3-900051-07-0. <http://www.R-project.org/>). Concordance rates were calculated using the 'gee' function from the 'drgee' package (Drgee: Doubly Robust Generalized Estimating Equations. R package. [program]. 1.1.0 version, 2015.), tetrachoric correlations were calculated using the 'polycor' package (Polycor: Polychoric and Polyserial Ccorrelations, R package, version. <http://CRAN.R-project.org/package=polycor> 0.7-8 version, 2010). Heritability analyses were performed using full likelihood specifications with a non-linear, box-constrained, optimiser ('nlminb') to find maximum likelihood solutions, and the SEs were numerically found using the Hessian (using the 'Hessian' command from the 'numDeriv' package) (numDeriv: Accurate Numerical Derivatives. R. [program]. 2012.9-1 version, 2012.).

### Ethics

This study was approved by the Regional Ethics Review Board in Stockholm (2006/633-31/4). This was a registry-based study, and for this reason no participant was contacted.<sup>19</sup>

## RESULTS

### Descriptive

Of the 107 912 twins included in the study, 513 had biopsy-verified CD. Thus, the prevalence of CD, defined as VA (Marsh III) at histopathology examination, was 0.48% in the Swedish twin cohort; the prevalence of different subgroups of twins is shown in [table 1](#). In [table 1](#), the birth years of the included individuals in the sample are presented with median and 1st and 99th percentiles. In [figure 1](#), the age at first biopsy-verified diagnosis is presented.

**Table 1** Basic descriptives of study population

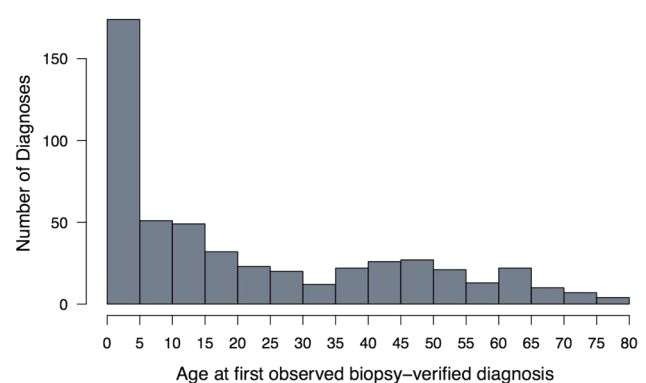
|                                   | Number of individuals | Number of CD cases | Prevalence of CD (%) | Birth years, median (98% interquartile boundaries) |
|-----------------------------------|-----------------------|--------------------|----------------------|----------------------------------------------------|
| All                               | 107 912               | 513                | 0.48                 | 1976 (1929 to 2007)                                |
| Male                              | 52 031                | 164                | 0.32                 | 1977 (1931 to 2007)                                |
| Female                            | 55 881                | 349                | 0.62                 | 1975 (1928 to 2007)                                |
| MZ                                | 30 222                | 140                | 0.46                 | 1976 (1929 to 2005)                                |
| Male                              | 13 698                | 36                 | 0.26                 | 1977 (1931 to 2005)                                |
| Female                            | 16 524                | 104                | 0.63                 | 1976 (1928 to 2005)                                |
| DZ                                | 77 690                | 373                | 0.48                 | 1975 (1930 to 2008)                                |
| Males from same-sexed pairs       | 16 066                | 64                 | 0.40                 | 1973 (1931 to 2006)                                |
| Female from same-sexed pairs      | 17 090                | 120                | 0.70                 | 1968 (1927 to 2005)                                |
| Males from opposite-sexed pairs   | 22 267                | 64                 | 0.29                 | 1980 (1931 to 2008)                                |
| Females from opposite-sexed pairs | 22 267                | 125                | 0.56                 | 1980 (1931 to 2008)                                |

CD, coeliac disease; DZ, dizygotic; MZ, monozygotic.

### Concordance rates and tetrachoric correlations

The highest concordance rate was seen in MZ twins ([table 2](#)), with lower concordance rates in same-sexed DZ twin pairs and the lowest concordance in opposite-sexed DZ twins. A similar pattern was seen when examining the tetrachoric correlations ([table 2](#)). This implies that genetic factors contribute to the development of CD.

We tested whether the MZ correlations, as well as the DZ correlations, could be assumed to be the same across sex. To do this, we fitted a model where the correlations were estimated separately in the MZ male, MZ female, DZ male, DZ female and DZ opposite sexed groups; we then proceeded to fit a model where the MZ male and MZ female correlations were constrained to be the same, and similarly all DZ groups were assumed to have the same correlation; the model including these constraints provided a good fit to the data compared with the model without constraints (difference in  $-2 \log$  likelihood ( $\chi^2$ )=5.36, difference in degrees of freedom (df)=3,  $p=0.147$ ). We performed the same test while adjusting the prevalence of



**Figure 1** Age of first biopsy-verified coeliac disease diagnosis in the current sample.

## Coeliac disease

**Table 2** Concordance rates and tetrachoric correlations

|               | Concordant non-CD pairs | Discordant pairs | Concordant CD pairs | Concordance rate (95% CI) | Tetrachoric correlation (95% CI) |
|---------------|-------------------------|------------------|---------------------|---------------------------|----------------------------------|
| MZ            | 15 005                  | 72               | 34                  | 0.49 (0.38 to 0.59)       | 0.89 (0.84 to 0.94)              |
| Male–male     | 6819                    | 24               | 6                   | 0.33 (0.13 to 0.53)       | 0.82 (0.69 to 0.95)              |
| Female–female | 8186                    | 48               | 28                  | 0.54 (0.42 to 0.65)       | 0.91 (0.86 to 0.96)              |
| DZ            | 38 491                  | 335              | 19                  | 0.10 (0.06 to 0.14)       | 0.51 (0.42 to 0.60)              |
| Male–male     | 7975                    | 52               | 6                   | 0.19 (0.06 to 0.32)       | 0.67 (0.50 to 0.83)              |
| Female–female | 8433                    | 104              | 8                   | 0.13 (0.05 to 0.22)       | 0.54 (0.39 to 0.69)              |
| Female–male   | 22 083                  | 179              | 5                   | 0.05 (0.01 to 0.10)       | 0.40 (0.24 to 0.56)              |

CD, coeliac disease; DZ, dizygotic; MZ, monozygotic.

CD by birth year; again, the constrained model provided a good fit to the data compared with the unconstrained model ( $\chi^2=5.07$ ,  $df=3$ ,  $p=0.167$ ). These results lent support for combining sexes in further analyses.

### Heritability

We proceeded to estimate the heritability, under the assumption that there were no sex differences in the correlations. The heritability was estimated at 75% (95% CI 55% to 96%), and the remaining variation in liability to disease was explained by C—shared environment, 14% (95% CI –5% to 33%) and E—non-shared environment, 11% (95% CI 6% to 15%; [table 3](#)). We performed the same analyses while adjusting for birth year, and the estimates were practically identical ([table 3](#)).

### Non-HLA genetics

To test the influence of the DQ2/DQ8 HLA alleles, we fitted models where the probability of carrying these necessary cause alleles was incorporated (see online supplementary appendix A for details). Under the assumption that a necessary cause allele was present in 25% of the population,<sup>17</sup> we fitted an ACE model, first crude, and then adjusted for birth year ([table 4](#)). The heritability, now representing all genetic effects except that of the HLA locus, was somewhat attenuated compared with the crude analysis (ie, the analysis without taking HLA into consideration); the non-HLA heritability was estimated at 68% (95% CI 40% to 96%), the shared environment at 17% (95% CI –9% to 43%) and the non-shared environment at 15% (95% CI 9% to 22%). Additionally, adjusting for birth year did not change the results markedly ([table 4](#)).

### Sensitivity analyses

We estimated the heritability (not accounting for HLA) in a subsample of individuals born in 1970–2002 (N=51 514). The estimates were consistent with the previous results: the heritability was estimated at 70% (95% CI 46% to 93%), the shared

environment at 21% (95% CI 0% to 43%) and the non-shared environment at 9% (95% CI 4% to 14%).

We estimated the heritability (not accounting for HLA) in the full sample, while adjusting for prevalence differences in those born between 1982 and 1997. When an indicator of being born in this period was included in the model, the regression coefficient for additional linear adjustment for birth year became non-significant. Thus, in addition to allowing for differences in prevalence between sexes, we only allowed for a period effect which was different between genders. Although individuals born in the period had a higher prevalence than the rest of the population, differences in the heritability estimates were slight; the heritability was estimated at 77% (95% CI 57% to 98%), the shared environment at 12% (95% CI –7% to 31%) and the non-shared environment at 11% (95% CI 6% to 16%). Finally, we fitted a series of models where the prevalence of the necessary cause (HLA) was allowed to vary from 4% to 80%. When applicable, we allowed for dominance genetic effects; the result is presented in [figure 2](#).

### DISCUSSION

By analysing a large nationwide cohort of twins and taking HLA genetics into account, we estimated the heritability of CD. Furthermore, our data demonstrate that non-shared environmental factors are important to CD development. To our knowledge, this is the first study to account for the role of HLA genetics in heritability estimates.

The presence of necessary genetic factors violates the assumption of the classic twin model that postulates that many factors of small effect underlie the disease liability (the infinitesimal model). Presence of any of the risk haplotype DQ2/DQ8 is believed to be necessary for the development of CD. Earlier twin studies<sup>10 11</sup> have not properly taken this into account. Our approach to estimate the non-HLA heritability, that is, the heritability from genetic effects except that of the HLA locus, is analogous to estimating the heritability for breast cancer only among women. If no sex distinction would be made in a twin

**Table 3** Heritability of coeliac disease; explained variance of liability to disease

| Model                         | A (95% CI)       | C (95% CI)       | E (95% CI)*     |
|-------------------------------|------------------|------------------|-----------------|
| ACE                           | 75% (55% to 96%) | 14% (–5% to 33%) | 11% (6% to 15%) |
| ACE, adjusted for birth year† | 76% (55% to 96%) | 13% (–6% to 32%) | 11% (6% to 16%) |

\*SE obtained by the delta method.

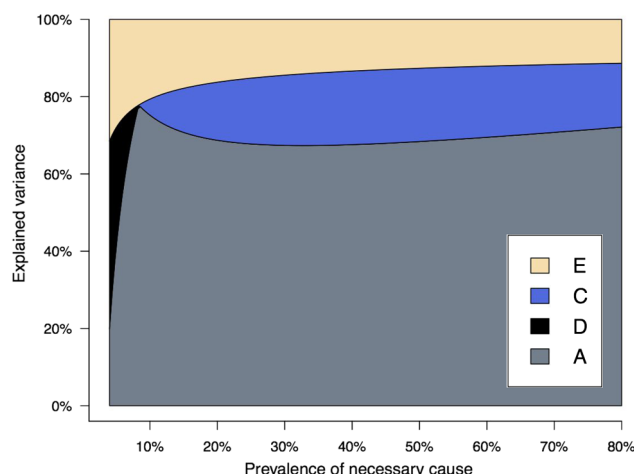
†Birth year adjusted as linear effect, separately for males and females. A, additive genetic; C, shared environment; E, non-shared environment.

**Table 4** Heritability of coeliac disease accounting for a genetic carrier (DQ2/DQ8) present in 25% of the population

| Model                         | A (95% CI)      | C (95% CI)       | E (95% CI)*    |
|-------------------------------|-----------------|------------------|----------------|
| ACE                           | 68% (40 to 96%) | 17% (–9 to 43%)  | 15% (9 to 22%) |
| ACE, adjusted for birth year† | 69% (40 to 97%) | 16% (–10 to 42%) | 15% (9 to 22%) |

\*SE obtained by the delta method.

†Birth year adjusted as linear effect, separately for males and females. A, additive genetic; C, shared environment; E, non-shared environment.



**Figure 2** Genetic and environmental sources of variance over a range of values for the necessary cause. A, additive genetic; D, dominance genetic; C, shared environment; E, non-shared environment.

heritability analysis of breast cancer, we would get biased estimates since breast cancer will (almost) exclusively occur in women. On the other hand, considering the importance of HLA had only minor effects on the heritability estimates. Taking the Swedish epidemic into account<sup>18</sup> did not influence our results.

Our concordance rate in MZ twins was lower (0.49) than that in the earlier Italian study (0.83).<sup>11</sup> This may have several explanations. First, we cannot rule out that some co-twins have not been screened for CD, and are in fact concordant for CD (false negative). Our nationwide dataset is based on real-life data (in contrast, Nisticò *et al*<sup>11</sup> visited all non-affected twins and screened them). Second, concordance rates may differ between countries, just as earlier research has shown strikingly different prevalence estimates of CD between countries.<sup>20</sup>

We found a heritability of 68% when considering HLA. This compares with highly varying estimates, depending on the assumed CD prevalence in the population (57% and 87% with CD prevalence assumed at 1/1000 and 1/91, respectively). In contrast, we calculated our heritability using data from a population with a known prevalence of diagnosed CD (0.48%). We found that common and unique environmental factors contributed 17% and 15%, respectively (in the non-HLA heritability analysis). Our estimate of shared environmental variance was similar to that of the 1/91 CD prevalence model presented by Nisticò *et al*,<sup>11</sup> where shared environment explained 12% of disease liability. We also suggest that non-shared environmental factors play a non-negligible role in CD (15%), while both models in the Italian study suggested that non-shared environmental factors contribute at the most 1% (95% CI 0% to 5%).<sup>11</sup>

### Strengths and limitations

This study has several strengths. Foremost is the large study base. We included more than 107 000 twins (of which 513 had diagnosed CD). The large number of study participants allowed for stratification on sex, where we found a concordance rate for MZ twins of 0.33 in males and 0.54 in females. By employing a novel approach to account for the strong impact of the HLA locus, we found that the heritability estimate in CD was only moderately affected by the violation of the underlying assumptions in heritability analysis. The diagnosis of CD was based on biopsy record data, which up until 2012<sup>21</sup> were the reference standard for diagnosis in both adults and children and have

remained so in adults<sup>1 3</sup> (but with an option to abstain from biopsy in a subset of children). Some 95% of individuals with VA have CD in a Swedish setting,<sup>13</sup> accuracy exceeding that of having a physician-assigned diagnosis of CD in the Swedish inpatient register.<sup>15</sup> Our use of biopsy reports is also likely to have a high sensitivity for diagnosed CD, as during the time of data collection >95% of all adult gastroenterologists and paediatricians reported that they performed a biopsy before assigning a diagnosis of CD to the patient.<sup>13</sup> A positive CD serology was not requested for the diagnosis of CD. However, a previous validation of the cohort found that some 88% of patients with available serology data were positive for transglutaminase, endomysium or gliadin antibodies.<sup>12</sup> Finally, data on CD and twin status were obtained independently.

Although our detection of diagnosed CD is virtually complete,<sup>13</sup> an important limitation of our data is that we cannot detect all individuals with CD, since a significant proportion of CD cases remains undiagnosed. We had no data on presenting symptoms in the twins included in this analysis, but an earlier validation found that in 118 individuals with biopsy-verified CD from our original cohort, 36% had diarrhoea and 35% had anaemia.<sup>13</sup> Although we paid much attention to the violation of the basic assumption of non-necessary alleles, some assumptions remain in estimation of the heritability: The assumption of equally shared environment (C) in MZ and DZ twin pairs (equal environments assumption) has not been tested in the current study, and may, if violated, bias the heritability estimate (likely upwards). Potential deviations from additive genetic effects of alleles have not been, and cannot be, investigated in the current data. Although additive genetic effects in twin models capture parts of dominance and epistatic genetics,<sup>22 23</sup> we have no means to infer whether dominance and epistatic genetics are present, nor to which extent they affect CD, and thus the heritability estimates may be biased. Finally, the E component in our model included measurement errors and that may have impacted on our findings.

### CONCLUSION

In conclusion, our study confirms a high heritability in CD, even when considering HLA in the statistical model. Since the concordance rate among MZ twins was 0.49, our study suggests that also non-shared environmental factors may contribute to the risk of CD.

**Contributors** ICMJE criteria for authorship read and met: JFL, RKH, BL, JH, CW and PKEM. Agreed with the manuscript's results and conclusions. Approved the final version of the manuscript: JFL, RKH, BL, JH, CW and PKEM. Designed the study: JFL, RKH and PKEM. Analysed the data: RKH. Wrote the first draft of the paper: JFL and RKH. Contributed to the writing of the paper: BL, JH, CW and PKEM. Contributed to the design of study and interpretation of the data analyses: BL, JH, CW. Responsible for data integrity: JFL, RKH and PKEM. Obtained funding: JFL.

**Funding** JFL was supported by grants from the Swedish Society of Medicine and the Swedish Research Council.

**Competing interests** None declared

**Ethics approval** This project (2006/633-31/4) was approved by the Regional Ethical Review Board in Stockholm (Karolinska Institutet), Sweden on 14 June 2006.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** Data can be obtained through the Swedish National Board of Statistics upon request.

### REFERENCES

- Ludvigsson JF, Leffler DA, Bai JC, *et al*. The Oslo definitions for coeliac disease and related terms. *Gut* 2013;62:43–52.
- Dubé C, Rostom A, Sy R, *et al*. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology* 2005;128:S57–67.

## Coeliac disease

- 3 Ludvigsson JF, Bai JC, Biagi F, *et al*. Diagnosis and management of adult coeliac disease: guidelines from the British Society of Gastroenterology. *Gut* 2014;63:1210–28.
- 4 Elfström P, Granath F, Ekström Smedby K, *et al*. Risk of Lymphoproliferative Malignancy in Relation to Small Intestinal Histopathology Among Patients With Coeliac Disease. *J Natl Cancer Inst* 2011;103:436–44.
- 5 Lundin KE, Scott H, Hansen T, *et al*. Gliadin-specific, HLA-DQ(alpha 1\*0501,beta 1\*0201) restricted T cells isolated from the small intestinal mucosa of coeliac disease patients. *J Exp Med* 1993;178:187–96.
- 6 Liu E, Lee HS, Aronsson CA, *et al*. Risk of pediatric coeliac disease according to HLA haplotype and country. *N Engl J Med* 2014;371:42–9.
- 7 Lionetti E, Castellana S, Francavilla R, *et al*. Introduction of gluten, HLA status, and the risk of coeliac disease in children. *N Engl J Med* 2014;371:1295–303.
- 8 Vriezinga SL, Auricchio R, Bravi E, *et al*. Randomized feeding intervention in infants at high risk for coeliac disease. *N Engl J Med* 2014;371:1304–15.
- 9 Neale MC, Cardon LR. *Methodology for Genetic Studies of Twins and Families (Nato Science Series D:)*. Dordrecht: Kluwer Academic Publishers, 1992.
- 10 Greco L, Romino R, Coto I, *et al*. The first large population based twin study of coeliac disease. *Gut* 2002;50:624–8.
- 11 Nisticò L, Fagnani C, Coto I, *et al*. Concordance, disease progression, and heritability of coeliac disease in Italian twins. *Gut* 2006;55:803–8.
- 12 Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, *et al*. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 2009;24:659–67.
- 13 Ludvigsson JF, Brandt L, Montgomery SM, *et al*. Validation study of villous atrophy and small intestinal inflammation in Swedish biopsy registers. *BMC Gastroenterol* 2009;9:19.
- 14 Ludvigsson JF, Brandt L, Montgomery SM. Symptoms and signs in individuals with serology positive for coeliac disease but normal mucosa. *BMC Gastroenterol* 2009;9:57.
- 15 Smedby KE, Akerman M, Hildebrand H, *et al*. Malignant lymphomas in coeliac disease: evidence of increased risks for lymphoma types other than enteropathy-type T cell lymphoma. *Gut* 2005;54:54–9.
- 16 Lichtenstein P, De Faire U, Floderus B, *et al*. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J Intern Med* 2002;252:184–205.
- 17 Abadie V, Sollid LM, Barreiro LB, *et al*. Integration of genetic and immunological insights into a model of coeliac disease pathogenesis. *Annu Rev Immunol* 2011;29:493–525.
- 18 Ivarsson A, Persson LA, Nyström L, *et al*. Epidemic of coeliac disease in Swedish children. *Acta Paediatr* 2000;89:165–71.
- 19 Ludvigsson JF, Håberg SE, Knudsen GP, *et al*. Ethical aspects of registry-based research in the Nordic countries. *Clin Epidemiol* 2015;7:491–508.
- 20 Mustalahti K, Catassi C, Reunanen A, *et al*. The prevalence of coeliac disease in Europe: results of a centralized, international mass screening project. *Ann Med* 2010;42:587–95.
- 21 Husby S, Koletzko S, Korponay-Szabó IR, *et al*. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 2012;54:136–60.
- 22 Keller MC, Medland SE, Duncan LE. Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behav Genet* 2010;40:377–93.
- 23 Mäki-Tanila A, Hill WG. Influence of gene interaction on complex trait variation with multilocus models. *Genetics* 2014;198:355–67.



## Heritability of non-HLA genetics in coeliac disease: a population-based study in 107 000 twins

Ralf Kuja-Halkola, Benjamin Lebwohl, Jonas Halfvarson, Cisca Wijmenga, Patrik K E Magnusson and Jonas F Ludvigsson

*Gut* 2016 65: 1793-1798 originally published online May 20, 2016  
doi: 10.1136/gutjnl-2016-311713

---

Updated information and services can be found at:  
<http://gut.bmj.com/content/65/11/1793>

---

*These include:*

### References

This article cites 22 articles, 8 of which you can access for free at:  
<http://gut.bmj.com/content/65/11/1793#BIBL>

### Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

### Topic Collections

Articles on similar topics can be found in the following collections  
[Coeliac disease](#) (537)

---

### Notes

---

To request permissions go to:  
<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:  
<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:  
<http://group.bmj.com/subscribe/>