Published in final edited form as:

Inflamm Bowel Dis. 2012 February ; 18(2): 312–322. doi:10.1002/ibd.21781.

Transmission distortion in Crohn's disease risk gene *ATG16L1* leads to sex difference in disease association

Linda Y. Liu^{1,2,3}, Marc A. Schaub⁴, Marina Sirota^{1,2,3}, and Atul J. Butte^{1,2,3,*}

Linda Y. Liu: lyl@stanford.edu; Marc A. Schaub: marc.schaub@cs.stanford.edu; Marina Sirota: msirota@stanford.edu; Atul J. Butte: abutte@stanford.edu

¹ Stanford Center for Biomedical Informatics Research, Stanford University School of Medicine, Stanford, California, United States of America

² Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, California, United States of America

³ Lucile Packard Children's Hospital, Palo Alto, California, United States of America

⁴ Computer Science Department, Stanford University, Stanford, California, United States of America

Abstract

Background—Crohn's disease (CD), an inflammatory disease of the bowel, affects millions of people around the world. Evidence suggests that disease onset and pathogenesis differ between males and females. Yet no comprehensive efforts exist to assess the sex-specific genetic architecture of CD.

Methods—We used genotyping data from a cohort of 1748 CD cases and 2938 controls to investigate 71 meta-analysis-confirmed CD risk loci for sex differences in disease risk. We further validated the significant results in separate cohorts of 968 CD cases and 2809 controls, and performed a meta-analysis across datasets.

Results—The SNP rs3792106 (C/T) in *ATG16L1* showed a significant sex effect with p-value 6.9×10^{-13} and allelic odds ratio 1.48 in females, and p-value 0.013 and odds ratio 1.22 in males (odds ratio heterogeneity p-value 0.037). Surprisingly, the difference was found to arise from a discrepancy in allele frequencies between male and female controls (p-value 0.0045) rather than cases. We found similar results for this SNP in the separate validation data sets. Using 155 HapMap 3 trios, we detected significant maternal over-transmission of the T allele at rs3792106 (p-value 0.027).

Conclusions—Our results indicate that different transmission patterns between sexes may sustain the disparate allele frequencies at rs3792106 in healthy populations, and furthermore that a virus-risk variant mechanism implicated in CD alters the distribution in diseased patients. To our knowledge, this is the first report of sex-specific CD association in *ATG16L1*. The possible implications in Crohn's disease and basic human biology present interesting areas for future investigation.

*Corresponding Author: Atul J. Butte, MD, PhD, Stanford University School of Medicine, 251 Campus Drive MS-5415, Room X-163 Stanford, CA 94305-5415, Phone: (650) 723-3465, Fax: (650) 723-7070, abutte@stanford.edu.

The authors declare that there are no conflicts of interest related to this manuscript.

LYL & MAS conceived the study, LYL designed the study, LYL and AJB obtained the data, LYL & MAS & MS analyzed the data, LYL performed statistical analysis, LYL & MAS & MS wrote the paper, AJB supervised the study and revised the manuscript.

Keywords

Inflammatory bowel disease; *ATG16L1*; transmission distortion; sexual dimorphism

Introduction

Crohn's disease (CD) is a form of autoimmune inflammatory bowel disease that affects over 3 million individuals in Europe and North America. Genetic predisposition is a major component of disease susceptibility, although environmental factors such as smoking are also associated with increased risk of CD (1). There is also substantial evidence of sex-related differences in the presentation and course of both adult- and pediatric-onset Crohn's disease. Female CD patients require more frequent ileocecal resections and experience significantly shorter time to disease recurrence following resections (2). Additionally, pediatric female CD patients experience more severe symptoms compared to pediatric male patients (3). Previous studies have found sex-specific effects in autosomal polymorphisms associated with the disease, including within the HLA region on chromosome 6 (4), and an interaction between the IL-6 promoter on chromosome 7 and early-onset of CD in males (5).

There is some previous evidence suggesting that transmission ratio distortion in CD susceptibility genes may contribute to sex effects. For instance, studies have shown that the possible protective effect of the R30Q allele in the *DLG5* gene for female pediatric CD patients (6) may be due to sex-dependent transmission distortion, but *DLG5* has not fully replicated in recent genome-wide scans (7). There is also widespread evidence of transmission distortion effects in multiple sclerosis (MS), another autoimmune disease with female predominance (8). Researchers showed that the major histocompatibility locus contains a MS risk allele with sex-specific risk and maternal over-transmission to female offspring (9). They further hypothesized that epigenetic modifications and gene-environment interactions may mediate parent-of-origin effects at this locus (10).

In this study, we build on previous examples of sex-specific polymorphisms associated with CD (4, 5) as well as growing evidence that many complex human diseases have a complicated underlying genetic architecture involving many more sex-specific loci than was previously assumed (11). We extend the previous analyses to systematically evaluate all confirmed CD susceptibility loci for sex differences in disease risk. Recent meta-analyses of CD genome-wide association studies have increased the number of CD susceptibility loci to 71 (12). We tested these 71 loci using genotyping data generated from the Wellcome Trust Case Control Consortium (WTCCC) study of Crohn's disease (13). We then validated our results using data from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Inflammatory Bowel Disease Genetics Consortium study (14). Surprisingly, our results indicated that a locus in an intronic region of the gene *ATG16L1* (Autophagy related 16-like 1) has a sex-specific disease association that is due in large part to different allele distributions in healthy male and female control subjects. We used family trios from the recently-released HapMap Phase III genotyping data (15) and found evidence of parent-specific transmission of the allele to offspring in diverse healthy populations. Our observations provide additional explanations for the variability in CD susceptibility even in patients with the same genetic variant, and suggest a role for a sex-specific interaction of a genetic variant with inflammatory processes in determining risk and progression of Crohn's disease.

Materials and Methods

Genotyping study participants and data quality control

We obtained the data for our analysis from the original Wellcome Trust Case Control Consortium (WTCCC) study of 2,000 cases for each of seven complex human diseases and 3,000 shared controls (1,500 from the 1958 British Birth Cohort, and 1,500 blood donors recruited specifically for the project) (13). Access to the data was authorized by the Wellcome Trust Case Control Consortium. The original study performed genotyping using the Affymetrix GeneChip 500K Mapping Array Set, which comprises 500,568 SNPs. We used genotyping data only from the Crohn's disease (CD) and control cohorts. We applied quality control criteria at both the SNP level and the individual level. These criteria exclude SNPs with a minor allele frequency (MAF) less than 1%, SNPs that are not in Hardy-Weinberg equilibrium in the controls, SNPs that show significant differences between the two control populations, and individuals missing genotype calls or showing non-European ancestry. We excluded all 31,011 SNPs and 256 CD and 64 control individuals that were excluded during quality control checks in the original WTCCC study. The WTCCC study also manually inspected the raw signal intensities of all SNPs that showed a significant association between cases and control in order to detect false positives caused by spurious genotype calling or other genotyping quality issues. An additional 578 SNPs were excluded during at this stage, and we also excluded those SNPs from our analysis. In order to limit the impact of genotype calling errors, we combine the genotype calls made by the two algorithms used by the WTCCC, Chiamo (Marchini *et al.*, in preparation) and BRLMM (Affymetrix). For each SNP, we individually consider the genotype calls of both programs for each individual; if the calls made by the two algorithms differ, or if one of the two algorithms does not make a call, then the genotype of the individual for this SNP is considered unknown. Individual participants with more than 5% of unknown genotypes are excluded from our analysis, and SNPs with more than 5% of unknown genotypes are also excluded.

Selection of genome-wide significant SNPs from 71 confirmed loci

We started with the list of 71 replicated CD SNPs published by Franke et al (12). We intersected this list with the list of all SNPs passing quality control filters described above. This resulted in 24 SNPs that were commonly measured in the published meta-analysis and also on the Affymetrix platform used by the WTCCC study. We used a chi-square test with two degrees of freedom to compute disease association test statistics and p-values for these 24 SNPs and applied a genome-wide (based on 301,275 SNPs) Bonferroni correction to the SNP p-values. We limited our analysis to SNPs with a Bonferroni-corrected p-value < 0.05 in males, females, or combined sexes. Since none of the 24 commonly measured SNPs passed this filter, we also expanded our analysis to include any genes reported by the meta-analysis. We looked for SNPs in those genes on the Affymetrix chip, regardless if the exact SNPs were reported in the previous study. We used NCBI's dbSNP database to assign SNPs to genes if they were designated as being in a gene or near a gene (16). 44 genes reported by Franke et al were represented by SNPs on the Affymetrix chip, and 4 of those genes contained a total of 11 genome-wide significant SNPs. Those were the only SNPs tested for sex effects. For the SNP found to have significant sex effects, we also plotted the normalized signal intensity data generated from the Affymetrix intensity ('CEL') files by the WTCCC. We manually verified that there is no suspicious difference in signal intensity between male and female samples (Supplementary Figure 1). To obtain the linkage disequilibrium measurement between SNPs rs3792106 and rs3792109, we used data from the HapMap project and linkage disequilibrium values for the CEU population (17).

Validation cohorts

To confirm the finding for rs3792106, we obtained data from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Inflammatory Bowel Disease Genetics Consortium study (14). The data consisted of 968 cases and 995 unrelated controls. Access to the data was authorized by the NIDDK after obtaining IRB approval. The study performed genotyping using the Illumina HumanHap300 Genotyping BeadChip, which measures 317,503 SNPs. We only analyzed data for rs3792106. We additionally obtained data from the International Multiple Sclerosis Genetics Consortium (IMSGC), which consisted of 931 family trios (2,793 total individuals) in which the child has multiple sclerosis and the parents are healthy (18). Access to the data was authorized by the IMSGC. The study performed genotyping using the Affymetrix GeneChip 500K Mapping Array Set. We only analyzed data for rs3792106.

Statistical Analysis

We calculated male and female disease association p-values separately for the SNPs passing our filtering criteria (described above). Using CD cases and controls of one sex, we performed a chi-square test with two degrees of freedom. We also computed sex-specific odds ratios and used Woolf's test to assess heterogeneity of male and female odds ratios for our SNP of interest (19). Woolf's test is a method for testing the heterogeneity of 2×2 contingency tables over multiple strata. The method tests whether the log odds ratios are the same in all strata. For our case, the strata are male and female and we test if the allelic odds ratios (ORs) for disease association are equal between male and female for a single SNP. This test is sensitive both to ORs in opposite directions (the allele is protective for one sex and risky for the other) as well as in the same direction but of different magnitude. We represent allelic counts for males and females separately for one SNP in 2×2 contingency tables such as the following:

	Cases	Controls
Has allele 'A'	a	b
No allele 'A'	c	d

We then compute a pooled OR using the Mantel-Haenszel method:

$$OR_{MH} = \frac{\sum_{i=1}^k \frac{a_i d_i}{T_i}}{\sum_{i=1}^k \frac{b_i c_i}{T_i}}, \text{ where } k=2 \text{ strata and } T_i = a_i + b_i + c_i + d_i$$

Then the Woolf's chi-square statistic is: $\chi^2_{(1)} = \sum_{i=1}^k \frac{(\ln OR_i - \ln OR_{MH})^2}{\text{var}(\ln OR_i)}$, where $k = 2$ and $\text{var}(\ln OR_i) = \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}$.

We additionally performed logistic regression with a genotype-sex interaction term to assess the effect of sex on rs3792106 genotype. Using the following multivariate logistic regression model for a single SNP, we model the probability of the outcome variable (the cohort to

which an individual belongs, case or control) as a function of predictor variables X_1 and X_2 (genotype and sex).

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

We use p_i to represent the probability of belonging to the disease class, and model the natural logarithm ($\ln$) of the odds of the probabilities. Letting m represent the minor allele and M the major allele in the controls, we coded genotype (X_1) as $MM=0$, $Mm=1$, and $mm=2$. Sex (X_2) was coded as male=1, female=2. We represented the effect of sex on genotype by including the interaction term $X_1 X_2$. We tested the significance of the coefficient β_3 of the interaction term and report p-values.

To test genotype frequency differences between male and female control cohorts, we used a chi-square test with two degrees of freedom, and to test allele frequency differences, we used a chi-square test with one degree of freedom.

We performed inverse variance-weighted fixed effect meta-regression (20) on rs3792106 allele counts across three cohorts (WTCCC, NIDDK, and IMSGC). We used control sets from all three studies, and CD cohorts from the WTCCC and NIDDK studies, and conducted separate meta-analyses in control and case cohorts. We computed the odds ratio in each individual study as the odds of being male and having the 'T' allele. The odds ratios are weighted by the inverse of the standard error of the odds ratio, then combined to estimate the meta-effect. The meta-effect that we estimate with the fixed effect regression model is then interpreted as the effect of sex on allele frequency across all studies in the analysis. We used this model to estimate the meta-effect for both the control analysis and CD analysis. We used the package 'meta' in the R to do this analysis (21).

Transmission analysis using HapMap 3 data

We obtained genotyping data from the HapMap Phase III, which includes individuals in 240 family trios from five populations (15). The 663 individuals with genotyping samples passing quality control include 83 ASW (African ancestry in Southwest USA), 165 CEU (Utah residents with Northern and Western European ancestry), 171 MKK (Maasai in Kinyawa, Kenya), 77 MXL (Mexican ancestry in Los Angeles, CA), and 167 YRI (Yoruba in Ibadan, Nigeria). We excluded all individuals who were not part of a family trio of mother, father and child, and we also excluded trios missing any of the three members. We required every member of a trio to have a known genotype for rs3792106. Trios that satisfied all of these conditions are termed "complete" and we only used complete trios for transmission analysis. After combining complete trios from the five populations, we had 155 trios for our analysis. To be able to trace transmission of alleles from parents to offspring, we used only trios where the pairing of the parents constitutes an informative mating. For each type of analysis we performed, the definition of informative mating varied. To test transmission disequilibrium for male versus female children, we used trios where both parents are heterozygous, and used a chi-square test to compare the allele distribution in children versus the expected distribution of 50% C, 50% T. For parental origin analysis, we used trios where exactly one parent is heterozygous, to test the proportion of times fathers and mothers transmitted each allele. In every case since we know the expected allele distribution in the children, we used chi-square tests to assess the significance of the observed allele distribution.

Results

In this study, we carried out sex-specific analysis of genome-wide association study (GWAS) data for Crohn's disease, and focused in particular on single nucleotide polymorphisms (SNPs) in genes that show previous evidence of disease association. Figure 1 shows the overall process of our study.

***ATG16L1* shows evidence of sex-specific risk loci in WTCCC data**

We obtained genome-wide association study data from the Wellcome Trust Case Control Consortium (13). After quality control filtering (see Methods), 301,275 SNPs and 1694 Crohn's Disease (CD) and 2928 control individuals remained in our analysis. The CD cohort was 61% female and 39% male, and the control cohort was 51% female and 49% male. We intersected the list of 71 confirmed CD SNPs (12) with the 301,275 SNPs and found that 24 of the 71 SNPs were measured on the Affymetrix platform used by the WTCCC. However, none of these 24 SNPs achieved a genome-wide significant disease risk Chi-square p-value in the WTCCC cohort based on a Bonferroni correction on 301,275 SNPs (see Methods). As different studies may report different SNPs as "tags" for genes, we decided to expand our analysis to include all genes linked to the 71 loci and found that 44 of those genes were measured by SNPs on the Affymetrix chip. We analyzed all SNPs linked to those 44 genes and discovered that 11 SNPs in four genes (*ATG16L1*, *BSN*, *IL23R*, and *NOD2*) attained genome-wide significance in the WTCCC cohort. Table 1 shows the male and female disease association p-values and odds ratios for the 11 SNPs tested for sex differences. Only the five SNPs in *ATG16L1* (rs3792106, rs6431654, rs10210302, rs3828309, and rs6752107) show large sex differences in disease association. Figure 2 shows SNPs on chromosome 2 that are in or near *ATG16L1*. Clearly, the five SNPs rs3792106, rs6431654, rs10210302, rs3828309, and rs6752107 are significantly associated to CD risk in females, while showing non-significant disease association in males. We observe a cluster of significant p-values in females, with gradual increasing and decreasing significance on either side. This phenomenon is most likely due to the linkage disequilibrium in this region, and provides evidence of a biological effect rather than an artifact of genotyping. A similar effect is also visible in males, but to a much smaller degree.

rs3792106 has significant allele frequency differences between healthy males and females

The SNP rs3792106 in *ATG16L1* has the most extreme difference between male and female disease association (Table 1), so we chose to focus on this single SNP for further analysis. We noted that rs3792106 is in linkage disequilibrium with rs3792109, the SNP reported in conjunction with *ATG16L1* in the meta-analysis. The R-square between these two SNPs is 0.8 (see Methods). In WTCCC, the female disease association p-value for rs3792106 is 6.85×10^{-13} , which is more significant than the p-value for the combined-sex cohort (4.77×10^{-12}), even with the smaller sample size when considering females only. The male p-value for this SNP is 0.013, which is not genome-wide significant. In addition, we observe different effect sizes in males and females. The Woolf test for heterogeneity of effect sizes is significant, with p-value 0.037. Additionally, in a logistic regression model, the sex*genotype interaction term for rs3792106 is significant, with p-value 0.03. Taking a closer look at the genotype and allele frequencies of this SNP in CD cases and controls, we discovered that while male and female CD patients had relatively similar distributions, male and female controls had significantly different allele frequencies (Table 2). The frequency of male and female control CC genotypes is 6% higher in males, and CT genotype frequency is 4.55% higher in females. Overall, the frequency of the C allele is nearly 4% greater in male controls than in female controls. The two degree-of-freedom chi-square test between male and female control genotype frequencies has p-value 0.0025, and the one degree-of-freedom chi-square test between male and female control allele frequencies has p-value 0.0045. The

corresponding tests in CD male and female cases are not significant (p-values of 0.28 and 0.64, respectively).

Meta-analysis of rs3792106 allele frequency in three GWAS data sets

To validate the observed allele frequencies for rs3792106, we performed a meta-analysis using GWAS data from WTCCC and two additional European cohorts. We analyzed genotyping data of 968 CD cases and 995 unrelated controls from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Inflammatory Bowel Disease Genetics Consortium (14). We also used a set of 931 family trios from the International Multiple Sclerosis Genetics Consortium (IMSGC) characterized by multiple sclerosis (MS)-affected children and healthy parents (18). We only used the healthy parents (931 males and 931 females) as an additional control set. We only tested the single SNP rs3792106 in the NIDDK and IMSGC data sets. The allele frequency distribution between healthy male and female controls seen in WTCCC was also observed in these data sets (Figure 3A; Supplementary Tables 1 and 2). We performed a meta-analysis of the sex-specific allele distributions in the three control data sets using inverse variance weighting and a fixed effect regression model (see Methods). We found the three data sets to be homogeneous with a meta-effect size of 0.88 and 95% confidence interval (CI) of [0.82, 0.95] (Figure 3B). We therefore concluded that there is a significant sex effect on rs3792106 allele distributions in healthy individuals across these three data sets. We also performed a meta-analysis across the two CD cohorts from WTCCC and NIDDK. As in the WTCCC data, we observed no sex difference in allele distributions in the NIDDK CD patients (Figure 3C). The fixed-effect meta-regression model showed the two data sets to be homogeneous with a meta-effect size of 1.01 and 95% CI of [0.90, 1.14] (Figure 3D). Thus, there is no sex difference in allele frequencies of CD patients across these two data sets. Since the three control cohorts and two case cohorts are respectively from homogeneous distributions, we can compute combined allele distributions in the control and case cohorts. Figure 3E shows the total counts for the combined control and case cohorts. The one degree-of-freedom chi-square test for male versus female controls is significant, while the male and female allele distributions in cases are identical.

Transmission distortion in rs3792106 in HapMap 3 trios

Previous studies have reported evidence of transmission distortion at a locus on chromosome 10 as a possible mechanism for maintaining different allele frequencies between healthy males and females as well as sex-specific Crohn's disease association (6, 7). Additionally, there has been evidence of transmission distortion in the vicinity of rs3792106 on chromosome 2 in the Framingham population (22). In the Framingham study, the SNP reported (rs1000548) is only 44KB away from rs3792106. Therefore, we hypothesized that transmission distortion at the SNP locus rs3792106 may contribute to the difference we observed in allele frequencies between healthy males and females. To test this, we obtained genotype data from the International HapMap Project, as a previous study demonstrated that use of HapMap cohorts enabled detection of transmission distortion (23). The data contained 240 healthy trios (father, mother and child) from five populations: ASW (African ancestry in Southwest USA), CEU (Utah residents with Northern and Western European ancestry), MKK (Maasai in Kinyawa, Kenya), MXL (Mexican ancestry in Los Angeles, CA), and YRI (Yoruba in Ibadan, Nigeria) (15). After performing filtering (described in Methods), 155 complete trios remained for our analysis. For this investigation, we combined all five populations into a single cohort. Although the allele frequencies for this SNP differ slightly between the populations, we were only interested in testing transmission effects between parents and children, and therefore could combine the different cohorts. Table 3 presents the genotype and allele frequencies for rs3792106 in the HapMap 3 trios. Although the difference in allele frequencies between male and female parents is less

than 1%, there is a 13% difference in allele frequencies between male and female children in the opposite direction as the parents. This was a further clue that a transmission effect is present. We investigated two possibilities: differential allele transmission based on the sex of the child, and differential allele transmission based on the parent of origin. Table 4 contains results from both transmission analyses. We detected notable (although non-significant) differences in allele transmission to male versus female children. In cases when both parents are heterozygous for the SNP, female children were twice as likely to receive the C allele than the T allele. Male children were more likely to receive the T allele than the C allele. These effects are consistent with the allele frequencies seen in the children. In addition, we surprisingly detected significant transmission distortion based on the parental origin of alleles. For this analysis, we only used cases when one parent is heterozygous and the other parent is homozygous. We detected significant over-transmission of the T allele from mothers to children of both sexes (p-value 0.027). Mothers were twice as likely to transmit the T allele than the C allele. There is a slight over-transmission of the C allele from fathers, but this is not significant (p-value 0.34).

Discussion

In this work, we have demonstrated that variants in the Crohn's disease susceptibility gene *ATG16L1* display sexual dimorphism in disease risk. *ATG16L1* is located on chromosome 2 and encodes a component of a large protein complex essential for delivery of cytoplasmic proteins and organelles to the lysosomes for degradation, a process known as autophagy (24). In particular, the SNP rs3792106 is significantly associated to CD in females only, with a significant difference in effect size between men and women. In fact, the association for this SNP in WTCCC females is even stronger than the association in the larger combined-sex cohort, and the effect is nearly absent in males. Therefore, the effect of the SNP in females essentially drives the observed disease association in the combined cohort. We observe a similar effect for four other SNPs in tight linkage disequilibrium on the gene *ATG16L1* and propose that polymorphisms in this gene may in fact be risk variants for Crohn's disease only for females. This novel hypothesis should be investigated in other CD genome-wide association studies, particularly those involving individuals of other ethnic groups.

The idea that *ATG16L1* is a female-specific risk gene informs efforts to explain the observed sex differences in the presentation and course of CD. We observed that in three large cohorts of healthy control individuals, there is a significant difference in allele distributions between males and females at rs3792106. We make the conclusion that a possible cause of allele frequency differences is transmission disequilibrium between healthy men and women. This result is fascinating in light of recent research demonstrating that an allele may have opposite effects on the risk of a complex disease depending on the parent from which it was inherited (25). Therefore, the fact that mothers under-transmit the CD risk allele at this locus could interplay with different effects on disease risk based on parental origin. Additionally, the difference in allele distributions between parents and children in large numbers of family trios leads to the hypothesis that rs3792106 has some deleterious effect on survival or reproduction. These observations could be used to build a model of allele transmission at this locus in which transmission probabilities are dependent on parental origin (see Figure 1).

The difference in age of onset between male and female CD patients is also interesting. There is a male-bias of CD incidence in pediatric patients, and a female-bias of CD in adult patients (5, 26). Researchers have hypothesized that *DLG5*, a primary progesterone target gene, has a protective effect only in female children but then loses its protective effect in adult women due to the hormonal changes that occur during female puberty (6). Similarly,

male early-onset CD could be due to increased transcription of IL-6 and up-regulation of inflammatory pathways, an effect that may be repressed by estrogen in females (5). Conversely, *ATG16L1* has only been associated with CD in adult patients and has not been shown to influence disease risk in childhood-onset CD (27). These results, along with our finding that *ATG16L1* increases risk only in females, are in alignment with the female bias of CD incidence in adult patients. These clinical and genetic examples present further evidence of underlying physiological and biological sex differences that should be further investigated to reveal new insights into the disease mechanism of CD.

There have been other previous studies demonstrating clinical sex differences in CD due to an interaction with another disease process. For example, there is evidence of a sex-specific interaction between inflammatory bowel disease and colon cancer from a prospective cohort study. The authors reported that having inflammatory bowel disease conferred a significantly higher risk of colon cancer to men than to women (28). A separate study suggested that variants in the p53/mdm2 apoptosis network (implicated in several human tumor types) are sexually dimorphic in their effect on CD risk (29).

A recent study demonstrates that yet another type of interaction, this time between variants in *ATG16L1* and a murine norovirus, induces pathologic abnormalities similar to CD in mice (30). Interestingly, mice that carry the mutation but are not exposed to the virus do not exhibit changes in intestinal phenotype. These results suggest the possibility of a similar virus-plus-susceptibility gene interaction involving *ATG16L1* in humans. Such an interaction presents an interesting hypothesis for the roughly equal allele distributions we observed in male and female WTCCC and NIDDK CD cases. If there is a subtype of CD cases in humans for which both a mutation and a viral infection are needed, then the patient population having this subtype will all have the mutation, irrespective of their sex. This would reduce the overall difference in allele frequency between male and female cases. As infection rates might be different between men and women, this could potentially explain the allele distributions we observe. Our results thus highlight the importance of carefully considering sex effects in any future analysis of the interaction between viral infections and the *ATG16L1* gene in humans. The known differences in autophagy and general immune responses between men and women (31–35), as well as the genetic differences we identified in healthy controls, indicate that a potential virus-plus-susceptibility gene interaction at *ATG16L1* in human may be sex-specific, a hypothesis that should be considered when looking for such interactions.

Despite much being known about the effect of polymorphisms in *ATG16L1* on Crohn's disease pathogenesis, our finding is a novel demonstration of sexual dimorphism in the effect of this gene on disease risk. Our result concerning transmission distortion is interesting especially since it holds true across five diverse populations. However, the fact that we had to combine five cohorts in order to have a sufficient sample of trios highlights a sample size issue in using HapMap data for this kind of analysis. If we had used the CEU cohort alone (since those individuals are closest in ethnicity to the WTCCC and NIDDK individuals), it would not have provided enough trios in each category of informative matings. Additionally, we had to remove many trios from consideration due to missing data or missing family relationships. Even with the 155 trios that were remaining, we only had around 40 trios for each sub-type of transmission analysis we performed. Also, we have no data on disease status for any HapMap individuals, so it is unclear how the transmission phenomenon that we observed relates to Crohn's disease. It may be an independent biological phenomenon that may have developed due to selective forces on sex-specific fitness conferred by the variant, or the parent-specific transmission may impact CD susceptibility in a sex-specific manner. To investigate these questions, similar transmission analyses should be performed on large numbers of family trios with CD-affected children as

well as control trios. The results we present here should be further confirmed in subsequent studies, but are an interesting avenue for further investigation into the sex differences in manifestation of Crohn's disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge the following individuals and funding sources for their contribution to this work: Joel Dudley, Rong Chen, Alex Morgan, David Ruau, Hewlett Packard Foundation, and Lucile Packard Foundation for Children's Health. We thank Jay Pasricha for helpful comments on the manuscript. We also thank Alex Skrenchuk and Boris Oskotsky from Stanford University for computer support.

L.Y.L. is funded by a Hewlett family Stanford Graduate Fellowship and a National Science Foundation Graduate Research Fellowship.

M.A.S. is funded by a Richard and Naomi Horowitz Stanford Graduate Fellowship.

M.S. is funded by the US National Library of Medicine (T15 LM007033).

A.J.B. is funded by the US National Institute of General Medical Sciences (R01 GM079719).

This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113.

The Inflammatory Bowel Disease Genetics Consortium (IBDGC) study was conducted by the IBDGC Investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The data from the IBDGC reported here were supplied by the NIDDK Central Repositories. This manuscript was not prepared in collaboration with Investigators of the IBDGC study and does not necessarily reflect the opinions or views of the IBDGC study, the NIDDK Central Repositories, or the NIDDK.

This study makes use of data generated by the International Multiple Sclerosis Genetics Consortium (IMSGC). A full list of the investigators who contributed to the generation of the data is available from www.imsgc.org.

References

1. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology*. 2004; 126:1504–1517. [PubMed: 15168363]
2. Wagtmans MJ, Verspaget HW, Lamers CB, et al. Gender-related differences in the clinical course of Crohn's disease. *Am J Gastroenterol*. 2001; 96:1541–1546. [PubMed: 11374696]
3. Gupta N, Bostrom AG, Kirschner BS, et al. Gender differences in presentation and course of disease in pediatric patients with Crohn disease. *Pediatrics*. 2007; 120:e1418–1425. [PubMed: 18055660]
4. Fisher SA, Hampe J, Macpherson AJ, et al. Sex stratification of an inflammatory bowel disease genome search shows male-specific linkage to the HLA region of chromosome 6. *Eur J Hum Genet*. 2002; 10:259–265. [PubMed: 12032734]
5. Sagiv-Friedgut K, Karban A, Weiss B, et al. Early-onset Crohn disease is associated with male sex and a polymorphism in the IL-6 promoter. *J Pediatr Gastroenterol Nutr*. 2010; 50:22–26. [PubMed: 19934771]
6. Biank V, Friedrichs F, Babusukumar U, et al. DLG5 R30Q variant is a female-specific protective factor in pediatric onset Crohn's disease. *Am J Gastroenterol*. 2007; 102:391–398. [PubMed: 17156146]
7. Friedrichs F, Brescianini S, Annese V, et al. Evidence of transmission ratio distortion of DLG5 R30Q variant in general and implication of an association with Crohn disease in men. *Hum Genet*. 2006; 119:305–311. [PubMed: 16446977]

8. Kantarci OH. Sex-stratified inheritance of MS: New horizons from studies in MHC region. *Neurology*. 2011; 76:210–212. [PubMed: 21209375]
9. Chao MJ, Herrera BM, Ramagopalan SV, et al. Parent-of-origin effects at the major histocompatibility complex in multiple sclerosis. *Hum Mol Genet*. 2010; 19:3679–3689. [PubMed: 20634196]
10. Chao MJ, Ramagopalan SV, Herrera BM, et al. MHC transmission: Insights into gender bias in MS susceptibility. *Neurology*. 2011; 76:242–246. [PubMed: 21209377]
11. Ober C, Loisel DA, Gilad Y. Sex-specific genetic architecture of human disease. *Nat Rev Genet*. 2008; 9:911–922. [PubMed: 19002143]
12. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010; 42:1118–1125. [PubMed: 21102463]
13. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature*. 2007; 447:661–678. [PubMed: 17554300]
14. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet*. 2007; 39:596–604. [PubMed: 17435756]
15. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010; 467:52–58. [PubMed: 20811451]
16. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001; 29:308–311. [PubMed: 11125122]
17. A haplotype map of the human genome. *Nature*. 2005; 437:1299–1320. [PubMed: 16255080]
18. Hafler DA, Compston A, Sawcer S, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med*. 2007; 357:851–862. [PubMed: 17660530]
19. Woolf B. On estimating the relation between blood group and disease. *Ann Hum Genet*. 1955; 19:251–253. [PubMed: 14388528]
20. Borenstein, M. *Introduction to meta-analysis*. Chichester, West Sussex, U.K.; Hoboken: John Wiley & Sons; 2009.
21. Schwarzer, G. R package version 16–1. 2010. meta: Meta-Analysis with R.
22. Paterson AD, Waggott D, Schillert A, et al. Transmission-ratio distortion in the Framingham Heart Study. *BMC Proc*. 2009; 3 (Suppl 7):S51. [PubMed: 20018044]
23. Santos PS, Hohne J, Schlattmann P, et al. Assessment of transmission distortion on chromosome 6p in healthy individuals using tagSNPs. *Eur J Hum Genet*. 2009; 17:1182–1189. [PubMed: 19259136]
24. Mizushima N, Kuma A, Kobayashi Y, et al. Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. *J Cell Sci*. 2003; 116:1679–1688. [PubMed: 12665549]
25. Kong A, Steinthorsdottir V, Masson G, et al. Parental origin of sequence variants associated with complex diseases. *Nature*. 2009; 462:868–874. [PubMed: 20016592]
26. Kappelman MD, Rifas-Shiman SL, Kleinman K, et al. The prevalence and geographic distribution of Crohn's disease and ulcerative colitis in the United States. *Clin Gastroenterol Hepatol*. 2007; 5:1424–1429. [PubMed: 17904915]
27. Van Limbergen J, Russell RK, Nimmo ER, et al. Autophagy gene ATG16L1 influences susceptibility and disease location but not childhood-onset in Crohn's disease in Northern Europe. *Inflamm Bowel Dis*. 2008; 14:338–346. [PubMed: 18088053]
28. Soderlund S, Granath F, Brostrom O, et al. Inflammatory bowel disease confers a lower risk of colorectal cancer to females than to males. *Gastroenterology*. 2010; 138:1697–1703. [PubMed: 20167217]
29. Zimmer V, Widmann T, Muller M, et al. Genotypic interaction and gender specificity of common genetic variants in the p53/mdm2 network in Crohn's disease. *Digestion*. 2010; 81:246–251. [PubMed: 20110711]

30. Cadwell K, Patel KK, Maloney NS, et al. Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. *Cell*. 2010; 141:1135–1145. [PubMed: 20602997]
31. Fish EN. The X-files in immunity: sex-based differences predispose immune responses. *Nat Rev Immunol*. 2008; 8:737–744. [PubMed: 18728636]
32. Meier A, Chang JJ, Chan ES, et al. Sex differences in the Toll-like receptor-mediated response of plasmacytoid dendritic cells to HIV-1. *Nat Med*. 2009; 15:955–959. [PubMed: 19597505]
33. Zhang H, Bain JL, Caskey CP, et al. Effects of gender on serum biomarkers of systemic inflammation coincident to experimentally-induced periapical lesions. *Arch Oral Biol*. 2010
34. Du L, Hickey RW, Bayir H, et al. Starving neurons show sex difference in autophagy. *J Biol Chem*. 2009; 284:2383–2396. [PubMed: 19036730]
35. Sobolewska A, Gajewska M, Zarzynska J, et al. IGF-I, EGF, and sex steroids regulate autophagy in bovine mammary epithelial cells via the mTOR pathway. *Eur J Cell Biol*. 2009; 88:117–130. [PubMed: 19013662]

Figure 1.

Study flowchart. Analysis of Wellcome Trust Case Control Consortium (WTCCC) genotyping data led to discovery of a sex-specific SNP in *ATG16L1*, rs3792106. Healthy male and female controls from WTCCC, NIDDK and IMSGC data sets have significantly different allele distributions, which prompted investigation of transmission distortion at this locus using 155 healthy family trios in five cohorts from the International HapMap Project (HapMap 3). Based on our observations from all the data sets, a model similar to the one diagrammed here could be proposed for allele transmission for the SNP rs3792106. The dashed arrows from female parents represent different transmission probabilities from male parents.

Figure 2.

Manhattan plot showing male and female disease association p-values separately for SNPs in *ATG16L1* that are genome-wide significant in combined male and female CD cohort. The $-\log_{10}$ of the disease association p-values for females are plotted on the positive y-axis, and the $\log_{10}$ of the disease association p-values for males are plotted on the negative y-axis. The horizontal red and blue lines indicate Bonferroni thresholds. Chromosomal base position is plotted on the x-axis, with the position of *ATG16L1* shown in green.

Figure 3.

Meta-analysis results for SNP rs3792106 in WTCCC, NIDDK and IMSGC data sets.

A) Allele counts and frequencies for rs3792106 in three control cohorts.

B) Estimate for effect of sex on allele distribution from inverse variance-weighted fixed effect meta-regression in controls. Weight refers to the weight each study is given in the meta-regression model, based on the inverse of the variance of the allele effect size. I^2 and τ^2 measure inter-study variance. P-values are from Cochran's Q test of heterogeneity. OR = odds ratio; S.E. = standard error of odds ratio; CI = confidence interval.

C) Allele counts and frequencies for rs3792106 in two CD cohorts.

D) Estimate for effect of sex on allele distribution from inverse variance-weighted fixed effect meta-regression in cases.

E) Combined allele counts and frequencies for rs3792106 in three control cohorts and two CD cohorts. Also included are results of Chi-square tests between male and female allele distributions. Asterisks indicate significant p-values.

Table 1

Disease association p-values and odds ratios for 11 SNPs in genes with confirmed disease risk association in WTCCC data set and Franke et al meta-analysis. Combined p-value, male p-value and female p-value are calculated from case-control disease association tests using WTCCC data. Odds ratios (OR) refer to case-control allelic odds ratios in WTCCC.

SNP	Combined p-value	Female p-value	Male p-value	Combined OR	Female OR	Male OR	Gene
rs3792106	4.77×10^{-12}	6.85×10^{-13}	0.013	1.35	1.48	1.22	<i>ATG16LI</i>
rs6431654	7.75×10^{-14}	1.15×10^{-12}	0.0021	1.39	1.51	1.26	<i>ATG16LI</i>
rs10210302	8.10×10^{-14}	9.21×10^{-13}	0.0025	1.38	1.51	1.25	<i>ATG16LI</i>
rs3828309	5.90×10^{-13}	4.52×10^{-12}	0.0036	1.37	1.50	1.25	<i>ATG16LI</i>
rs6752107	1.59×10^{-13}	1.13×10^{-12}	0.0036	1.38	1.51	1.25	<i>ATG16LI</i>
rs9858542	4.33×10^{-8}	5.93×10^{-6}	0.0037	1.25	1.30	1.22	<i>BSN</i>
rs11805303	6.75×10^{-12}	2.00×10^{-5}	2.70×10^{-7}	1.39	1.33	1.46	<i>IL23R</i>
rs17375018	5.02×10^{-8}	3.37×10^{-5}	0.0021	1.31	1.34	1.26	<i>IL23R</i>
rs2201841	6.79×10^{-11}	1.03×10^{-4}	4.08×10^{-7}	1.36	1.29	1.45	<i>IL23R</i>
rs17221417	1.04×10^{-10}	3.53×10^{-7}	3.22×10^{-4}	1.37	1.38	1.33	<i>NOD2</i>
rs3135499	4.02×10^{-8}	0.0013	1.30×10^{-5}	1.29	1.24	1.36	<i>NOD2</i>

Table 2

A) Genotype counts and frequencies for rs3792106 in WTCCC Crohn's disease and control cohorts.
B) Allele counts and frequencies for rs3792106.
Also included are results of Chi-square tests between male and female genotype or allele distributions. Asterisks indicate significant p-values.

A						
Crohn's disease genotypes				Control genotypes		
	CC	CT	TT	CC	CT	TT
Male	272 (41.03%)	309 (46.61%)	82 (12.37%)	510 (35.34%)	700 (48.51%)	233 (16.15%)
Female	452 (43.84%)	440 (42.68%)	139 (13.48%)	436 (29.36%)	788 (53.06%)	261 (17.58%)
	Chi-square p-value between M/F cases: 0.28			Chi-square p-value between M/F controls: 0.0025*		

B				
Crohn's disease alleles			Control alleles	
	C	T	C	T
Male	853 (64.33%)	473 (35.67%)	1720 (59.6%)	1166 (40.4%)
Female	1344 (65.18%)	718 (34.82%)	1660 (55.89%)	1310 (44.11%)
	Chi-square p-value between M/F cases: 0.64		Chi-square p-value between M/F controls: 0.0045*	

Table 3

A) Genotype counts and frequencies for rs3792106 in HapMap 3 cohort. All trios from ASW, CEU, MKK, MXL, and YRI populations are combined.

B) Allele counts and frequencies for rs3792106.

A				
		CC	CT	TT
All trios	Male	55 (23.40%)	113 (48.09%)	67 (28.51%)
	Female	60 (26.09%)	116 (50.43%)	54 (23.48%)
Parents only	Male	40 (25.81%)	75 (48.39%)	40 (25.81%)
	Female	39 (25.16%)	75 (48.39%)	41 (26.45%)
Children only	Male	15 (18.75%)	38 (47.50%)	27 (33.75%)
	Female	21 (28.00%)	41 (54.67%)	13 (17.33%)

B			
		C	T
All trios	Male	223 (47.45%)	247 (52.55%)
	Female	236 (51.30%)	224 (48.70%)
Parents only	Male	155 (50.00%)	155 (50.00%)
	Female	153 (49.35%)	157 (50.65%)
Children only	Male	68 (42.50%)	92 (57.50%)
	Female	83 (55.33%)	67 (44.67%)

Table 4

A) Transmission analysis for rs3792106 based on sex of child, using informative trios from HapMap3. The observed count of C and T alleles in male and female children are compared to what is expected based on a double heterozygous mating. Also included are results of Chi-square tests between observed and expected allele distributions in children.

B) Transmission analysis for rs3792106 based on parental origin of alleles. The observed count of C and T alleles in children are compared to what is expected based on each type of mating. Also included are results of Chi-square tests between observed and expected allele distributions in children. Asterisks indicate significant p-values.

A									
Sex-of-child analysis									
Male children					Female children				
Observed		Expected		Observed		Expected			
C	T	C	T	C	T	C	T		
Heterozygous parents		15	23	19	19	21	11	16	16
CT father × CT mother									
		Chi-square p-value = 0.19		Chi-square p-value = 0.077					

B									
Parental origin analysis									
				Observed		Expected			
Allele transmitted from mother				C	T	C	T	Chi-square p-value	
CT mother × CC father				7	18	12.5	12.5	0.028 *	
CT mother × TT father				6	9	7.5	7.5	0.44	
Total transmitted from mother				13	27	20	20	0.027 *	
				Observed		Expected			
Allele transmitted from father				C	T	C	T	Chi-square p-value	
CT father × CC mother				11	8	9.5	9.5	0.49	
CT father × TT mother				12	9	10.5	10.5	0.51	
Total transmitted from father				23	17	20	20	0.34	