



Published in final edited form as:

*Nat Genet.* 2007 May ; 39(5): 596–604. doi:10.1038/ng2032.

## Genome-wide association study identifies five novel susceptibility loci for Crohn's disease and implicates a role for autophagy in disease pathogenesis.

John D. Rioux<sup>1,2</sup>, Ramnik J. Xavier<sup>3</sup>, Kent D. Taylor<sup>4</sup>, Mark S. Silverberg<sup>5</sup>, Philippe Goyette<sup>1</sup>, Alan Huett<sup>3</sup>, Todd Green<sup>2</sup>, Petric Kuballa<sup>3</sup>, M. Michael Barmada<sup>6</sup>, Lisa Wu Datta<sup>7</sup>, Yin Yao Shugart<sup>8</sup>, Anne M. Griffiths<sup>9</sup>, Stephan R. Targan<sup>4</sup>, Andrew F. Ippoliti<sup>4</sup>, Edmond-Jean Bernard<sup>10</sup>, Ling Mei<sup>4</sup>, Dan L. Nicolae<sup>11</sup>, Miguel Regueiro<sup>12</sup>, L. Philip Schumm<sup>13</sup>, A. Hillary Steinhart<sup>5</sup>, Jerome I. Rotter<sup>4</sup>, Richard H. Duerr<sup>6,12</sup>, Judy H. Cho<sup>14,16</sup>, Mark J. Daly<sup>2,15,16</sup>, and Steven R. Brant<sup>7,8,16</sup>

<sup>1</sup>Université de Montréal and the Montreal Heart Institute, Research Center, 5000 rue Belanger, Montreal, Quebec H1T 1C8, Canada

<sup>2</sup>The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

<sup>3</sup>Gastrointestinal Unit and Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, MA 02114, USA

<sup>4</sup>Medical Genetics Institute and IBD Center, Cedars-Sinai Medical Center, 665 W, 8700 W. Beverly Blvd., Los Angeles, CA 90048, USA

<sup>5</sup>Mount Sinai Hospital IBD Centre, University of Toronto, 441-600 University Avenue, Toronto, Ontario M5G 1X5, Canada

<sup>6</sup>University of Pittsburgh, Graduate School of Public Health, Department of Human Genetics, Crabtree A300, 130 Desoto Street, Pittsburgh, PA 15261, USA

<sup>7</sup>Johns Hopkins University, Department of Medicine, Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center, 1503 East Jefferson Street, Baltimore, MD 21231, USA

<sup>8</sup>Johns Hopkins University, Bloomberg School of Public Health, Department of Epidemiology, 615 E. Wolfe Street E6036, Baltimore, MD 21205, USA

<sup>9</sup>The Hospital for Sick Children, Department of Pediatrics, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

<sup>10</sup>Université de Montréal and CHUM - Hôtel-Dieu de Montréal, 3875 Saint-Urbain Street, Montreal, Quebec H2W 1V1, Canada

<sup>11</sup>University of Chicago, Department of Medicine, 5801 South Ellis, Chicago, IL 60637, USA

---

Corresponding author: John D. Rioux Montreal Heart Institute 5000 Belanger Street Montreal, Quebec, Canada, H1T 1C8 E-mail: rioux@broad.mit.edu.

Collection of patient samples and clinical information was performed by S.R.B., J.H.C., R.H.D., E.-J. B., M.R., J.D.R., J.I.R., M.S., A.H.S., L.W.D., Y.Y.S. and K.D.T. Design of the genome-wide study was performed by S.R.B., J.H.C., M.J.D., R.H.D., J.D.R., and M.S. Design of the replication study was performed by J.H.C., M.J.D., P.G., J.D.R., J.I.R., S.R.B. and K.D.T. Analysis of the genome-wide and replication data was performed by T.G. and P.G., respectively, with contributions from M.M.B. and D.N., under the supervision of M.J.D. and J.D.R. Analysis of the Cedars-Sinai ATG16L1 genetic data was performed by L.M. under the supervision of K.D.T. and J.I.R. Molecular and cellular biology experiments were performed by P.K. and A.H., under the supervision of R.J.X. The manuscript was written by M.J.D., J.D.R. and R.J.X., with contributions from S.R.B., J.H.C., R.H.D., M.S., A.H.S., Y.Y.S., D.L.N. and K.D.T. The analysis and manuscript writing efforts of this multi-center study was coordinated by J.D.R.

<sup>12</sup>University of Pittsburgh, School of Medicine, Department of Medicine, Division of Gastroenterology, Hepatology and Nutrition, UPMC Presbyterian, Mezzanine Level, C-Wing, 200 Lothrop Street, Pittsburgh, PA 15213, USA

<sup>13</sup>University of Chicago, Department of Health Studies, 5841 S. Maryland Avenue, Chicago, IL 60637, USA

<sup>14</sup>Yale University, Department of Medicine, Division of Gastroenterology, IBD Center, 300 Cedar Street S155A, New Haven, CT 06519, USA

<sup>15</sup>Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, MA 02114, USA

<sup>16</sup>These authors share senior position.

## Abstract

We present a genome-wide association study of ileal Crohn's disease (CD) and two independent replication studies that identify five novel regions of association to CD. Specifically, in addition to the previously established CARD15 and IL23R associations, we report strong associations with independent replication to variation in the genomic regions encoding the PHOX2B, NCF4 and ATG16L1 genes, as well as a predicted gene on 16q24.1 (FAM92B) and an intergenic region on 10q21.1. We further demonstrate that the ATG16L1 gene is expressed in intestinal epithelial cell lines and that functional knock down of this gene abrogates autophagy of *Salmonella typhimurium*. Together these findings suggest that autophagy and host cell responses to intra-cellular microbes are involved in the pathogenesis of CD.

---

Crohn's disease (CD) and ulcerative colitis (UC) represent the two common forms of idiopathic inflammatory bowel disease (IBD), each with a prevalence of roughly 100-150 per 100,000 individuals of European ancestry<sup>1</sup>. CD most commonly involves the ileum and colon but can affect any region of the gut. UC always involves the rectum, and inflammation may extend as far as the cecum in a contiguous pattern<sup>2</sup>. Strong familial aggregation, twin studies and established genetic associations attest to the important role of genetics in IBD pathogenesis<sup>3-5</sup>. There is also very strong evidence that the enteric microflora plays a central role in the initiation and maintenance of disease. Therefore, like most complex trait diseases, IBD results from a combination of genetic and non-genetic risk factors, where each individual factor may be expected to have a relatively modest effect on disease risk.

While a combination of genome-wide linkage, candidate gene and targeted association mapping studies have been successful in the identification of CD-associated genetic variants in CARD15 and the *IBD5* haplotype, these explain only a small fraction of the heritability of CD<sup>6-8</sup>. We therefore embarked upon a genome-wide association (GWA) study of CD in order to find additional genetic risk factors. Phenotypes for both CD and UC vary considerably among individuals, primarily with regard to sites of inflammation, disease behavior, severity and extraintestinal manifestations. Furthermore, CD site and behavior are likely under genetic control based on clustering within affected sibling pairs<sup>9</sup>, as well as specific observations that CARD15 mutations are a greater risk factor for ileal CD and stricturing behavior<sup>10</sup>. Therefore we have exclusively focused on patients with CD involving the ileal region of the small intestine (with or without other sites of involvement) in an attempt to minimize clinical and genetic heterogeneity. Based on an interim analysis approximately halfway through this study, we identified, confirmed and published the discovery of genetic variants in the IL23R gene that significantly influence risk to developing CD and UC<sup>11</sup>. Specifically at that point, 567 non-Jewish ileal CD cases had been scanned and analyzed. Here we report the results of the completed scan, with genotyping of 988 ileal CD patients and 1007 controls performed using the Illumina HumanHap300 Genotyping BeadChip, replication of the significantly associated

SNPs in independent familybased and case-control cohorts, and preliminary functional data that begins to place these novel genes into biological context and suggests potential pathogenic mechanisms.

## RESULTS

### Genome-wide association analysis of CD cases and controls

In order to identify genetic factors influencing risk to develop CD, we typed 308,332 SNPs using DNA samples from 988 patients with ileal CD and 1007 controls that were collected by the members of the North American NIDDK IBD Genetics Consortium. Overall, 98.6% of genotypes were called in this experiment; however, before analysis, quality filtering of both samples and SNPs were performed to insure robust association tests (see Methods). After performing these data filtering measures, we had a final data set of genotypes consisting of 304,413 SNPs in 946 cases and 977 controls with an average call rate of 99.35%. Our baseline analysis was a Cochran-Mantel-Haenszel (CMH) test using two groups (Jewish and non-Jewish) to accommodate potential subtle differences in the genetic background of the two groups. The median chi-square observed was slightly inflated so we corrected all statistics using a genomic control (GC) factor of 1.056.

This GC-corrected CMH analysis produced a significant excess of positive associations – for example 27 SNPs in 9 distinct genomic regions were associated with  $p < 10^{-5}$ , a level which we would informally expect to observe roughly 3 times by chance in this scan given the number of tests performed (See Fig. 1 and Table 1). The two most significant loci, with multiple SNPs being significantly associated, corresponded to the *CARD15* and *IL23R* genes that have previously been confirmed as IBD genes. The remaining regions identified with extreme p-values were all novel, showed no unusual pattern of missing data or departure from Hardy-Weinberg equilibrium (HWE) in controls (of the best SNPs in the top 23 regions examined further below, all had HWE- $p > .01$  in controls and 18 were missing less than 1% genotypes), and many were supported by nearby correlated SNPs also showing association. It should be noted that the confirmed *IBD5* CD risk haplotype was not among the top loci in this study of ileal CD as it only had modest evidence of association ( $p < 0.01$ ).

### Two independent replication studies confirm novel IBD loci

While certainly encouraging, even an overall high-quality data set and robust analyses cannot be assumed to perfectly filter out all artifactual sources of false positives. We therefore tested the most significantly associated SNP in the 23 independently associated regions with 5  $p < 5 \times 10^{-5}$  in an independent set of 650 individuals with ileal CD and their parents (Replication Cohort #1), to attempt replication of these findings in a robust family-based setting and with an alternative genotyping technology (iPlex; Sequenom). Although we did not succeed in typing three of these SNPs (and the replication of *CARD15* and *IL23R* had already been established and are not considered here) the remaining 18 SNPs had high quality genotype data that passed our quality thresholds. After removal of DNAs with low genotype call rates and families with an excess of Mendel errors, the final analyses included 530 trios. Association testing of this final dataset from Replication Cohort #1 identified that six out of the 18 SNPs had significant evidence of replication (Table 1; see Supplementary Table 1 online): rs2241880 ( $p=0.00089$ ); rs16853571 ( $p=0.00037$ ); rs224136 ( $p=0.00021$ ); rs4821544 ( $p=0.019$ ); rs8050910 ( $p=0.024$ ); rs11617463 (0.037). To provide further evidence for the association of these replicated SNPs, we typed these six in another independent cohort of 353 patients with ileal CD and 207 controls (Replication Cohort #2) on yet another independent genotyping platform (Taqman; Applied Biosystems). One of these failed to type but the other five were successfully typed and provided additional evidence of replication (Table 1; see Supplementary Table 1 online). Considering the two replication studies together (using Mantel-Haenszel

statistics), two of these SNPs were unequivocally confirmed as novel CD risk factors (rs2241880,  $p < 5 \times 10^{-8}$ ; rs224136,  $p < 5 \times 10^{-7}$ ). In combination with the genome-wide scan data, these two SNPs are considerably in excess of the most conservative genome-wide thresholds (rs2241880,  $p < 10^{-13}$ ; rs224136,  $p < 10^{-10}$ ). Three additional loci (rs16853571, rs4821544, and rs8050910) showed very strong evidence of replication ( $p < 0.01$ ). Even when setting aside the two most significantly associated novel SNPs (rs2241880 and rs224136), observing replication of 3/16 SNPs with  $p < 0.01$  is highly unlikely to occur by chance ( $p = 0.0005$ ). Since there is only a 1% chance of observing even two such results by chance, it is highly likely that these three SNPs represent true CD susceptibility variants, but further confirmation is required to achieve the same level of certainty as attained for the IL23R, rs2241880 and rs224136 SNPs identified in this study.

### Association mapping implicates novel IBD genes

Given the significant evidence of replication for these five novel loci, we examined the linkage disequilibrium structure surrounding the replicated SNPs using the genotype data from the GWA screen, and examined the association mapping data from the GWA, replication studies as well as some additional higher density typing in order to map these association signals to specific genomic regions and genes. As shown in Supplementary Figure 1 online, in four of these instances, the association mapping implicates specific genes, whereas in the fifth it implicates an intergenic region.

Specifically, the LD structure and association mapping around the most associated SNP (GWA,  $p = 6.4 \times 10^{-8}$ ; Replication,  $p = 4.1 \times 10^{-8}$ ), rs2241880, implicates a region on chromosome 6q27.1 containing a single gene known as ATG16 autophagy related 16-like 1 (*S. cerevisiae*) or ATG16L1 gene. In fact, this SNP encodes a non-synonymous amino acid change – an alanine to threonine substitution in exon 8 – Ala197Thr. Logistic regression analyses conditional on the Ala197Thr in the family-based samples indicates that this coding variant can fully explain the association signal to this locus – therefore we consider this to be the causal risk variant. In all of the datasets tested in this study, the threonine allele is the minor allele and has a protective effect; in the GWA dataset the estimated odds ratio is 0.692 (95% CI of 0.608-0.788). Interestingly, the threonine is the ancestral allele and is the major allele in the YRI, JPT, and HCB samples typed in the International HapMap Project (see Supplementary Fig. 2 online). ATG16L1 is part of a family of genes involved in autophagy – a biological process involved in protein degradation, antigen processing, regulation of cell signaling and many other pathways essential to the initiation and regulation of the inflammatory response – and therefore this association suggests that this biological process is likely to have an important role in disease pathogenesis.

The next SNP with incontrovertible association to ileal CD is rs224136 (GWA,  $p = 7.9 \times 10^{-6}$ ; Replication,  $p = 2.9 \times 10^{-7}$ ). The association mapping in this case, however, implicates a genomic region on chromosome 10q21.1 of approximately 70 kb that does not contain any known protein coding genes. This associated intergenic region is flanked on the centromeric side by the zinc finger gene known as ZNF365 and on the telomeric side by a predicted protein (c10orf22) and by the known early growth response gene EGR2, also a zinc finger protein. Although there is little known about the ZNF365 and c10orf22 genes, mutations in EGR2 are known to lead to the development of congenital hypomyelinating neuropathy (MIM#605253) or of Charcot-Marie-Tooth disease (MIM#607678). Potentially more relevant, however, is the observation that EGR2 appears to be a negative regulator of T cell activation in mice<sup>12</sup>, and although speculative, genetic variation within this intergenic region could play a role in the control of expression of EGR2 and influence biological function.

The replicated SNP rs16853571 (GWA,  $p = 7.7 \times 10^{-7}$ ; Replication,  $p = 8.4 \times 10^{-3}$ ) is located in the promoter region of the Paired-Like Homeobox 2B (PHOX2B) gene on chromosome 4p13 (see

Supplementary Fig. 1 online) - 2143 bp upstream of the first known exon and 10 bp upstream of a putative alternate first exon. PHOX2B is a homeodomain transcription factor, predominantly expressed in differentiating neurons. Human mutations in the PHOX2B gene have been implicated in disorders of neural crest development, congenital central hypoventilation syndrome and Hirschsprung disease<sup>13,14</sup>. Inactivation of PHOX2B in mice disrupts noradrenergic differentiation throughout the nervous system. Autonomic ganglia fail to develop properly and these mice have striking defects in enteric neurons<sup>14-16</sup>.

The genomic region identified by rs4821544 (GWA,  $p=2.9E-05$ ; Replication,  $p=9.0E-03$ ) also contains a single gene known as NCF4 (rs4821544 itself is in intron 1). NCF4 encodes the 7 p40phox protein that has previously been demonstrated to play an important role in NADPH oxidase activity and the generation of reactive oxygen species (ROS) production upon phagocytosis<sup>17,18</sup>; both important for mounting an effective anti-microbial response.

Finally, rs8050910 (GWA,  $p=3.3E-05$ ; Replication,  $p=8.5E-03$ ) is located within the 4<sup>th</sup> intron of the predicted gene FAM92B located on chromosome 16q24.1. This gene prediction is supported by multiple mRNAs and spliced ESTs, as well as significant conservation across multiple species (data not shown). This gene encodes a protein with no known function or recognizable motifs.

It should be noted that when using a stratified analysis approach, none of these three loci appeared to have any statistically significant epistatic interactions with the CARD15, IBD5 or IL23R genes (data not shown). Though some association with UC is seen for rs224136, little or no evidence is seen at the other newly identified SNPs (see Supplementary Table 1 online).

### RNA and protein expression analyses provide initial biological context to novel genes

Since we were interested in determining more about ATG16L1's function and biological context, we examined the expression distribution of ATG16L1 and other known autophagy components by quantitative real-time RT-PCR in a variety of epithelial and immune cell lines (see Fig. 2A, B and C). ATG5, 7 and 16 were broadly expressed, with SW480 cells having the lowest overall mRNA abundance relative to GAPDH controls. Apart from the markedly elevated levels of ATG7 seen in THP-1 cells, expression of each of the three components followed a similar pattern in the cells tested, as might be suggested by the stoichiometric relationship between these components in the homeostatic autophagy process. We also investigated the expression of ATG16L1 in primary human immune cells, to gain a better understanding of its role in the association with CD (see Fig. 2D). Amongst samples tested, we found that ATG16L1 was most highly expressed in the T cell compartment, with both CD4+ and CD8+ cells showing high levels (approximately 13- and 10-fold higher than placental control RNA, respectively). CD19+ cells (B cells) also showed almost 5-fold greater expression of ATG16L1 than both mononuclear cells and placental control RNA. Thus it is possible that ATG16L1 and its variants play roles in both the epithelial- and immune-driven aspects of CD.

PHOX2B is known to be expressed in the autonomic nervous system and enteric neurons<sup>16</sup>. Attempts to determine expression using real-time RT-PCR were unsuccessful, most likely due to expression being confined to specific cell types and therefore overall transcript levels being low in the tissues studied. To overcome this limitation we obtained a specific antibody and stained mouse and human gut samples (ileum and colon respectively). These clearly showed that the PHOX2B protein was expressed in a specific subset of epithelial cells, possibly neuroendocrine cells (See Supplementary Fig. 3 online).

NCF4 encodes p40phox, a component of the NADPH oxidase complex required for optimal ROS generation in immune cells. Expression is restricted to haematopoietic cells including neutrophils, monocytes, eosinophils, mast cells and basophils (See Supplementary Figure 4

and reference<sup>19</sup>). The p40phox protein serves to enhance delivery of p47 and p67phox to the membrane and thus promotes high-level ROS generation. However, its role in cells that do not generate high-levels of ROS and lack expression of p47 and p67phox remains unknown.

### RNA-interference directed at ATG16L1 prevents autophagy

Although the yeast homologue of human ATG16L1 is required for autophagy<sup>20</sup>, it was not known whether the mammalian isoforms were similarly essential. We addressed this by utilizing oligo-based siRNA directed against ATG16L1 isoforms 1 and 2. Using co-transfection of a flag-tagged ATG16L1 expression plasmid and siRNA oligos we established specific knockdown of ATG16L1 in HEK293 cells (see Fig. 3A). In order to study the role of ATG16L1 in autophagy during a well-characterized host-pathogen interaction, we performed ATG16L1 knockdown in HeLa cells, followed by *S. Typhimurium* infection. We confirmed efficient knockdown of endogenous ATG16L1 in HeLa cells by real-time RT-PCR (see Fig. 3B). Since siRNA 2 was the most effective, this was used in subsequent experiments, co-transfected with a plasmid expressing GFP-LC3, a marker for the autophagic compartment<sup>21</sup>. It is known that a subpopulation of internalized salmonella are targeted by autophagy in this system<sup>22</sup>. Infections with *S. Typhimurium* showed a marked difference between control and siRNA 2 -treated cells: control cells targeted a mean of 17.5% (+/- 1.3 s.e.m.) of intracellular bacteria to autophagic LC3+ vacuoles within 1 hour, while siRNA 2 transfection reduced this to only 2% (+/- 0.5 s.e.m) (see Fig. 3C). The autophagic targeting rates in control-treated and ATG16L1-knockdown cells correspond well to those seen in ATG5<sup>+/+</sup> and ATG5<sup>-/-</sup> mouse embryonic fibroblasts (20% and 3%, respectively)<sup>22</sup>, confirming that, like ATG5, ATG16L1 is required for autophagy. Confocal imaging of infected cells revealed a near-complete lack of LC3 localisation with intracellular bacteria in ATG16L1-knockdown cells, despite the presence of abundant cytoplasmic LC3-GFP and complete envelopment of targeted bacteria in control cells (see Fig. 3D). In addition, we utilized classical autophagic stimuli to demonstrate the inhibitory effect of ATG16L1 knockdown (see Supplementary Fig. 5 online); both serum starvation and rapamycin treatment for 24 hours induced LC3-GFP vesicles only in control siRNA transfected cells, not those with ATG16L1 knockdown. Inhibition of lysosome fusion by ammonium chloride also induced rapid accumulation of LC3+ vesicles in control cells, but these structures were not found in ATG16L1-siRNA treated cells (see Supplementary Fig. 5 online).

## DISCUSSION

The potential of GWA studies to discover modest genetic risk factors in complex disease has been widely debated for several years. The simultaneous emergence of resources, such as the data from the International HapMap Project and dramatic leaps in genotyping technology (both clearly motivated by this potential), has now enabled the actual test of this hypothesis. In the case of CD, years of research effort by a substantial international community of researchers had identified only two uniformly replicated, genuinely associated, genetic risk factors for this disease (CARD15 and IBD5). Encouragingly, after only the first, moderately-sized GWA study and modest replication attempts, compelling evidence that this list has now grown substantially has been provided, as unequivocally strong evidence of replication has emerged from this study for variation at IL23R, ATG16L1 and an intergenic region on 10q. Moreover, the excess of additional low p-values in this study – in particular three additional top-ranking SNPs (in PHOX2B, NCF4, and FAM92B) which replicated at  $p < .01$  – along with the fact that other, large GWA attempts are ongoing, offers the potential that the list of replicated loci will grow even further in a short period of time. In fact, since the submission of the current study, reports describing a genome-wide association studies of 735 German patients with CD and of 547 Belgian patients with CD have been published<sup>23,24</sup>. In the German study, after association analyses of 7,159 nonsynonymous SNPs, the Ala197Thr in ATG16L1 was found to be

associated with CD, confirming this gene's involvement in CD susceptibility. In contrast to our findings, however, this group reports an epistatic interaction with CARD15. It should be noted, however, that rather than an aberrant joint frequency estimates in cases (as would be expected in the case of epistasis promoting a rare disease), the interaction signal in that study was mostly due to deviation from expectation in the control genotypes. Although this is possible for some genetic models, it requires a careful investigation. Furthermore, since different IBD genes show different degrees of phenotypic specificity, genes which promote the same specific sub-phenotype may be perceived as interacting if that specificity is not accounted for in the analyses. In the Belgian study, using the Illumina HumanHap300 Genotyping BeadChip, the authors identified an associated SNP (rs1373692) that implicated the prostaglandin receptor EP4 (PTGER4) gene in CD pathogenesis. In the current study, we also found evidence of association of the rs1373692 SNP with ileal CD ( $p=5.37E-04$ ), as well as to a SNP (rs4613763;  $p=1.21E-04$ ) less than 40kb away, providing the first replication of this finding. Full release and integration of results from these and other ongoing studies seems likely to add further to this burgeoning list of confirmed associations.

It is without doubt, however, that there are limitations to the GWA approach, just like any other genetic approach, and therefore complementary strategies for gene discovery will continue to be necessary<sup>25</sup>. One limitation is related to our ability to type large enough cohorts to have sufficient statistical power to detect loci of very modest genetic effect<sup>25</sup>. One example from our current study is the lack of genome-wide significant association of the IBD5 haplotype, despite the fact that this is a confirmed CD risk haplotype<sup>26</sup> and despite the fact that in a targeted replication study performed with the Replication Cohort #1 presented in the current study we find significant evidence of replication<sup>27</sup>. The strength of the association of the IBD5 in the current GWA study, it should be noted however, is consistent with modest risk conferred by this locus<sup>28</sup>. While recognizing the limitations inherent to the GWA study design, these current and future association findings also have the potential for dramatically changing our understanding of the biological processes that are crucial to disease pathogenesis, given the fact that genome-wide searches are not biased by our current understanding of pathogenic mechanisms. A prime example of this is our discovery that the ATG16L1 gene is associated with IBD, given that this gene, or the biological pathway of autophagy to which it belongs, was not previously implicated in IBD pathogenesis before the current generation of genome scans.

Autophagy is a constitutive process required for proper cellular homeostasis and organelle turnover. However, recent data has revealed a role for autophagy in innate and adaptive immune responses to pathogens<sup>29</sup>. This places autophagy as a gate-keeper for innate immune ligands and antigen presentation from intracellular compartments. So far the role of autophagy in innate immune recognition has been little studied, but several systems have been identified which show a potential role in host-defense<sup>30-32</sup>. We have exploited a model system using the invasive bacterium *S. Typhimurium* in cultured human epithelial cells<sup>22</sup> to demonstrate that in addition to classical autophagic stimuli, such as serum starvation and rapamycin, ATG16L1 is required for autophagic targeting of a subset of intracellular *Salmonella* exposed to the cytoplasm. While we do not propose that this represents the physiological stimulus involved in CD, we do believe it provides initial evidence that ATG16L1 variants with altered autophagic efficiency/efficacy might alter either bacterial replication and immune control, or delivery of antigens to adaptive immune pathways. For example, interactions between innate immunity, the inflammasome and autophagy have been proposed for *Legionella pneumophila*<sup>33</sup>. The proposed mechanism relies upon a NOD-LRR family member (Naip5) to detect invading *L. pneumophila* and subsequently activate autophagy, or if Naip5 signalling is sustained, proinflammatory cell death via caspase 1. Associations between the response to microbial products and chronic intestinal inflammation have been noted for flagellin, with both colitic mice and CD patients exhibiting elevated anti-flagellin serum IgG<sup>34</sup>. Since it is now known that TLR7-mediated detection of viral replication

is autophagy-dependant<sup>35</sup>, it is tempting to speculate that differential autophagic effectiveness might enhance defects in NOD-LRR family based signaling from intracellular pathogens.

Adaptive responses are also influenced by autophagic processing, since HLA class II molecule loading occurs in lysosomes and autophagy delivers cytoplasmic components into this compartment. Antigen loading via this pathway has been observed in professional antigen presenting cells as well as epithelial cells<sup>36</sup>. In cells with low levels of endocytosis, such as epithelial cells, this is likely to be an important route for antigen presentation and immune surveillance<sup>37</sup>. It is possible that differential rates or substrate specificities induced by variation within the autophagic apparatus will result in differential antigen presentation, as has been observed for B cells in starvation stress<sup>37</sup>. In many models of CD and other inflammatory disorders, T-cells are the primary effectors and autophagy is a vital process for T cell maintenance and homeostasis, as demonstrated by the increased cell death seen in ATG5<sup>-/-</sup> T cells and their inability to proliferate following stimulation<sup>38</sup>. T cells may also be stimulated directly by bacterial products in the gut microenvironment, demonstrated by the discovery of the I2 superantigen<sup>39</sup>. Given the role of autophagy in controlling T-cell death and proliferation, ATG16L1-mediated alterations in the proliferative and survival abilities of T cells provide a possible explanation for some of the T-cell driven pathology observed in CD. The relationships between T cell survival and regulation and the ATG16L1 variants discussed here will be an important area of further study.

Although our observation of potential association of the NCF4 gene to ileal CD is not unequivocally confirmed, the known functions of this gene warrant further examination. In particular, previous studies of the p40phox protein encoded by the NCF4 gene have demonstrated that it plays an important role in NADPH oxidase activity and the generation of reactive oxygen species (ROS) production upon phagocytosis<sup>17,18</sup>; both important for mounting an effective anti-microbial response. An alteration in gene expression could potentially have an effect on phagosome function, less effectively kill phagocytosed microbes, and could result in prolonged immune activation, incomplete pathogen clearance or might influence TLR activation and antigen presentation due to inappropriate ROS generation<sup>40,41</sup>. The constant exposure to bacterial challenge and high antigen load in the gut may pose special difficulties for even slightly altered innate immune responses, resulting in gut-restricted phenotypes. Such gut restriction has been previously observed for much more severe phox deficiencies in a number of chronic granulomatous disease (CGD) patients, whose initial symptoms resulted in a diagnosis of CD. Thus we propose that mutations in NCF4 might interfere with the stoichiometry of interaction with p67 and other p47phox interacting partners resulting in relatively moderate phagosome dysfunction phenotypes sufficient to trigger the onset of IBD.

Although additional functional studies are required to place all of these novel loci into a biological context that will enable a deeper understanding of their roles in normal homeostasis and disease pathogenesis, there can be little question that the *opportunity* to make advances in our understanding and eventual treatment of CD is being offered by the successful application of robust GWA studies.

## METHODS

### Subject ascertainment and diagnostic classification

The Screening Cohort used for this GWA study consisted of 988 patients with ileal CD and 1007 controls from the North American NIDDK IBD Genetics Consortium (IBDGC). Cases and geographically matched controls were ascertained through the Cedars-Sinai Medical Center, Johns Hopkins University, University of Chicago, University of Montreal, University of Pittsburgh, and the University of Toronto Genetics Research Centers (GRCs), with



additional age and ethnicity-matched controls provided by the New York Health project. In all cases, informed consent was obtained using protocols approved by the local institutional review board in all participating institutions. The diagnosis of IBD requires a) one or more symptoms of diarrhea, rectal bleeding, abdominal pain, fever, or complicated perianal disease, b) occurrence of symptoms on two or more occasions separated by at least 8 weeks or ongoing symptoms of at least 6 weeks duration, and c) objective evidence of inflammation from radiologic, endoscopic, and histologic evaluation. Ileal CD involvement was defined as mucosal ulceration, cobblestoning, stricturing or bowel wall thickening from endoscopy reports, barium X-rays, operative reports and/or pathology resection specimen reports. This was an inclusive definition where individuals with either “ileal only” or “ileo-colonic” were part of the “ileal CD” category. We previously reported the association between the IL23R gene and IBD from a genome-wide study of 567 CD patients and 571 controls<sup>11</sup>. These samples were included in the current study as well as an additional 421 ileal CD and 436 control individuals that were ascertained in an identical manner.

Replication Cohort #1 consisted of 883 nuclear families, which were independent of the Screening Cohort. These study subjects were collected by the same IBDGC Genetics Research Centers as described above, were previously described<sup>11</sup> and included 650 individuals with ileal CD and their parents. The Replication Cohort #1 was typed for 19/23 SNPs that had significant association ( $p < 5E-05$ ) in the GWA screen (the CARD15 and IL23R had previously been typed and two other SNPs failed at the assay design stage).

Replication Cohort #2 consisted of 978 patients with IBD (353 of which had ileal CD as defined above) and 207 control individuals that were ascertained as part of ongoing genetic studies at the Inflammatory Bowel Center at Cedars-Sinai Medical Center, Los Angeles, California. Recruitment of subjects has been approved by the Cedars-Sinai Medical Center Institutional Review Board. Replication Cohort #2 was typed for the six SNPs that had significant evidence of replication ( $p < 0.05$ ) in Replication Cohort #1 (rs2241880, rs16853571, rs224136, rs4821544, rs8050910, and rs11617463).

## Genotyping methods

For the genome-wide association study, approximately 750ng of genomic DNA was used to genotype each sample on the Illumina HumanHap300 Genotyping BeadChip (Illumina, San Diego) at the Feinstein Institute for Medical Research. Samples were processed according to the Illumina Infinium 2 assay manual. Briefly, each sample was whole-genome amplified, fragmented, precipitated and resuspended in appropriate hybridization buffer. Denatured samples were hybridized on prepared HumanHap300 beadchips for a minimum of 16 hours at 48°C. Following hybridization, the beadchips were processed for the single base extension reaction, staining and imaging on an Illumina Bead Array Reader. Normalized bead intensity data obtained for each sample was loaded into the Illumina Beadstudio 2.0 software which converted fluorescent intensities into SNP genotypes. The replication study of the all of the putative loci (Table 1), and the additional genotyping of the *ATG16L1* region was performed using primer extension chemistry and mass spectrometric analysis (iPlex assay, Sequenom, San Diego) using Sequenom Genetics Services (Sequenom, San Diego). TaqMan MGB technology was used to perform the genotyping of the six variants with nominal replication in replication study #1 using the design available from Applied Biosystems and following the manufacturer's recommendations (Bulletin #4317594) in the Cedars-Sinai Inflammatory Bowel Disease Center cohort. Quality of the SNP data was checked by reproducibility with 5% of samples duplicated.

## Analysis of GWA data

Overall, 98.6% of genotypes were called in this experiment; however, before analysis, quality filtering of both samples and SNPs were performed to insure robust association tests. Based on an evaluation of empirical distributions, data quality, and likely introduction of false positive associations, we required that samples pass a 93% genotyping call rate threshold and SNPs pass a 95% call rate threshold in order to be included in the analysis (see Supplementary Fig. 6 online). The data from the genome-wide association studies were used to detect possible relatedness in the case-control cohorts. The HWE test was performed on the genotype data from controls and we investigated the relationship between HWE and genotype yield (call rate). The call rate distribution suggested a 95% and 93% genotype completion threshold for inclusion of samples and SNPs, respectively in the genetic association analyses. At sample call rates of 90% and below, there was an elevation in observed heterozygosity suggesting either a bias in missing data or the presence of false heterozygous calls. To avoid this, we selected a threshold greater than 90 at a point consistent with the tail of successful samples. Below SNP success of 95% there was an excess of markers out of HWE (P-value <0.001 in controls) and a significant bias in missing data between cases and controls. We looked at segments of the data just 15 above and below this cutoff and observed that SNPs with a 90-95% call rate had much more substantial inflation than those in the 95-97% call rate range (genomic control correction 1.34 vs. 1.16) indicative of significant excess of false positives due to lower data quality/biases in missing data so we opted not to lower this threshold further. As a final step, using identity-by-state (IBS) counts from this data set, we identified 8 duplicate samples and 10 additional pairs of samples that shared 10% or more of their genome by descent (first cousins through full siblings) - we eliminated one member of each of these 18 pairs of samples. The data from the Jewish and non-Jewish case-control cohorts were analyzed jointly using a Cochran-Mantel-Haenszel chi-squared test; we used the test as implemented in R (<http://www.r-project.org/>) with the option that gives an exact p-value.

## Analysis of replication study data

For the replication study, the most significantly associated SNP from the top 23 ranking independently associated regions (p-value= $5 \times 10^{-5}$  in the combined screening analyses) were selected. Of these, 2 failed in assay design, one failed to type, and 2 were located within previously reported associated loci, leaving 18 SNPs in the replication analysis. In addition to these, 19 additional SNPs were selected to test the variability in the ATG16L1 region. This set of SNPs was evaluated for replication in a family-based ileal-CD cohort composed of 650 mother-father-affected offspring trios. The genotype data from the family-based cohorts were used in determining Mendelian inconsistencies and departures from HWE. After removal of failing SNPs (monomorphic, call rate below 75%, and causing a high number of Mendel errors), families with excess Mendelian inconsistencies and uninformative families (where a parent or the sole affected offspring had a genotyping call rate below 90%), there remained 530 affected trios (from 431 independent nuclear families) with both parents and at least one affected offspring genotyped. Following the removal of additional SNPs causing excess Mendel errors, the final genotyping call rate for each of the 18 replication markers, as well as the 19 ATG16L1 tagging SNPs, ranged from 90-100%, and Mendelian inconsistencies were below 4 for all markers and less than 4 per family. Single marker association tests were performed in the family data using PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/>). We then tested the replicated SNPs in the set of 530 independent IBD trios (post-QC) from the NIDDK IBDGC and in the independent Cedars-Sinai cohort of 350 patients with ileal CD and 207 controls, and onetailed p-values are reported. The analysis of ATG16L1 SNPs conditional on Ala197Thr was performed by logistic regression in the family-based sample using the software WHAP (<http://pngu.mgh.harvard.edu/~purcell/whap>).

### mRNA analysis of ATG16L1

RNA templates were purified from cell lines using the RNeasy miniprep kit (Qiagen). In the case of human primary immune cell subsets RNA was purchased from a commercial supplier (Human Blood Fractions MTC Panel, Clontech). In all cases cDNA was made from RNA templates using a standard reverse transcriptase protocol (iScript, BioRad), with 500 ng of RNA per reaction. The resulting template cDNAs were diluted 1:10 in nuclease-free water and 10  $\mu$ l of this template used for each 40  $\mu$ l qPCR reaction. Quantitative RT-PCR was performed using real-time SYBR Green method on a BioRad iQ5 thermal cycler, using iQ SYBR Green Supermix (BioRad) and specific primers. Primer sequences used were selected using MGH Primerbank (<http://pga.mgh.harvard.edu/primerbank/>); all primer sequences are shown in Supplementary Table 2. PCR products were visualized on a 2% agarose gel to confirm correct band sizes. Each reaction was performed in duplicate, with final calculations resulting from means of duplicate wells. Normalization for cDNA quantity was performed with GAPDH control primers for each template and final abundance figures adjusted to yield an arbitrary value of 1 for control templates (SW480 for cell lines or placental control RNA for primary cells) using the delta-delta Ct method.

### Immunohistochemistry for localization of PHOX2B in mouse and human gut tissues

Paraffin-embedded tissues from both mouse ileum and human colon biopsies were dewaxed, rehydrated, blocked in serum and stained using rabbit anti-PHOX2B antibody (Sigma), incubated overnight at a dilution of 1:500. Tissues were subsequently incubated with biotinylated anti-rabbit secondary antibody, Vector ABC reagent and finally developed with DAB substrate (Vector Laboratories) as has been previously described<sup>42</sup>. Sections were mounted and observed under phase contrast to visualize tissue architecture.

### Cell lines and bacterial strain

HeLa cells were obtained from ATCC and cultured in DMEM supplemented with 10% iron-supplemented calf serum (CSFe) (Hyclone) and 20  $\mu$ g ml<sup>-1</sup> gentamicin sulfate (Sigma). *Salmonella enterica* sbsp. *enterica* serovar Typhimurium (*S. Typhimurium*), strain SL1344 was transformed with a DsRed2 expression plasmid (Clontech, CA) (DsRed2 is under the control of the *lac* promoter, which is constitutive in the absence of *lacI*) to facilitate bacterial visualization. The plasmid was maintained with 100  $\mu$ g ml<sup>-1</sup> ampicillin selection during culture in LB broth.

### siRNA directed against ATG16L1

We obtained modified siRNA duplexes (Stealth siRNA) directed against ATG16L1 (siRNA 1 & 2) as well as a non-targeting scrambled sequence (siRNA control) from a commercial source (Invitrogen). Cells were co-transfected in the absence of antibiotics with siRNAs and plasmids using Lipofectamine (Invitrogen) according to the manufacturer's instructions. We optimized 17 siRNA conditions using a flag-tagged overexpression ATG16L1 construct in HEK293 cells, allowing us to easily assess knockdown by Western blotting. HEK293 cells grown in 12 well plates were transfected with 500 ng Flag-ATG16L1 expression plasmid and 20 pmol siRNA duplex. For knockdown of endogenous ATG16L1, HeLa cells were plated onto 18mm glass coverslips in 12-well plates at a density of  $1 \times 10^5$  well<sup>-1</sup> and allowed to grow for 24 hours prior to transfection. Each well received 500ng of GFP-LC3 plasmid and 20 pmol siRNA duplex. Four hours later the medium was changed again and the cells were allowed to grow for a further 48 hours. Knockdown was confirmed for each experiment using real-time quantitative RT PCR with specific primers to ATG16L1 and normalized to GAPDH controls, with RNA isolated from specimen wells.

## Induction of autophagy and bacterial infection

Autophagy was induced in HeLa cells using serum starvation or rapamycin. Cells were transfected as described above and after 48 hours the medium was changed to either 1% serum, or 10% serum plus 200 nM rapamycin for a further 24 hours. Ammonium chloride treatment was performed with a final concentration of 50 mM for 2 hours in 1% serum medium. *S. Typhimurium* infections were performed as previously described, with slight modifications<sup>43</sup>. Briefly, *S. Typhimurium* SL1344 carrying a DsRed2 expression plasmid was grown overnight in LB broth containing 100 µg ml<sup>-1</sup> ampicillin at 37°C with aeration and subcultured at a dilution of 1:33 for a further 3 hours in LB. This culture was further diluted in DMEM 10% CSFe without antibiotics, to yield an m.o.i. of 100. Infections were allowed to proceed for 20 minutes, cells were washed once in complete medium containing 100 µg ml<sup>-1</sup> gentamycin sulfate, and then incubated in fresh high-gentamycin medium for 1 hour.

## Microscopy

Infected, autophagy-induced and control HeLa cells were fixed in 4% formalin in PBS for 15 minutes, washed in PBS and mounted in aqueous mountant (Polysciences, PA). Slides were then either viewed under wide-field fluorescence illumination (Zeiss Axioplan) for counting, or laser scanning confocal microscopy (BioRad Radiance 2000) was used to obtain high-resolution z-stacks, which were subsequently projected onto single images using LSM Image software (Carl Zeiss GmbH, Germany). The total number of bacteria per cell, and the number of LC3-GFP positive bacteria were assessed in randomly chosen fields with at least 100 cells counted for each condition. The numbers of LC3-GFP positive bacteria were then calculated as a percentage of total bacteria. Significance was assessed using the two-tailed, unequal variance Student's T-test.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

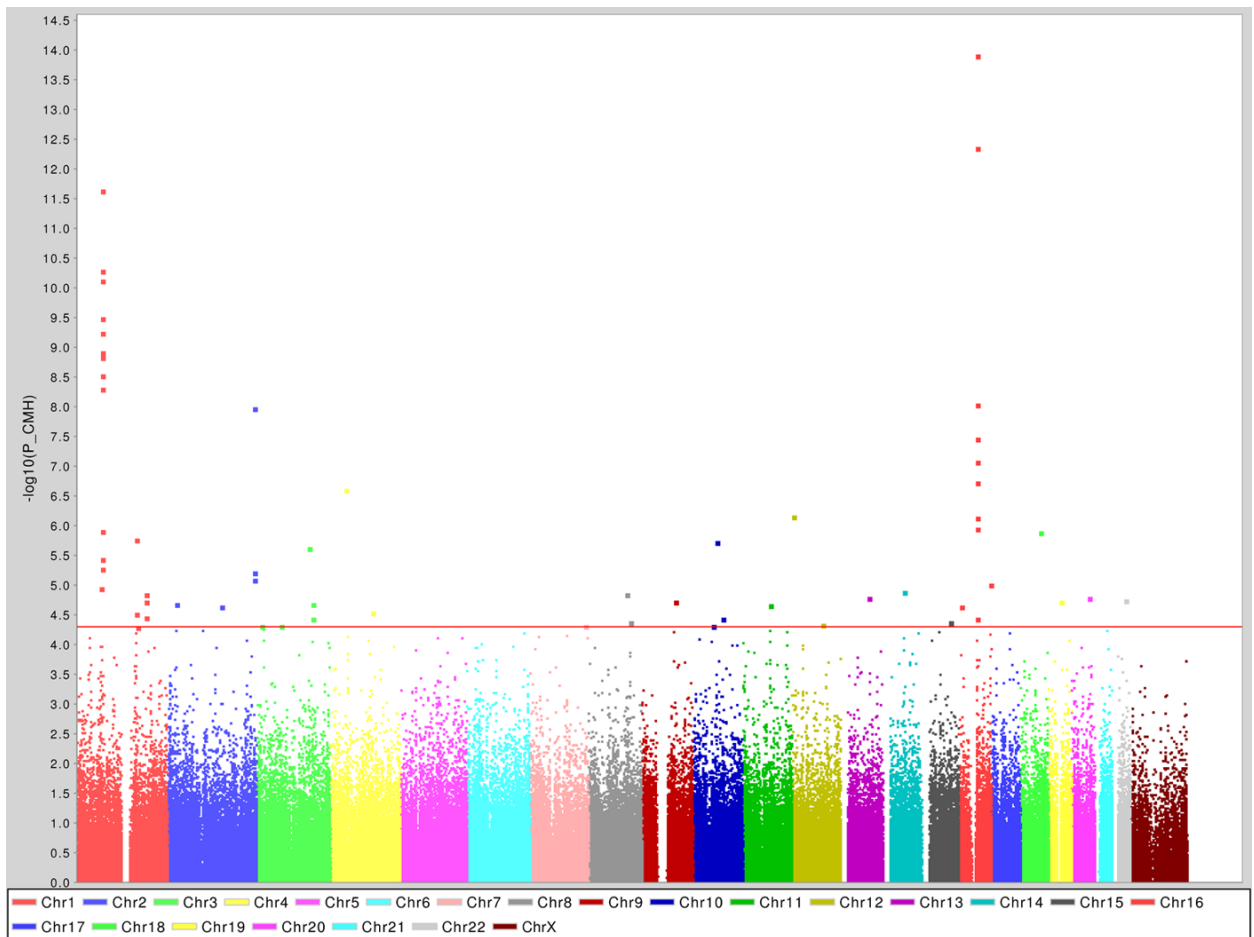
We thank the patients and their families for participating in these studies. We are grateful to all of the clinicians, research nurses and study coordinators for their essential contributions to the work. The authors would like to thank D. Caplan, G. Charron, C. Labbé, C. Lefebvre, and D. Miclaus for their help in the preparation of the manuscript, J. Adams for help with immunohistochemistry, A. Landry expression studies and to D. Altshuler for his critical reading of the manuscript. The National Institute of Diabetes and Digestive and Kidney Diseases IBD Genetics Consortium is funded by the following grants: DK62431 (S.R.B.), DK62420 (R.H.D.), DK62432 (J.D.R.), DK62423 (M.S.S.), DK62413 (K.D.T.), and DK62422 and DK62429 (J.H.C.). A.H.S. is on the Scientific Advisory Boards of Shire Pharmaceuticals, Schering (Canada), and Procter & Gamble Pharmaceuticals. The work on the Cedars-Sinai cohort was supported by project 1 of DK 46763 (J.I.R.). R.J.X. is supported by the following grants AI062773 and DK43351.

## REFERENCES

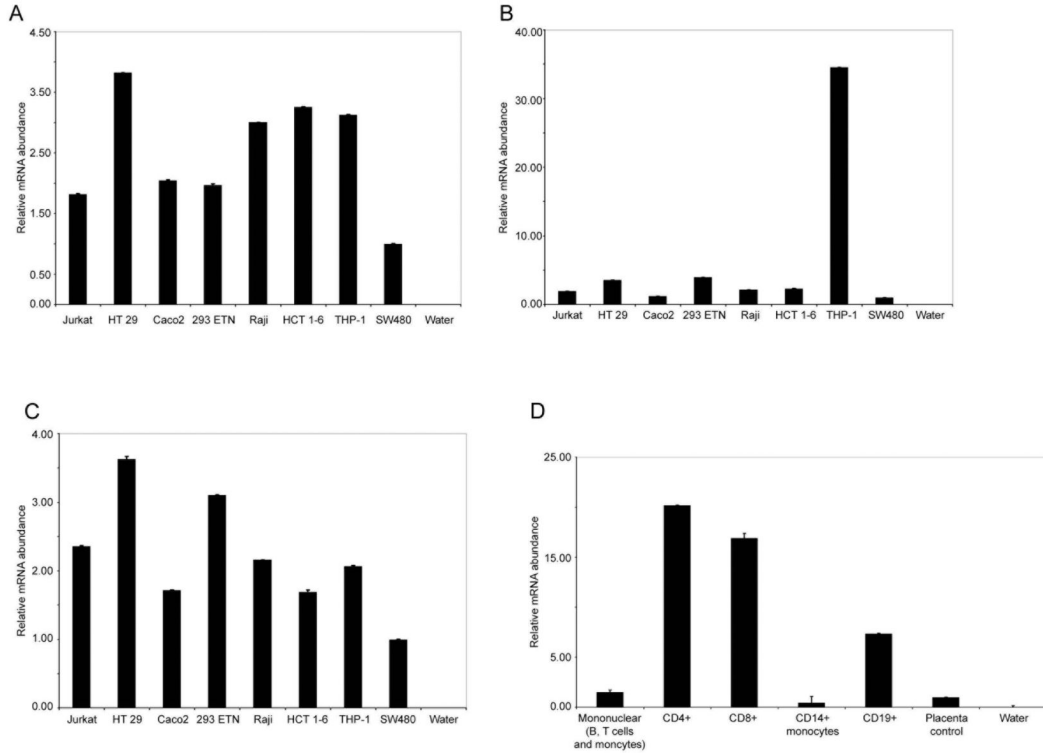
1. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 2004;126:1504–17. [PubMed: 15168363]
2. Podolsky DK. Inflammatory bowel disease. *N Engl J Med* 2002;347:417–29. [PubMed: 12167685]
3. Halfvarson J, Bodin L, Tysk C, Lindberg E, Jarnerot G. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 2003;124:1767–73. [PubMed: 12806610]
4. Thompson NP, Driscoll R, Pounder RE, Wakefield AJ. Genetics versus environment in inflammatory bowel disease: results of a British twin study. *Bmj* 1996;312:95–6. [PubMed: 8555939]
5. Tysk C, Lindberg E, Jarnerot G, Floderus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 1988;29:990–6. [PubMed: 3396969]

6. Hugot JP, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;411:599–603. [PubMed: 11385576]
7. Ogura Y, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;411:603–6. [PubMed: 11385577]
8. Rioux JD, et al. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001;29:223–8. [PubMed: 11586304]
9. Bayless TM, et al. Crohn's disease: concordance for site and clinical type in affected family members--potential hereditary influences. *Gastroenterology* 1996;111:573–9. [PubMed: 8780559]
10. Economou M, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JP. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am J Gastroenterol* 2004;99:2393–404. [PubMed: 15571588]
11. Duerr RH, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006;314:1461–3. [PubMed: 17068223]
12. Safford M, et al. Egr-2 and Egr-3 are negative regulators of T cell activation. *Nat Immunol* 2005;6:472–80. [PubMed: 15834410]
13. Cargnin F, et al. PHOX2B regulates its own expression by a transcriptional autoregulatory mechanism. *J Biol Chem* 2005;280:37439–48. [PubMed: 16144830]
14. Pasqualetti M, Rijli FM. Homeobox gene mutations and brain-stem developmental disorders: learning from knockout mice. *Curr Opin Neurol* 2001;14:177–84. [PubMed: 11262732]
15. Brunet JF, Pattyn A. Phox2 genes - from patterning to connectivity. *Curr Opin Genet Dev* 2002;12:435–40. [PubMed: 12100889]
16. Pattyn A, Morin X, Cremer H, Goridis C, Brunet JF. The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives. *Nature* 1999;399:366–70. [PubMed: 10360575]
17. Ellson CD, et al. Neutrophils from p40phox<sup>-/-</sup> mice exhibit severe defects in NADPH oxidase regulation and oxidant-dependent bacterial killing. *J Exp Med* 2006;203:1927–37. [PubMed: 16880254]
18. Suh CI, et al. The phosphoinositide-binding protein p40phox activates the NADPH oxidase during FcγRIIA receptor-induced phagocytosis. *J Exp Med* 2006;203:1915–25. [PubMed: 16880255]
19. Zhan S, et al. Genomic structure, chromosomal localization, start of transcription, and tissue expression of the human p40-phox, a new component of the nicotinamide adenine dinucleotide phosphate-oxidase complex. *Blood* 1996;88:2714–21. [PubMed: 8839867]
20. Kuma A, Mizushima N, Ishihara N, Ohsumi Y. Formation of the approximately 350- kDa Apg12-Apg5-Apg16 multimeric complex, mediated by Apg16 oligomerization, is essential for autophagy in yeast. *J Biol Chem* 2002;277:18619–25. [PubMed: 11897782]
21. Kabeya Y, et al. LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *Embo J* 2000;19:5720–8. [PubMed: 11060023]
22. Birmingham CL, Smith AC, Bakowski MA, Yoshimori T, Brumell JH. Autophagy controls Salmonella infection in response to damage to the Salmonella-containing vacuole. *J Biol Chem* 2006;281:11374–83. [PubMed: 16495224]
23. Hampe J, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 2007;39:207–11. [PubMed: 17200669]
24. Libioulle, C., et al. A novel susceptibility locus for Crohn's disease identified by whole genome association maps to a gene desert on chromosome 5p13.1 and modulates the level of expression of the prostaglandin receptor EP4. *PLoS Genet.* 2007.
25. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005;6:109–18. [PubMed: 15716907]
26. Daly MJ, Rioux JD. New approaches to gene hunting in IBD. *Inflamm Bowel Dis* 2004;10:312–7. [PubMed: 15290928]
27. Silverberg MS, et al. Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. *Eur J Hum Genet* 2007;15:328–35. [PubMed: 17213842]
28. Rioux JD, Abbas AK. Paths to understanding the genetic basis of autoimmune disease. *Nature* 2005;435:584–9. [PubMed: 15931210]

29. Deretic V. Autophagy in innate and adaptive immunity. *Trends Immunol* 2005;26:523–8. [PubMed: 16099218]
30. Amer AO, Byrne BG, Swanson MS. Macrophages rapidly transfer pathogens from lipid raft vacuoles to autophagosomes. *Autophagy* 2005;1:53–8. [PubMed: 16874021]
31. Nakagawa I, et al. Autophagy defends cells against invading group A *Streptococcus*. *Science* 2004;306:1037–40. [PubMed: 15528445]
32. Singh SB, Davis AS, Taylor GA, Deretic V. Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science* 2006;313:1438–41. [PubMed: 16888103]
33. Swanson MS. Autophagy: eating for good health. *J Immunol* 2006;177:4945–51. [PubMed: 17015674]
34. Lodes MJ, et al. Bacterial flagellin is a dominant antigen in Crohn disease. *J Clin Invest* 2004;113:1296–306. [PubMed: 15124021]
35. Lee HK, Lund JM, Ramanathan B, Mizushima N, Iwasaki A. Autophagy-dependent viral recognition by plasmacytoid dendritic cells. *Science* 2007;315:1398–401. [PubMed: 17272685]
36. Schmid D, Pypaert M, Munz C. Antigen-loading compartments for major histocompatibility complex class II molecules continuously receive input from autophagosomes. *Immunity* 2007;26:79–92. [PubMed: 17182262]
37. Dengjel J, et al. Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. *Proc Natl Acad Sci U S A* 2005;102:7922–7. [PubMed: 15894616]
38. Pua HH, Dzhagalov I, Chuck M, Mizushima N, He YW. A critical role for the autophagy gene *Atg5* in T cell survival and proliferation. *J Exp Med* 2007;204:25–31. [PubMed: 17190837]
39. Dalwadi H, Wei B, Kronenberg M, Sutton CL, Braun J. The Crohn's disease-associated bacterial protein I2 is a novel enteric T cell superantigen. *Immunity* 2001;15:149–58. [PubMed: 11485746]
40. Matsue H, et al. Generation and function of reactive oxygen species in dendritic cells during antigen presentation. *J Immunol* 2003;171:3010–8. [PubMed: 12960326]
41. Nakahira K, et al. Carbon monoxide differentially inhibits TLR signaling pathways by regulating ROS-induced trafficking of TLRs to lipid rafts. *J Exp Med* 2006;203:2377–89. [PubMed: 17000866]
42. Maison SF, Adams JC, Liberman MC. Olivocochlear innervation in the mouse: immunocytochemical maps, crossed versus uncrossed contributions, and transmitter colocalization. *J Comp Neurol* 2003;455:406–16. [PubMed: 12483691]
43. Beuzon CR, et al. *Salmonella* maintains the integrity of its intracellular vacuole through the action of SifA. *Embo J* 2000;19:3235–49. [PubMed: 10880437]

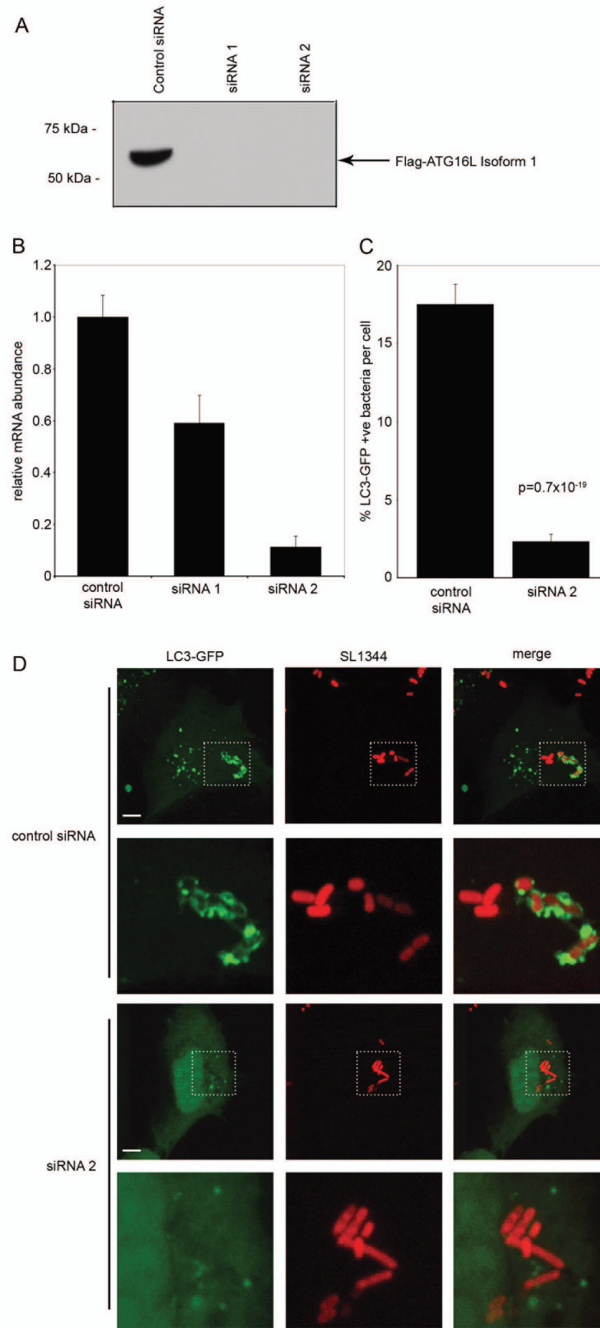


**Figure 1. Genome-wide association results for 946 ileal CD patient and 977 control samples**  
 Single marker association results for the combined Non-Jewish and Jewish cohorts using the Cochran-Mantel-Haenszel (CMH) test. Each chromosome is depicted as a different color and the blue and red lines indicate the suggestive and significant thresholds, respectively, based on a Bonferroni correction for multiple testing. A modest correction factor ( $\lambda$  1.06-1.07) is necessary to apply in order to correct for stratification between cases and controls, but does not modify the current results.



**Figure 2. Expression of autophagy components in human cell-lines and primary immune cells**  
 Quantitative real-time PCR was used to determine the expression patterns of ATG5, 7 and 16L1 in a variety of human immune and epithelial cell lines. Panel A - Expression pattern of ATG5 in human cell lines. ATG5 seems broadly expressed in the cell lines studied, with between 1.5- and 4-fold variations in expression level across the cell lines. Panel B - Expression pattern of ATG7 in human cell lines. ATG7 expression levels appeared broadly comparable to those of ATG5, with the exception of THP1 cells, which exhibited a 35-fold higher level than SW480 cells. Panel C - Expression pattern of ATG16L1 in human cell lines. ATG16L1 expression was also broadly similar to those of ATG5 and 7 (with the exception of ATG7 expression in THP1 cells) with moderate variations in expression levels across the cell lines tested (1.5- to 3.5-fold). Panel D - Expression pattern of ATG16L1 in human primary immune cells. In a resting human immune cell RNA panel (Clontech, CA) ATG16L1 showed a clear peak of expression in the T cell compartment, with both CD4+ and CD8+ cells expressing high levels of transcript (15- to 20-fold more than control samples). Modest levels were detected in CD19+ cells, and low levels in CD14+ monocytes and unfractionated mononuclear cells. In all cases, real-time quantitative RTPCR reactions were performed in duplicate and the means plotted, error bars represent 1 standard deviation. Reactions containing no template RNA were performed to control for reaction contamination (water controls). Expression levels were normalized by comparison to GAPDH controls and arbitrary relative expression units plotted where SW480 or placental RNA is equal to 1 (for cell lines, or primary cells respectively). All RNAs were isolated from resting cells in the absence of stimulation or activation.





**Figure 3. ATG16L1 is required for autophagy of *Salmonella Typhimurium***

**Panel A.** Specific siRNA knockdown of overexpressed, Flag-tagged ATG16L1 is achieved in HEK293 cells within 48 hours of transfection with oligo duplexes. Protein levels of Flag-ATG16L1 were undetectable by Western blotting following treatment with specific siRNA constructs, but expression was maintained with control duplexes.

**Panel B.** Endogenous ATG16L1 mRNA is efficiently knocked-down by siRNA 2 in HeLa cells within 48 hours of transfection. HeLa cells were transfected with siRNA duplexes and allowed to grow for 48 hours, RNA was then isolated and real-time quantitative RT-PCR performed. Data was normalized to GAPDH controls and siRNA control mRNA levels were set at 1.0. Compared to control duplex, siRNA 2 yielded an 89% reduction in mRNA transcript.

RT-PCR reactions were performed in triplicate and means are plotted with error bars representing 1 standard deviation, results are representative of two independent experiments.

**Panel C.** Knockdown of ATG16L1 prevents effective autophagy of *S. Typhimurium* in HeLa cells. 48 hours following co-transfection with control siRNA or duplex 2 and LC3-GFP plasmid, HeLa cells were infected for 1 hour with *Salmonella* Typhimurium SL1344, fixed and microscopically examined. The mean % of bacteria per cell encapsulated by LC3+ membranes (autophagosomes) and standard errors are shown. Bacterial counts were pooled from two separate experiments, each counting a minimum of 100 infected cells. Significance was assessed using a two-tailed Students T-test, assuming unequal variances.

**Panel D.** Representative images of intracellular *S. Typhimurium* show normal autophagic encapsulation by LC3+ membranes in control siRNA-treated cells, but loss of autophagic targeting when ATG16L1 knockdown is affected. HeLa cells were transfected with LC3-GFP plasmid and siRNA and subsequently infected as described above. Channels are separated to show both the LC3-GFP fusion protein (autophagosomal marker) and *S. Typhimurium* SL1344 dsRed2 (SL1344), with a merged image in the final column. Rows 2 and 4 show magnified views of the areas marked in rows 1 and 3 respectively. Images were obtained using confocal microscopy and are flat projection of Z-stacks taken throughout the plane of the infected cell. Scale bars represent 5  $\mu\text{m}$ .

Summary of the GWA study performed in 946 patients with ileal CD and 977 controls and the replication studies performed in 530 trios with ileal CD (Replication #1) and 350 cases with ileal CD and 207 controls (Replication #2)

TABLE 1

RANK	#SNPs	CHR	RS#	GWA			REPLICATION #1		REPLICATION #2		COMBINED REPLICATION		GENE
				MAF iCD	MAF CTL	P-VALUE	T	U	MAF iCD	MAF CTL	OR	P-VALUE	
1	8	16	rs2076756	0.358	0.244	7.01E-14	P.T.	P.T.	P.T.	P.T.	P.T.	P.T.	CARD15
2	13	1	rs7517847	0.295	0.403	3.06E-12	P.T.	P.T.	P.T.	P.T.	P.T.	P.T.	IL23R
3	3	2	<b>rs2241880</b>	0.364	0.453	6.38E-08	<b>220</b>	<b>306</b>	0.353	0.478	0.68	4.1 E-08	ATG16L1
4	1	4	<b>rs1685371</b>	0.038	0.077	7.68E-07	<b>39</b>	<b>75</b>	0.057	0.047	0.69	0.0084	PHOX2B
5	1	12	rs886898	0.156	0.102	1.93E-06	121	136	N.D.	N.D.	N.D.	N.D.	-
6	2	1	rs2343331	0.279	0.212	2.49E-06	Failed	Failed	N.D.	N.D.	N.D.	N.D.	-
7	1	18	rs937815	0.054	0.094	3.25E-06	96	99	N.D.	N.D.	N.D.	N.D.	-
8	1	3	rs6439924	0.218	0.160	6.00E-06	166	140	N.D.	N.D.	N.D.	N.D.	-
9	1	10	<b>rs224136</b>	0.134	0.191	7.90E-06	<b>94</b>	<b>149</b>	0.140	0.230	0.60	2.9 E-07	intergenic
10	1	9	rs10821091	0.399	0.332	1.44E-05	274	252	N.D.	N.D.	N.D.	N.D.	-
10	1	14	rs1188157	0.487	0.417	1.58E-05	254	240	N.D.	N.D.	N.D.	N.D.	-
12	1	1	rs2819130	0.177	0.126	2.10E-05	130	144	N.D.	N.D.	N.D.	N.D.	-
13	1	11	rs2712800	0.373	0.441	2.38E-05	242	222	N.D.	N.D.	N.D.	N.D.	-
14	1	22	<b>rs4821544</b>	0.397	0.333	2.89E-05	<b>267</b>	<b>221</b>	0.374	0.339	1.19	0.0090	NCF4
15	1	2	rs6733000	0.081	0.124	3.03E-05	81	77	N.D.	N.D.	N.D.	N.D.	-
16	1	2	rs7603516	0.064	0.102	3.10E-05	73	62	N.D.	N.D.	N.D.	N.D.	-
17	1	16	<b>rs8050910</b>	0.388	0.458	3.28E-05	<b>221</b>	<b>271</b>	0.400	0.430	0.84	0.0085	FAM92B
18	2	1	rs2490271	0.206	0.152	3.44E-05	175	166	N.D.	N.D.	N.D.	N.D.	-
19	1	20	rs4810663	0.236	0.180	3.45E-05	182	178	N.D.	N.D.	N.D.	N.D.	-
20	1	8	rs10505007	0.400	0.332	3.78E-05	221	248	N.D.	N.D.	N.D.	N.D.	-
21	1	8	rs2044999	0.330	0.395	3.84E-05	NT	NT	N.D.	N.D.	N.D.	N.D.	-
22	1	9	rs4878061	0.418	0.485	4.64E-05	NT	NT	N.D.	N.D.	N.D.	N.D.	-
23	1	13	<b>rs11617463</b>	0.044	0.077	4.85E-05	<b>59</b>	<b>80</b>	Failed	Failed	N.D.	N.D.	-

#SNPs= number of significant SNPs (P-value < 3e-05) in GWA analysis for the given locus; MAF iCD= minor allele frequency in ileal CD patients in GWA study and Replication #2; MAF CTL= minor allele frequency in controls in GWA study and Replication #2; T= number of transmissions of the minor allele in Replication #1; U= number of non transmissions of the minor allele in Replication #1; NT= not tested (e.g. failed to be designed); Failed= failed genotyping; P.T. = previously tested in these samples and having confirmed replication. N.B. P-values for the GWA results are two-tailed and for the combined replication study one-tailed. All 23 loci were included in Replication #1. All six loci that were replicated in Replication #1 - having P-value < 0.05 (shown in

**bold** – were included in Replication #2. Mantel-Haenszel ORs and p-values are reported for the combined replication data only. For additional detail regarding the results from the two replication studies, please see Supplementary Table 1 online.