Published in final edited form as:

Gut. 2011 August ; 60(8): 1060–1067. doi:10.1136/gut.2010.227256.

Variants in *ZNF365* isoform D are associated with Crohn's Disease

Talin Haritunians¹, Michelle R. Jones^{2,3}, Dermot P.B. McGovern^{1,3,4}, David Q. Shih⁴, Robert J. Barrett⁴, Carrie Derkowski⁴, Marla C. Dubinsky^{4,5}, Debra Dutridge¹, Phillip R. Fleshner⁴, Andrew Ippoliti⁴, Lily King¹, Esther Leshinsky-Silver⁶, Arie Levine⁷, Gil Y. Melmed⁴, Emebet Mengesha¹, Eric A. Vasilaukas⁴, Shabnam Ziaee^{3,4}, Jerome I. Rotter¹, Stephan R. Targan⁴, and Kent D. Taylor¹

¹Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA

²Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA

³Graduate Program in Biomedical Science and Translational Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA

⁴Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA

⁵Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA

⁶Molecular Genetic Lab, Wolfson Medical Center, Holon, Israel

⁷Pediatric Gastroenterology Unit, Wolfson Medical Center, Sackler School of Medicine, Tel-Aviv, Israel

Abstract

Objective—Genome-wide association studies (GWAS) have identified multiple Crohn's disease (CD) susceptibility loci, including association with non-coding intergenic single nucleotide polymorphisms (SNPs) at 10q21.

Design—To fine-map the 10q21 locus, we genotyped 86 SNPs in 1632 CD cases and 961 controls and performed single marker and conditional analyses using logistic regression.

Results—We observed association with CD risk spanning eleven SNPs ($p < 0.001$). The most significant association observed was at the nonsynonymous SNP rs7076156 (Ala62Thr) in *ZNF365*. The alanine allele was over-represented in CD ($p = 5.23 \times 10^{-7}$; OR = 1.39 [1.22–1.58]); allele frequency 76% CD, 69.7% controls). Conditional analysis on rs7076156 nullified all other significant associations, suggesting that this is the causative variant at this locus. Four isoforms of *ZNF365* have previously been identified and rs7076156 is located in an exon unique to *ZNF365* isoform D. We demonstrated, using RT-PCR, expression of *ZNF365D* in intestinal resections from both CD subjects and controls. We identified markedly reduced *ZNF365D* mean expression levels in EBV-transformed lymphoblastoid cell lines (ELCLs) from CD subjects homozygous for

Corresponding author: Talin Haritunians, Medical Genetics Institute, Cedars-Sinai Medical Center, 8700 Beverly Blvd., Los Angeles, CA 90048, phone: 310-423-3295; fax: 310-423-1323, Talin.Haritunians@cshs.org.

Competing Interest. None to declare

"I Talin Haritunians The Corresponding Author of this article (the Contribution") has the right to grant on behalf of all authors and does grant on behalf of all authors, a license to the BMJ Publishing Group Ltd and its licensees, to permit this Contribution (if accepted) to be published in *Gut* and any other BMJ Group products and to exploit all subsidiary rights, as set out in our licence set out at: (<http://gut.bmj.com/site/about/licence.pdf>)"

the risk allele (Ala). A whole-genome microarray expression study further suggested that the Ala62Thr change in *ZNF365* isoform D is related to differential expression of the genes *ARL4A*, *MKKS*, *RRAGD*, *SUMF2*, *TDR1* and *ZNF148* in CD.

Conclusions—Collectively our data support the hypothesis that the nonsynonymous Ala62Thr SNP rs7076156 underlies the association between 10q21 and CD risk and suggests that this SNP acts by altering expression of genes under the control of *ZNF365* isoform D.

Keywords

GWAS; Crohn's Disease; *ZNF365*; genetics; association study

Introduction

It is hypothesized that Crohn's disease (CD) results from dysregulation of the immune response to commensal gut bacterial flora in genetically susceptible individuals [1]. Linkage-based positional candidate gene studies, in addition to individual genome-wide association studies (GWAS) and GWAS meta-analyses, have identified over 30 genomic regions that contribute to the genetic susceptibility to Crohn's disease [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. For some of these regions, candidate genes for CD pathogenesis and a few putative causal variants have been identified [4, 5, 12], implicating pathways in innate immunity, autophagy, and adaptive immunity in the pathophysiology of Crohn's disease. Yet the causative variation for most regions remains unclear. Consequently, GWAS are the first step in a multi-layered process that includes locus specific fine-mapping strategies, expression profiling, and functional studies to identify the role of a particular variant in disease pathogenesis [13]. Model systems, such as immortalized cell lines or animal models of disease, will greatly contribute to these studies, particularly when obtaining blood or tissue from patients with defined genotypes is difficult.

Three GWAS performed in populations of Northern European descent have reported a CD susceptibility locus at 10q21.2 in a region 3' to the zinc finger protein 365 gene (*ZNF365*; dbGENE ID 22891) [7, 8, 10]. Rioux *et al.* reported the first CD association with rs224136, a single nucleotide polymorphism (SNP) approximately 40kb downstream from *ZNF365* [7]. CD association with this locus has been confirmed, although with different SNPs, in a report from the Wellcome Trust Case Control Consortium and in a meta-analysis of three GWAS [8, 10]. All three reported associations have been with non-coding intergenic SNPs located closer to the 3' of the coding region of *ZNF365* than the coding region of *EGR2*, another gene within this 10q21 locus. There is little evidence that these SNPs affect either expression or function of any genes and thus, causal variants remain to be identified.

Zinc finger protein 365 (*ZNF365*) was originally identified from a human brain cDNA library in 1998 [14] and four transcript variants have been reported (*ZNF365* isoforms AD; GenBank NM_014951, NM_199450, NM_199451, NM_199452) [15]. Expression of these isoforms varies across species. For example, *ZNF365A* is the only isoform observed in rodents but this isoform is highly conserved among higher order species, including humans. The *ZNF365A*/Su48 protein has been implicated as a centrosomal protein in mitosis and is expressed in human brain, heart, lung, and pancreas [15, 16]. In contrast, *ZNF365* isoform D appears to have emerged during primate evolution as it has not been identified in rodents, but is present in monkeys and has been detected in hominoids [15, 17]. Expression of the *ZNF365D*/Talanin protein has been shown in the placenta, lung, liver, kidney, and pancreas of humans [15], yet studies of this isoform are scant as this transcript is not represented on the available Affymetrix expression chips (NCBI, GEO datasets). A SNP in *ZNF365D*,

rs7076156, has been associated with uric acid nephrolithiasis (UAN) in an isolated population from Sardinia [15, 18].

The aim of this report was to characterize the previously identified association between CD and *ZNF365* through a fine-mapping approach since the association evidence was strongest for this gene at the time we began our study. Our results led to a further study of the expression of *ZNF365D* in intestinal tissues from both CD and non-IBD subjects, in addition to evaluating differences in gene expression in a model system of EBV-transformed lymphoblastoid cell lines from CD and non-IBD subjects harboring either the risk or protective genotype of a non-synonymous SNP specific to *ZNF365* isoform D.

Materials and Methods

Study Subjects

A total of 1632 CD cases and 961 non-IBD controls were included in these analyses. CD subjects were recruited at the Cedars-Sinai Medical Center Inflammatory Bowel Disease (CS-IBD) Center ($n=1590$) and Wolfson Medical Center, Holon, Israel ($n=42$) after diagnosis using standard clinical, endoscopic, and histological features [19]. Cases were predominantly of European descent (95.8%), with remaining subjects of Hispanic or Asian descent. Controls, also of European descent, were recruited through the CS-IBD Center ($n=295$; unrelated acquaintances and spouses of IBD patients with no personal or family history of IBD or autoimmune disease); as part of the Pharmacogenetics and Risk of Cardiovascular disease (PARC) Study ($n=479$), a pharmacogenetic study of statin response [20, 21]; or from the National Laboratory for the Genetics of Israeli Populations at Tel-Aviv University ($n=187$) (<http://nlgip.tau.ac.il>; Tel-Aviv, Israel). All cases and controls provided informed consent prior to study participation following approval of participating centers' institutional review boards.

EBV-transformed lymphoblastoid cell lines

Epstein Barr Virus-transformed B-cell lines (EBV-lymphoblastoid cell lines) are established for all subjects giving consent for participation in the ongoing genetic studies at the CS-IBD Center. These cell lines are established from peripheral blood lymphocytes following standard methods [22, 23]. For the studies reported here, cells were cultured in RPMI 1640 with L-glutamine and HEPES (Mediatech, Inc. Manassas, VA) supplemented with 20% Fetal Calf Serum (Omega Scientific, Inc., Tarzana, CA); 0.25 $\mu\text{g/ml}$ amphotericin B (Gemini Bio-Products, West Sacramento, CA); 50 $\mu\text{g/ml}$ gentamycin (Gemini Bio-Products); and $4.7 \times 10^{-5}\text{M}$ β -mercaptoethanol (Sigma-Aldrich, St. Louis, MO). Cells were maintained at 37°C in 5% atmospheric CO_2 . Total RNA was extracted using RNeasy Mini Kit (Qiagen, Valencia, CA).

SNP Selection

We applied a fine-mapping approach to the known isoforms of *ZNF365*, including the previously reported intergenic associations in the surrounding region, thus covering the region of chromosome 10 between 63,798,139 to 64,219,617 bp on human build 36 [7, 8, 10]. Using the software Tagger as implemented in Haploview [24, 25] and data from the CEPH population in the International HapMap project, release 2 [26, 27], we designed a panel of 86 SNPs that were compatible with the Illumina Infinium technology and tagged Caucasian haplotypes with a frequency greater than 5%. We further added non-synonymous SNPs with a minor allele frequency $>3\%$ in the Caucasian population.

Genotyping and Quality Control

All genotyping was performed at the genotyping laboratory of the General Clinical Research Center and Medical Genetics Institute at Cedars-Sinai Medical Center using custom iSelect Infinium technology (protocol from Illumina, San Diego, CA) [28, 29]. Samples with genotyping success rates <98% or with gender discrepancies were excluded from analyses. The average genotyping rate of samples retained in the analyses was 99.9%. Twenty samples performed in duplicate yielded 100% concordance for genotypes called. SNPs were excluded if the test of Hardy-Weinberg equilibrium across the entire sample was $p < 10^{-3}$; if the genotyping failure rate was >10%; if the minor allele frequency was <3%; or if the SNP had been selected for genotyping but was not found in the new dbSNP build at the time of analysis (dbSNP 129). These quality control measures left 78 SNPs for analyses. Genotyping for this particular study was performed as part of a larger project including a total of 6604 SNPs post-quality control.

Association and Conditional Analyses

Single marker analysis for association with case/control status was performed using logistic regression implemented in PLINK v1.06 [30]. Conditional logistic regression analysis was used to include allele load for the SNP being conditioned upon in the regression equation, and was performed using the *condition* function in PLINK.

Intestinal Biopsies for Tissue Specimens for Expression Studies

Intestinal tissues were obtained from one non-IBD control undergoing surgery for dysplasia (tissue was taken from non-dysplastic region of the bowel) and three individuals undergoing surgery for Crohn's disease (Table 1). Procedures for subject recruitment, informed consent, and specimen procurement were in accordance with protocols approved by the Institutional Review Board for Human Subject Protection of the Cedars-Sinai Medical Center. Lamina propria cell isolation was performed as previously described [31], and epithelial cells were obtained from intestinal tissue as described previously [32].

cDNA Synthesis and Reverse-Transcription PCR (RT-PCR)

Since ZNF365D had been reported to be expressed in kidney, commercially available total RNA extracted from pooled human adult whole kidney tissue (Agilent Stratagene, La Jolla, CA) was used as a positive control for ZNF365D expression. Tissue was stored in RNAlater following manufacturers protocol (Ambion, Austin, TX) until total RNA was extracted using the RiboPure Kit (protocol from Ambion, Austin, TX). cDNA was synthesized from 500 ng of total RNA using Oligo-dT primers and the Omniscript RT Kit (Qiagen, Valencia, CA). The presence of the ZNF365D isoform was detected in a standard PCR reaction using the AccuPrime Taq DNA Polymerase System (Invitrogen, Carlsbad, CA). A single amplicon band at the expected size (118 bp) was seen with the AccuPrime PCR Buffer-I and ZNF365 isoform D specific PCR primers (Forward - 5' GTT TGG CGT TGG CAG TCA GGT AAT 3' and Reverse - 5' CAC AGC ACG ACT CTG CAA GTG TAT 3' in exon 4; Integrated DNA Technologies, San Diego, CA). Thermal cycling conditions were 5 min at 95°; followed by 35 cycles of: 30 sec at 95°, 30 sec at 58°, 60 sec at 68°; and a final extension for 7 min at 72°. PCR products were analyzed on a 3% NuSieve 3:1 agarose gel (Lonza Inc., Rockland, MD).

Whole-Genome Expression

RNA was isolated from the EBV-transformed lymphoblastoid cell lines of control subjects with rs7076156 AA (n=3) or GG genotype (n=3) and of CD subjects with AA (n=3) or GG genotype (n=9). Whole expression data from the latter additional six CD-GG subjects were available to this project because of other experiments unrelated to this gene that were carried

out in the same gene expression run. cRNA amplification and biotin labeling was performed with 500 ng of total RNA from the EBV-transformed lymphoblastoid cell lines using an Illumina TotalPrep RNA amplification kit (protocol from Ambion, Austin, TX). cRNA yield was determined using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE) and 1.5 micrograms of biotin-labeled cRNA were hybridized to HumanWG-6 v3 Expression Beadchips (Illumina). This expression system contains approximately 48,800 50-mer probes; each probe is hybridized on average 20–60 times per sample. Data visualization and quality control were carried out with the Gene Expression module of GenomeStudio software (Illumina). Data were transformed by a variance-stabilizing method and normalized by quantile normalization [33, 34] (implemented in the *lumi* package of R v2.11.1) [35]. Data were carried forward into analysis if the detection signal for all samples within at least one of the four groups (control-AA, CD-AA, control-GG, CD-GG) was positive (p-value for detection <0.01). Thus, 14,541 probes remained for analysis. The association of RNA expression was analyzed using a two-way ANOVA for this two-by-two factorial design across the four groups, as implemented in MultiExperiment Viewer (TM4 microarray software suite) [36]. We observed six results with F-ratios greater than 30 in the expression experiment; this compared with one result observed using simulation and suggested that the false positive rate for this experiment was on the order of 1/6.

Quantitative real-time PCR (qPCR)

cDNA was synthesized using 1 µg of total RNA from the EBV-transformed lymphoblastoid cell lines and Oligo-dT primers (as described above; Omniscript RT Kit; Qiagen). ZNF365D and β-actin transcripts were amplified by qPCR using TaqMan probes and primers (ZNF365D Forward and Reverse primers in exon 4, same as above; ZNF365D probe: 5' /56-FAM/AAC TGG AAG CAA CCG TGA CGT TGT GCT /3BHQ_1/ 3'; β-actin primers: Forward - 5' CGT GTG GCT CCC GAG GAG 3' and Reverse 5' GGA TAG CAC AGC CTG GAT AGC 3'; β-actin probe: 5' /56-FAM/CCC CGT GCT GCT GAC CGA GGC / 3BHQ_1/ 3'; Integrated DNA Technologies). PCR was performed in duplicate, using thermal cycling conditions of: 2 min at 50°, 2 min at 95°, followed by 45 cycles: 15 sec at 95° and 1 min at 60°. Assays were performed following a modified TaqMan Gene Expression Assays protocol (Applied Biosystems, Foster City, CA), using Platinum qPCR SuperMix-UDG (Invitrogen) in a Mastercycler Realplex² (Eppendorf, Hamburg, Germany) and analyzed using Realplex software. The experiment was repeated three times in duplicate and amount of transcript was analyzed and expressed as the percentage of β-actin. F-ratio and p-values for the two-way ANOVA were calculated in R.

Results

Association of CD and ZNF365

A panel of 86 SNPs was designed using Tagger software and data from the International HapMap project to span the four known isoforms of *ZNF365* (Figure 1 **top panel**) and 78 of these SNPs were included in the final analysis of 1632 CD cases and 961 non-IBD controls (Figure 2 and Table S1). Eleven SNPs exhibited significant associations ($p < 0.001$) with CD risk, with the peak association observed at two SNPs, rs7076156 (OR= 1.39; $p = 5.23 \times 10^{-7}$) and rs7071642 (OR= 1.38; $p = 9.38 \times 10^{-7}$) (Figure 2 and Table 2). Furthermore, rs7076156 and rs7071642 retain significance following the Bonferroni correction for the 78 SNP tests related to the hypothesis of an association between CD and *ZNF365* (corrected p-values were $p = 4.23 \times 10^{-5}$ and $p = 7.6 \times 10^{-5}$, respectively), as well as for the 6604 SNP tests of the entire iSelect run ($p = 0.003$ and $p = 0.006$, respectively). To address whether inclusion of Jewish subjects could potentially influence our observed association, analyses were repeated to include self-identified Caucasian non-Jewish (NJ) subjects only (resulting in $n = 863$ CD

cases and $n=654$ non-IBD controls) and results remain unchanged following removal of Jewish subset (Table S2). Rs7076156 and rs7071642 are in complete linkage disequilibrium in our dataset (LD; $r^2=0.997$; $D'=1.0$; Figure 1 **bottom panel**). SNP rs7076156 is a non-synonymous SNP (G>A; Ala62Thr) in exon 4, an exon unique to *ZNF365* isoform D (Figure 1 **top panel**) [15]. The major allele (G or alanine allele) confers risk for CD (OR= 1.39; 95% CI [confidence interval] = 1.22–1.58), with an allelic frequency of 76% in patients with CD and 69.7% in controls. We also confirmed association of CD with the previously reported SNPs rs10761659 ($p=3.56\times10^{-4}$) and rs10995271 ($p=2.06\times10^{-4}$) [8, 10].

Conditional association

In order to determine whether the multiple associations we detected were due to the high LD structure observed in this region, and to identify the SNP or SNPs with the largest contribution to CD susceptibility, we performed a conditional analysis (logistic regression with the genotype of each SNP in turn used as a covariate in the analysis of the association with CD for all others). Conditioning on the most significantly associated SNP rs7076156 in both the entire sample size, as well as the NJ only subset, reduced all other CD associations to non-significance and regressed all odds ratios (OR) towards 1 (Table 2 and Table S2). A similar result was obtained when conditioning on rs7071642 (data not shown).

Detection of ZNF365 isoform D in IBD-related tissues

ZNF365 isoform D is expressed in human placenta, lung, liver, kidney, and pancreas [15] and we were interested to see whether the expression of this isoform could be observed in CD-related tissue. RT-PCR was performed to evaluate the expression of ZNF365D in whole human kidney (a positive control tissue) and in human intestine samples obtained by surgery. Details of these subjects are given in Table 1. We confirmed the previously reported expression of ZNF365D in kidney and also detected expression in a colon sample from a non-IBD control, and in samples from involved (inflamed) and uninvolved regions of the colon, the terminal ileum, and ileal epithelial cells from three CD patients (Figure 3).

Detection of ZNF365 isoform D in EBV-transformed lymphoblastoid cell lines

Because we have established EBV-transformed lymphoblastoid cell lines (ELCLs) for each CD subject and control that were recruited at the CS-IBD Center and provided consent for the genetic study, we next tested whether ZNF365 isoform D is expressed in these cell lines as a possible model system for the study of the function of genetic variation in *ZNF365D*. Total RNA was extracted from ELCLs from four groups: both control and CD subjects harboring either rs7076156 AA or GG genotypes, in a two-by-two factorial design. We employed a quantitative real-time PCR method (qPCR) and observed that ZNF365 isoform D is expressed in ELCLs (Figure 4). Moreover, ZNF365D expression levels were independently associated with disease status and genotype ($p_{\text{Two-way ANOVA}}=0.002$ and 0.007 , respectively). Interestingly, the mean expression level of ZNF365D in ELCLs obtained from CD subjects, as well as non-IBD controls, homozygous for the risk allele was appreciably lower than the level of expression observed in disease or control subjects with the protective AA genotype (Figure 4). In addition, we observed a general reduction in ZNF365D expression in CD cases compared to controls (Figure 4).

Whole genome expression study

We next tested the hypothesis that the identified *ZNF365D* variant was associated with variation in gene expression by performing a whole genome expression study. RNA from the ELCLs from four groups identified above was isolated and hybridized to whole genome expression chips. Data from this two-by-two factorial design ($n=3$ control-AA; $n=3$ CD-AA; $n=3$ control-GG; $n=9$ CD-GG subjects) were normalized, transformed, and analyzed by

ANOVA. Results with F-ratio greater than 30 for either genotype, disease, or the interaction between genotype and disease, are reported in Table 3 (complete list of differentially expressed genes [F-ratio >20] is given in Table S3). We observed six results shown with F-ratios greater than 30, with one result observed by simulation, suggesting the false positive rate for this experiment was 1/6. The hybridization signals from three β -actin probes in this experiment are also given for comparison. The amount of fluorescence for ZNF365D was not high enough for a positive detection score by this array. We observed an association between the *ZNF365D* rs7076156 genotype and altered expression of *RRAGD*, *SUMF2*, *TDRD1*, and *ZNF148*, and an association for interaction between genotype, disease status, and altered expression of *ARL4A* and *MKKS* (Table 3).

Discussion

We have conducted a fine-mapping study to follow up the well-documented association between CD susceptibility and 10q21. Our design was to genotype a region spanning all known isoforms of *ZNF365*, the gene closest to and within haplotype blocks spanning the previously reported single-nucleotide polymorphism (SNP) associations [7, 8, 10]. From our analysis panel of 78 SNPs, two SNPs (rs7076156 and rs7071642) were in perfect linkage disequilibrium and were associated with CD with the greatest significance (Figure 1 **bottom panel** and Figure 2). Several lines of evidence suggested that these SNPs are important for CD susceptibility, including (1) our observation that these SNPs account for the association signal in our dataset by conditional analyses (Table 2); (2) SNP rs7076156 is a non-synonymous alanine to threonine alteration in *ZNF365* isoform D; (3) our observation that *ZNF365* isoform D transcripts are expressed in the small intestine and colon obtained from both CD subjects and non-IBD control (Figure 3); and (4) expression of *ZNF365D* is reduced in cell lines from CD subjects, with a further reduction observed with the associated risk genotype (Figure 4).

The major allele of rs7076156 confers risk for CD (OR= 1.39; 95% CI= 1.22–1.58), with a frequency of 76% in cases and 69.7% in controls. By conditional analyses, we observed that rs7076156 accounts for the observed association of other SNPs in the immediate region, including those previously reported to be associated with CD and replicated in this study (Table 2; rs10761659, $p= 3.56 \times 10^{-4}$; rs10995271, $p= 2.06 \times 10^{-4}$) [8, 10, 37]. Furthermore, while there was also LD between rs7076156 and the previously reported SNPs rs10995271 and rs10761659 ($r^2= 0.20$ and 0.38 , $D'=0.88$ and 0.91 , respectively), our conditional analyses suggested that this variant accounted for the associations of these previously identified SNPs. We did not confirm an association between CD and a reported CD susceptibility locus within this region, rs224136 (Table S1) [7]. Our observation is consistent with a German study that failed to demonstrate an association for this SNP with either CD or ulcerative colitis (UC) [38]. Recently, rs10995271 was shown to be associated with pediatric CD [39].

More recently, a polymorphism (rs10761670) further downstream (10q21.3) in the 5' flanking region of *EGR2* has been shown to be associated with susceptibility to systemic lupus erythematosus and rheumatoid arthritis [40]. In addition, two SNPs (rs1412554 and rs1509957) in complete LD with rs10761670 were demonstrated to affect the binding of transcription factors and modulate gene expression in *in vitro* luciferase assays [40], suggestive of a potential role for *EGR2* in autoimmune disease. However, rs7076156 in *ZNF365D* was not in strong LD ($r^2= 0.005$; $D'= 0.134$) with the SNPs reported by Myouzen *et al.* Furthermore, two SNPs (rs1475084 and rs10509173) in our genome-wide dataset [11] (data not shown) in strong LD ($r^2> 0.9$; $D'= 1.0$) with rs10761670, rs1412554, and rs1509957 did not exhibit an association with CD ($p< 0.2$).

We have also observed that EBV-transformed lymphoblastoid cell lines from both CD and non-IBD subjects express ZNF365 isoform D and are thus a potential model system for the further study of this variant. Interestingly, disease status and genotype were independently associated with differences in the level of ZNF365D expression, as (1) ELCLs from CD patients demonstrated lower levels of ZNF365 isoform D expression compared to non-IBD controls, and (2) CD patients harboring the risk GG genotype exhibited lower levels of expression compared to CD subjects with the protective genotype (Figure 4). As shown in Figure 4, the GG genotype of rs7076156 is associated with a reduction in ZNF365D expression when compared with the AA genotype in both controls as well as cases, and further reduction is observed when cell lines from CD patients are compared with cell lines from controls. This evidence supports the hypothesis that expression of the *ZNF365D* gene is reduced by the rs7076156 GG genotype, as well as other genetic factors that are operational in cell lines from CD cases when compared with controls. Furthermore, our genome-wide expression study suggests that these variants, along with disease status, are also associated with significant differences in the expression of other genes (Table 3). Since zinc finger proteins such as ZNF365 may act as transcription factors, our study supports the concept that the SNPs reported here, an intronic SNP (rs7071642) or an amino acid change functional variant (rs7076156), or both, result in alteration of the expression of various genes, including *ARL4A*, *MKKS*, *RRAGD*, *SUMF2*, *TDRD1*, and *ZNF148*, another transcription factor (Table 3). In hominids, ZNF148 (ZBP-89) exists as both a full-length protein and an alternatively spliced isoform lacking the amino terminus (ZNF148-delN) [41]. Interestingly, in human colorectal cell lines, the N-terminal domain of full-length ZNF148 has been shown to interact with histone acetyltransferase p300 during butyrate activation of the cyclin-dependent kinase inhibitor, p21^{waf1} [42], a protein with increased levels in colonic crypts of UC patients [43]. Furthermore, butyrate suppresses DSS-induced colonic inflammation [44]. Moreover, mice homozygous for ZNF148-delN exhibit increased susceptibility to DSS-induced colitis, further highlighting a role for ZNF148 in gastrointestinal homeostasis [41]. Taken together with our demonstration of the association of *ZNF365D* genotype with altered expression of ZNF148, these observations support a role for ZNF365D in gastrointestinal homeostasis through ZNF148. However, further studies are necessary to demonstrate the role of these molecules in CD pathogenesis more directly.

A complex pattern of alternative splicing for *ZNF365* has been suggested [17]. While the *ZNF365A-C* isoforms all utilize the same CpG island promoter, the *ZNF365D* isoform uses its own specific TATA-box promoter located in intron 7 of the full transcript [15]. Isoform *ZNF365D* likely first appeared as a gene transcript, even if not as a functional gene, in New and Old World monkeys [17]. *ZNF365D* is predicted to be an integral membrane protein with potential N- and O-linked glycosylation sites in the N-terminus, and *in silico* modeling has suggested that the functional rs7076156 variant (Ala62Thr) results in a modification of the secondary structure of the protein, as one of the alpha helices is ablated when the threonine allele is present [15]. Nevertheless, the function of the protein encoded by this gene is thus far not well defined.

In summary, using information from the Human Genome Project and the Human HapMap Project in our fine-mapping approach, we have followed up a CD susceptibility locus originally identified in several GWAS. We provide evidence that *ZNF365* is the gene for CD susceptibility in this region, that *ZNF365* isoform D is expressed in intestine, and that rs7076156 is the putative causative variant for CD within this gene. Utilizing ELCLs from CD and non-IBD subjects, we have further demonstrated that CD patients express reduced ZNF365D, and that the risk variant leads to a further reduction in expression. Finally, we have observed this variant to be associated with significant differences in the expression of several genes, including *ZNF148*, a potential regulator of gastrointestinal homeostasis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported in part by NCCR grant M01-RR00425 to the Cedars-Sinai General Research Center Genotyping core; NIH/NIDDK grant P01-DK046763; the Southern California Diabetes Endocrinology Research Center grant, DK063491; Cedars-Sinai Medical Center Inflammatory Bowel Disease Research Funds; the Feintech Family Chair in IBD (SRT); the Abe and Claire Levine Chair in Pediatric IBD (MD) and the Cedars-Sinai Board of Governors' Chair in Medical Genetics (JIR); with additional funding through grants DK062413 (DPBM), DK76984 (MD), and DK084554 (MD and DPBM). We would like to thank Dr. Ronald M. Krauss from the Children's Hospital of Oakland Research Institute for permission to use subjects from the Pharmacogenetics and Risk of Cardiovascular disease (PARC) Study (NIH/NHLBI grant U01-HL069757) as controls for our study.

Abbreviations

GWAS	Genome-wide association studies
CD	Crohn's disease
SNP	single nucleotide polymorphism
ELCLs	EBV-transformed lymphoblastoid cell lines
ZNF365	Zinc finger protein 365
ZNF365D	Zinc finger protein 365 isoform D
UAN	uric acid nephrolithiasis
CS-IBD	Cedars-Sinai Medical Center Inflammatory Bowel Disease
PARC	Pharmacogenetics and Risk of Cardiovascular disease
RT-PCR	Reverse-Transcription PCR
qPCR	Quantitative real-time PCR
UC	ulcerative colitis
NJ	non-Jewish

References

- Shih DQ, Targan SR. Immunopathogenesis of inflammatory bowel disease. *World J Gastroenterol*. 2008; 14:390–400. [PubMed: 18200661]
- Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*. 2001; 411:599–603. [PubMed: 11385576]
- Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*. 2001; 411:603–6. [PubMed: 11385577]
- Yamazaki K, McGovern D, Ragoussis J, et al. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet*. 2005; 14:3499–506. [PubMed: 16221758]
- Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006; 314:1461–3. [PubMed: 17068223]
- Hampe J, Franke A, Rosenstiel P, 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]
- 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]

8. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–78. [PubMed: 17554300]
9. Libioulle C, Louis E, Hansoul S, et al. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet*. 2007; 3:e58. [PubMed: 17447842]
10. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008; 40:955–62. [PubMed: 18587394]
11. McGovern DP, Jones MR, Taylor KD, et al. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Hum Mol Genet*. 2010
12. McGovern DP, Rotter JI, Mei L, et al. Genetic epistasis of IL23/IL17 pathway genes in Crohn's disease. *Inflamm Bowel Dis*. 2009; 15:883–9. [PubMed: 19235914]
13. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet*. 2008; 17:R156–65. [PubMed: 18852205]
14. Nagase T, Ishikawa K, Suyama M, et al. Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Research*. 1998; 5:355–64. [PubMed: 10048485]
15. Gianfrancesco F, Esposito T, Ombra MN, et al. Identification of a novel gene and a common variant associated with uric acid nephrolithiasis in a Sardinian genetic isolate. *Am J Hum Genet*. 2003; 72:1479–91. [PubMed: 12740763]
16. Hirohashi Y, Wang Q, Liu Q, et al. Centrosomal proteins Nde1 and Su48 form a complex regulated by phosphorylation. *Oncogene*. 2006; 25:6048–55. [PubMed: 16682949]
17. Gianfrancesco F, Esposito T, Casu G, et al. Emergence of Talanin protein associated with human uric acid nephrolithiasis in the Hominidae lineage. *Gene*. 2004; 339:131–8. [PubMed: 15363853]
18. Ombra MN, Forabosco P, Casula S, et al. Identification of a new candidate locus for uric acid nephrolithiasis. *Am J Hum Genet*. 2001; 68:1119–29. [PubMed: 11309680]
19. Mow WS, Vasiliauskas EA, Lin YC, et al. Association of antibody responses to microbial antigens and complications of small bowel Crohn's disease. *Gastroenterology*. 2004; 126:414–24. [PubMed: 14762777]
20. Simon JA, Lin F, Hulley SB, et al. Phenotypic predictors of response to simvastatin therapy among African-Americans and Caucasians: the Cholesterol and Pharmacogenetics (CAP) Study. *Am J Cardiol*. 2006; 97:843–50. [PubMed: 16516587]
21. Krauss RM, Mangravite LM, Smith JD, et al. Variation in the 3-hydroxyl-3-methylglutaryl coenzyme A reductase gene is associated with racial differences in low-density lipoprotein cholesterol response to simvastatin treatment. *Circulation*. 2008; 117:1537–44. [PubMed: 18332269]
22. Anderson MA, Gusella JF. Use of cyclosporin A in establishing Epstein-Barr virus-transformed human lymphoblastoid cell lines. *In Vitro*. 1984; 20:856–8. [PubMed: 6519667]
23. Pressman S, Rotter JI. Epstein-Barr virus transformation of cryopreserved lymphocytes: prolonged experience with technique. *Am J Hum Genet*. 1991; 49:467. [PubMed: 1651051]
24. De Bakker, PI. Tagger. 2004. <http://www.broad.mit.edu/mpg/tagger/>
25. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21:263–5. [PubMed: 15297300]
26. International HapMap Consortium. The International HapMap Project. *Nature*. 2003; 426:789–96. [PubMed: 14685227]
27. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 31 million SNPs. *Nature*. 2007; 449:851–61. [PubMed: 17943122]
28. Gunderson KL, Kuhn KM, Steemers FJ, et al. Whole-genome genotyping of haplotype tag single nucleotide polymorphisms. *Pharmacogenomics*. 2006; 7:641–8. [PubMed: 16768648]
29. Gunderson KL, Steemers FJ, Ren H, et al. Whole-genome genotyping. *Methods Enzymol*. 2006; 410:359–76. [PubMed: 16938560]
30. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–75. [PubMed: 17701901]

31. Prehn JL, Mehdizadeh S, Landers CJ, et al. Potential role for TL1A, the new TNF-family member and potent costimulator of IFN-gamma, in mucosal inflammation. *Clin Immunol.* 2004; 112:66–77. [PubMed: 15207783]
32. Panja A, Barone A, Mayer L. Stimulation of lamina propria lymphocytes by intestinal epithelial cells: evidence for recognition of nonclassical restriction elements. *J Exp Med.* 1994; 179:943–50. [PubMed: 8113686]
33. Lin SM, Du P, Huber W, et al. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.* 2008; 36:e11. [PubMed: 18178591]
34. Quackenbush J. Microarray data normalization and transformation. *Nat Genet.* 2002; 32 (Suppl): 496–501. [PubMed: 12454644]
35. Team RDC. Computing RFFS. Vienna, Austria: 2006. R: A language and environment for statistical computing.
36. Saeed AI, Bhagabati NK, Braisted JC, et al. TM4 microarray software suite. *Methods Enzymol.* 2006; 411:134–93. [PubMed: 16939790]
37. Franke A, Balschun T, Karlsen TH, et al. Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. *Nat Genet.* 2008; 40:713–5. [PubMed: 18438405]
38. Glas J, Seiderer J, Pasciuto G, et al. rs224136 on chromosome 10q21. 1 and variants in PHOX2B, NCF4, and FAM92B are not major genetic risk factors for susceptibility to Crohn's disease in the German population. *Am J Gastroenterol.* 2009; 104:665–72. [PubMed: 19262523]
39. Amre DK, Mack DR, Morgan K, et al. Susceptibility loci reported in genome-wide association studies are associated with Crohn's disease in Canadian children. *Aliment Pharmacol Ther.* 2010.1111/j.1365-2036.2010.04294.x
40. Myouzen K, Kochi Y, Shimane K, et al. Regulatory polymorphisms in EGR2 are associated with susceptibility to systemic lupus erythematosus. *Hum Mol Genet.* 2010; 19:2313–20. [PubMed: 20194224]
41. Law DJ, Labut EM, Adams RD, et al. An isoform of ZBP-89 predisposes the colon to colitis. *Nucleic Acids Res.* 2006; 34:1342–50. [PubMed: 16517939]
42. Bai L, Merchant JL. Transcription factor ZBP-89 cooperates with histone acetyltransferase p300 during butyrate activation of p21waf1 transcription in human cells. *J Biol Chem.* 2000; 275:30725–33. [PubMed: 10899165]
43. Hofseth LJ, Saito S, Hussain SP, et al. Nitric oxide-induced cellular stress and p53 activation in chronic inflammation. *Proc Natl Acad Sci U S A.* 2003; 100:143–8. [PubMed: 12518062]
44. Moreau NM, Champ MM, Gouptry SM, et al. Resistant starch modulates in vivo colonic butyrate uptake and its oxidation in rats with dextran sulfate sodium-induced colitis. *J Nutr.* 2004; 134:493–500. [PubMed: 14988436]

Summary Box

What is already known about this subject?

- To date, genome-wide association studies and meta-analyses have identified greater than 30 loci that contribute to the genetic susceptibility of Crohn's disease.
- Some putative causal variants have been identified, however a causative variant for most associated loci remains unclear.
- An association between Crohn's disease and intergenic variants at 10q21 has been established beyond doubt.

What are the new findings?

- The major allele (G) of a nonsynonymous SNP rs7076156 in *ZNF365* isoform D in the susceptibility locus at 10q21.2 is strongly associated with Crohn's disease risk.
- This is the first report demonstrating *ZNF365* isoform D expression in the colon of Crohn's disease subjects and a non-IBD control.
- *ZNF365* isoform D is further expressed in EBV-transformed lymphoblastoid cell lines established from patients and controls. Interestingly, we demonstrate lower levels of expression in Crohn's disease cases than controls, with a further decrease in expression in those Crohn's disease patients harboring the risk (GG) genotype.
- Association of *ZNF365* rs7076156 genotype and altered expression of several genes, including *ZNF148*, a transcriptional factor known to be involved in gut homeostasis, is demonstrated.

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

- Identification of 'causative' loci increases our understanding of CD pathogenesis and can identify novel pathways involved in disease etiology.
- An increased understanding of disease etiology highlights potential novel therapeutic targets for CD.

Figure 1. *ZNF365* Isoforms and Linkage Disequilibrium Between Genotyped SNPs

Top panel: Four reported isoforms of *ZNF365* (A–D: GenBank NM_014951, NM_199450, NM_199451, NM_199452, respectively). The asterisks indicate the positions of rs7071642 (intronic; light grey) and rs7076156 (in exon 4; black) of *ZNF365* isoform D. Bottom panel: Linkage disequilibrium (D') between SNPs in the analysis panel was calculated in all data (both CD and controls) combined using Haploview [25]. Region encompassing four known *ZNF365* isoforms (A–D) is noted. rs7076156 (in black) and rs7071642 (immediately adjacent in light grey) are each marked with an asterisk, as are the top hits reported for Barrett *et al.* [10], WTCCC *et al.* [8], and Rioux *et al.* [7].

Figure 2. Association of *ZNF365* and Crohn's disease

Logistic regression analysis of CD and *ZNF365*. Plot of position across chromosome 10q21 vs. $-\log_{10}(\text{p-value})$ from logistic regression as described in Methods. SNPs are represented as red triangles. Locations and p-values of previously reported associations are noted in black; locations and p-values of rs7076156 and rs7071642 are noted in blue. LD heat map indicates level of linkage disequilibrium (r-squared) between SNPs. Chromosomal position of leading and trailing SNPs, and location of *ZNF365D* within this region, are noted. (Generated with R package *snp.plotter*; <http://cbdb.nimh.nih.gov/~kristin/snp.plotter.html>)

Figure 3. RT-PCR analysis of ZNF365D transcripts

ZNF365D expression was detected in kidney and four independent intestinal biopsies: (1) uninvolved tissue from CD subject (Colon: CD U); (2) involved tissue from CD subject (Colon: CD I); (3) tissue from a non-IBD subject (Colon: non-IBD); as well as (4) terminal ileum whole tissue (Ileum: CD), including epithelial fraction (Ileum Epithelial: CD) and lamina propria fraction (Ileum LPMC: CD), from CD subject. LPMC = lamina propria mononuclear cells.

Disease Status		Genotype
Control	CD	
0.035 +/- 0.0083	0.021 +/- 0.007	AA
0.021 +/- 0.0053	0.004 +/- 0.0065	GG

Figure 4. Mean level of ZNF365D Expression in EBV-transformed lymphoblastoid cell lines
 ZNF365D mean expression levels (expressed as % β -actin) for control or CD subjects, harboring either rs7076156 AA (protective) or GG (risk) genotypes. These are depicted as a 3D plot (above), with mean values $\pm$ S.E. specified below. Association between ZNF365D expression level and disease status: F-ratio= 10.07; p-value Two-way ANOVA= 0.0022. Association between ZNF365D expression level and genotype: F-ratio= 7.58; p-value Two-way ANOVA= 0.0074. No association between ZNF365D expression level and interaction of disease status with genotype (F-ratio= 0.034; p-value Two-way ANOVA= 0.85).

Table 1

Intestinal Biopsies for Tissue Specimens for Reverse-Transcribed PCR

	Gender	Age	Ethnicity	Tissue	Biopsy Location	Reason for Surgery	Current Medication	Family History
CD	F	51	Caucasian	Uninvolved	Colon	CD resistant to therapy	None	None
CD	M	44	Hispanic	Involved	Colon	CD resistant to therapy	Prednisone	Autoimmune disease
Non-IBD	M	76	Caucasian	Normal	Colon	(A)	(A)	(A)
CD	(A)	(A)	Caucasian	Involved	Terminal Ileum	Small bowel stricture secondary to CD	Adalimumab	Autoimmune disease

(A)= Anonymized sample

Table 2

SNPs in 10q21.2 Associated with CD

Chr	SNP	Position	Risk Allele	P-value	Odds Ratio	Conditioned P-value	Conditioned Odds Ratio	Significance in Previous Reports	Gene *
10	rs10740085	64065801	G	4.33E-06	1.39	0.43	1.10		ZNF365
10	rs12768538	64076364	G	4.21E-04	1.24	0.18	1.10		ZNF365
10	rs7068361	64078373	A	2.04E-04	1.24	0.58	1.04		ZNF365
10	rs7071642	64084066	A	9.38E-07	1.38	NA	NA		ZNF365
10	rs7076156	64085190	G	5.23E-07	1.39	NA	NA		ZNF365
10	rs729739	64100308	G	7.38E-05	1.30	0.19	0.81		ZNF365
10	rs10995271	64108492	G	2.06E-04	1.25	0.08	1.12	1.90E-11 ¹	Intergenic
10	rs12766391	64111210	A	1.17E-04	1.26	0.06	1.14		Intergenic
10	rs10761659	64115570	G	3.56E-04	1.23	0.48	1.05	2.68E-07 ²	Intergenic
10	rs224120	64115766	G	5.84E-05	1.30	0.35	0.87		Intergenic
10	rs224123	64121239	G	6.38E-04	1.22	0.01	1.17		Intergenic

* Defined by the largest transcript dbGene

¹ Barrett *et al.* Nat Genet 2008;40:955–62;

² WTCCC *et al.* Nature 2007;447:661–78

Genes with Differential Expression (F-ratio>30) in EBV-transformed Lymphoblastoid Cell Lines Due to Variation in *ZNF365* isoform D and Disease Status

Table 3

Probe ID	Gene	Mean Expression (fluorescence units)				Genotype		Disease		Interaction		Gene Description
		AA control	AA CD	GG control	GG CD	F-ratio	P-value	F-ratio	P-value	F-ratio	P-value	
ILMN_1704842	ARL4A	137	145	146	134	0.38	NS	1.54	NS	38.82	2.20E-05	ADP-ribosylation factor GTP-binding protein
ILMN_2274180	MKKS	242	259	282	222	0.02	NS	10.53	5.87E-03	31.40	6.50E-05	Probable molecular chaperone; assists in protein-folding upon ATP hydrolysis
ILMN_1699772	RRAGD	3362	3038	2910	1929	32.42	5.50E-05	23.60	2.50E-04	8.33	1.20E-02	Ras-related GTP binding protein
ILMN_1685371	SUMF2	293	234	318	326	32.87	5.20E-05	6.93	2.00E-02	11.66	4.20E-03	Sulfatase-modifying factor localizing to lumen of the endoplasmic reticulum
ILMN_2175094	TDRD1	296	253	238	194	49.65	5.80E-06	27.16	1.30E-04	0.62	NS	Tudor domain containing protein
ILMN_2143822	ZNF148	1212	1175	1037	856	53.63	3.80E-06	11.78	4.00E-03	6.05	2.80E-02	Zinc finger transcription factor; see discussion
ILMN_2152131	ACTB	26772	26646	27776	24534	0.02	NS	1.87	NS	1.60	NS	Beta-Actin
ILMN_1777296	ACTB	27896	28830	29250	29485	2.07	NS	0.75	NS	0.19	NS	
ILMN_2038777	ACTB	25695	26437	25128	26676	0.08	NS	0.85	NS	0.06	NS	

F-ratio>30 for genotype and interaction of genotype with disease status

NS= Not significant at p< 0.05; GG= Risk allele; AA= Protective allele