Extended haplotype association study in Crohn’s disease identifies a novel, Ashkenazi Jewish-specific missense mutation in the NF-κB pathway gene, HEATR3

W Zhang1,4, KY Hu1,4, A Gusev2, N Warner3, SME Ng1, J Ferguson1, M Choi4, A Burberry5, C Abraham1, L Mayer5, RJ Desnick5, CJ Cardinale6, H Hakonarson6, M Waterman2, Y Chowers7, A Karban2, SR Brant8, MS Silverberg9, PK Gregersen10, S Katz11, RP Lifton4, H Zhao4,12, G Nuñez3, I Pe'er13, I Peter7 and JH Cho1,4

The Ashkenazi Jewish population has a several-fold higher prevalence of Crohn’s disease (CD) compared with non-Jewish European ancestry populations and has a unique genetic history. Haplotype association is critical to CD etiology in this population, most notably at NOD2, in which three causal, uncommon and conditionally independent NOD2 variants reside on a shared background haplotype. We present an analysis of extended haplotypes that showed significantly greater association to CD in the Ashkenazi Jewish population compared with a non-Jewish population (145 haplotypes and no haplotypes with P-value <10\(^{-3}\), respectively). Two haplotype regions, one each on chromosomes 16 and 21, conferred increased disease risk within established CD loci. We performed exome sequencing of 55 Ashkenazi Jewish individuals and follow-up genotyping focused on variants in these two regions. We observed Ashkenazi Jewish-specific nominal association at R755C in NOD2 on chromosome 21. Within the chromosome 16 region, R642S of HEATR3 and rs9922362 of BRD7 showed genome-wide significance. Expression studies of HEATR3 demonstrated a positive role in NOD2-mediated NF-κB signaling. The BRD7 signal showed conditional dependence with only the downstream rare CD-causal variants in NOD2, but not with the background haplotype; this elaborates NOD2 as a key illustration of synthetic association.

Genes and Immunity (2013) 14, 310–316; doi:10.1038/gene.2013.19; published online 25 April 2013

Keywords: haplotype association; Ashkenazi Jewish; Crohn’s disease; NF-κB signaling; synthetic association

INTRODUCTION

Crohn’s disease (CD) is a complex genetic disorder characterized by chronic intestinal inflammation resulting from a dysregulated host immune response to intestinal microbiota. The NOD2 gene, involved in innate immune responses to bacterial peptidoglycan, is strongly associated with CD and was initially identified through genetic linkage. Uncommon, loss-of-function coding mutations (Arg702Trp, Gly908Arg and Leu1007fsinsC) in NOD2 confer a 17.1-fold (95% confidence interval: 10.7–27.2) increased risk for disease with homozygous or compound heterozygote risk allele carriage. The advent of genome-wide association studies (GWAS) has resulted in the identification of 140 genetic loci that account for only 13.6% of the estimated heritability. Because genotyping platforms utilized thus far have focused on common variants, it is possible that untested rare mutations may contribute significantly to complex disorders and account for some portion of missing heritability. Furthermore, precise models of disease pathogenesis integrating multiple disease associations are largely lacking, reflecting in part the significant pathophysiological heterogeneity underlying the myriad associations reported thus far in CD. Approaches to reduce genetic complexity, such as through focused studies in selected populations, may be of benefit.

The Ashkenazi Jewish population has a several-fold higher prevalence of CD compared with non-Jewish European ancestry cohorts, with estimates of increased prevalence ranging between 4.3- to 7.7-fold.5,6 The Ashkenazim have a unique genetic history, characterized by population bottlenecks, expansions and endogamy. However, the associated polymorphisms identified thus far through GWASs, as well as at NOD2, do not account for the higher disease prevalence in the Ashkenazim. Although the
basis for the increased frequency of CD in the Ashkenazim is not known, it is possible that unidentified, uncommon variants, unique to or at a higher frequency within Ashkenazi Jews, contribute to higher disease prevalence. An additional possibility, given the greater linkage disequilibrium observed within the Ashkenazim, is that multiple functional polymorphisms inherited in cis contribute to disease more commonly within Ashkenazi Jewish compared with non-Jewish European ancestry populations.

Such a configuration of mutations would be consistent with the well-documented presence of longer haplotype blocks with greater levels of linkage disequilibrium, relative to non-Jewish populations. As such, association testing to identify significant extended regions may point the way to understanding the unique genetic architecture underlying CD in Ashkenazi Jews.

Studying haplotype structure in the context of disease association also informs our conception of synthetic association, which has been hypothesized to be a major contributor to genome-wide association signals. Under this framework, molecular evolution, which creates varying amounts of haplotype diversity, allows for significant disease association near clusters of rare causal variants. For example, in CD, NOD2 is distinguished by three uncommon functional mutations whose carriers are a subset of individuals who have a common background haplotype. The multiple hierarchical levels in the tree-based genealogy, originally proposed by Dickson et al., are reflected in various stringencies for the clustering parameters in an analysis of contemporary haplotype diversity; for analyses with strict parameters, causal variants may not lie within the associated haplotypes themselves. Using haplotypes may be a particularly powerful approach for tagging functional variants; a recent study showed improved power over a single-variant approach. Toward this end, we present an analysis using extended shared haplotypes as a filter for prioritizing novel variants and detecting conditional dependence structure.

**RESULTS**

**Extended haplotype analysis**

To address the presence and relevance to disease of extended linkage disequilibrium stretches, we performed a haplotype-based analysis of a cohort comprising 397 Ashkenazi CD cases and 431 controls, and 547 non-Ashkenazi CD cases and 549 controls. Our analysis of associated extended haplotypes showed that there were markedly distinct patterns of haplotype association signals in the Ashkenazim compared with the non-Jewish cohort (Figure 1).

We observed 145 haplotypes demonstrating nominal evidence for association with $P$-values $< 10^{-3}$, including three distinct regions with $P$-values $< 10^{-4}$ (Supplementary Table 2). At this threshold, no regions were found to be significant in the non-Jewish cohort. Comparing the 145 haplotypes with 71 previously identified CD loci in a GWAS-based meta-analysis, we found that there were four haplotypes overlapping three established loci (Table 1); no haplotypes overlapped the 69 additional CD-associated loci in a larger immunochip-based study. Among all loci, the most significant haplotype association signal was found at chromosome 16q12, in a region that includes the NOD2 gene. Three of these four haplotypes showed higher allele frequencies in CD cases, corresponding to increased disease risk, whereas the remaining haplotype on chromosome 2 appeared to confer a protective function.

**Chromosome 16 haplotypes.** Two of the significant haplotypes in established CD loci (chr16_hap6721 and chr16_hap6830) were located contiguously on chromosome 16, spanning 48.88–49.18 and 49.18–49.48 Mb, respectively (Table 1). Haplotype chr16_hap6830 was carried by 25 CD Ashkenazi cases and 4 Ashkenazi controls, and chr16_hap6721 was carried by 15 CD cases and no controls. The 15 CD cases that carried chr16_hap6721 were completely a subset of the 25 cases that carried chr16_hap6721. Importantly, of these 15 CD-extended haplotype carriers, 12 also carried the Gly908Arg polymorphism within the NOD2 gene. Of the three major disease-associated polymorphisms, only the Gly908Arg variant is present at a significantly higher frequency in Ashkenazi compared with non-Ashkenazi CD cases.

In response to stimulation by a component of bacterial peptidoglycan, muramyl dipeptide (MDP), NOD2 signaling leads to the activation of the NF-$\kappa$B (nuclear factor of kappa light polypeptide gene enhancer in B cells) family of transcription factors. In addition to the established association signal at

---

**Figure 1.** Manhattan plot of haplotype associations in non-Ashkenazi Jewish (a) and Ashkenazi Jewish (b) European Ancestry cohorts. In the non-Jewish analysis, no regions demonstrated evidence for association. In contrast, in the Ashkenazim, we observed 145 haplotypes demonstrating nominal evidence for association with $P$-values $< 10^{-3}$, including three distinct regions with $P$-values $< 10^{-4}$. Associated haplotypes that overlapped with established CD loci, and which are detailed in Table 1, are labeled.
NOD2, the associated region on chromosome 16 also contained the BRD7 (bromodomain containing 7) gene, which has demonstrated genome-wide significant association at rs9922362 ($P$-value $= 3.26 \times 10^{-16}$) in a GWAS data set, separate from the one used in the haplotype analysis, comprising 907 Ashkenazi CD and 2345 Ashkenazi controls (Figure 2). BRD7 lies within 350 Kb of NOD2 and encodes a protein involved in chromatin remodeling. Notably, the GWAS data set did not show any evidence for independent association in CYLD (cylindromatosis), which is immediately adjacent to NOD2, as has previously been reported.

Chromosomes 2 and 21 haplotypes. Like the chromosome 16 region, the haplotype on chromosome 21, chr21_hap8015 (chr21: 44.47–44.65 Mb) also spans multiple immune-disease-related genes, such as ICOSLG (inducible T-cell co-stimulator ligand) and AIRE (autoimmune regulator). The chromosome 2 haplotype, chr2_hap10940 (chr2: 61.18–61.87 Mb), which had a higher frequency in controls compared with cases, was within 125 Kb of COMMD1 (copper metabolism (Murr1) domain containing 1), a regulator of the NF-$\kappa$B signaling pathway.

Exome sequencing
To further explore the haplotype regions of interest, we performed exome sequencing on 55 Ashkenazi Jewish samples, including 50 CD cases and 5 healthy controls. The samples' Jewish ancestry was validated genetically using GWAS data, and the case samples included two carriers, each of the risk haplotypes on chromosomes 16 and 21. Within the associated haplotype regions of chromosomes 16 and 21, we identified 40 previously unreported missense or nonsense polymorphisms among all 55 Ashkenazi Jewish individuals sequenced (Supplementary Table 3). None of the coding variants identified in the chromosome 2 haplotype were predicted by Polyphen-2 to be probably damaging in function, and no further investigation of these single-nucleotide polymorphisms (SNPs) was conducted.

Chromosome 21 haplotype. We identified a number of previously unreported missense or nonsense uncommon mutations within the TRPM2 (transient receptor potential cation channel, subfamily M, member 2) gene, including three polymorphisms, Arg755Cys, Gln953Stop and Thr1347Met, predicted by PolyPhen-2 to be probably damaging in function. Genotyping of these three SNPs was performed using Taqman assays on a follow-up Ashkenazi cohort of 1220 cases and 1167 healthy controls (Supplementary Table 1). Table 2 summarizes the association evidence for the three TRPM2 variants. We observed nominal evidence for association ($P$-value $= 0.0015$) for the most common of these variants, Arg755Cys in the Ashkenazi Jewish CD cohort. At this SNP, in a comparably sized non-Jewish cohort (915 non-Jewish European ancestry CD cases and 818 matched healthy controls), we observed no evidence for association ($P$-value $= 0.725$). In contrast to Arg755Cys, we observed no evidence for association in the Jewish cohort for either Gln953stop or Thr1347Met in TRPM2.

Of interest, the Gln953stop variant is carried by an individual carrying the chromosome 21 risk haplotype (Table 1) and is specific to the Ashkenazi Jewish population, not being observed in a subset of the non-Jewish European ancestry samples comprising 192 CD cases and 192 healthy controls. Additional studies with

Table 1. Associated haplotypes overlapping with established CD loci in Ashkenazi Jewish samples

| Haplotype ID     | Chr a | Start position a | End position a | P-value | MAF case (N = 397) | MAF control (N = 431) | Established CD loci (Mb) |
|------------------|-------|------------------|----------------|---------|--------------------|------------------------|--------------------------|
| chr2_hap10940    | 2     | 61 184 852       | 61 865 369     | $4.63 \times 10^{-04}$ | 0.0038             | 0.024                  | 60.76–61.87              |
| chr16_hap6721    | 16    | 48 885 095       | 49 184 879     | $5.04 \times 10^{-05}$ | 0.019              | 0                      | 49.02–49.41              |
| chr16_hap6830    | 16    | 49 184 879       | 49 479 621     | $3.17 \times 10^{-05}$ | 0.032              | 0.0046                 | 49.02–49.41              |
| chr21_hap8015    | 21    | 44 471 849       | 44 657 340     | $5.26 \times 10^{-04}$ | 0.014              | 0                      | 44.41–44.52              |

Abbreviations: CD, Crohn's disease; Chr, chromosome; ID, identification; MAF, minor allele frequency. aChromosomal coordinates in hg18.

Figure 2. Single-marker associations in the chromosome 16 haplotype region. We observed genome-wide significant associations at rs2076756 ($P$-value $= 2.32 \times 10^{-20}$) in NOD2 and rs9922362 ($P$-value $= 3.26 \times 10^{-16}$) in BRD7. Positions of two nominally associated Ashkenazi Jewish haplotypes are shown by black bars at top.
larger cohorts will be required to definitively determine a role for TRPM2 in Ashkenazi Jewish CD.

**Chromosome 16 haplotypes.** Within the region encompassing the chromosome 16 risk haplotypes (Table 1), we identified 21 missense or nonsense mutations among the 55 individuals undergoing exome sequencing (Supplementary Table 3). Although most of the carriers of the chromosome 16 risk haplotypes also carry the NOD2 Gly908Arg variant, the converse is not true; in the initial GWAS cohort, only 15 of the 68 Gly908Arg carriers (22%) also carried the extended chromosome 16 risk haplotypes. We then asked whether carriers of the extended risk haplotype might carry additional risk alleles. Importantly, two extended chromosome 16 risk haplotype carriers also carried a previously unidentified missense mutation, Arg642Ser, in HEATR3. The HEATR3 gene encodes a 680 amino-acid protein, and the Arg642Ser polymorphism is highly conserved between species and predicted to be probably damaging to gene function by PolyPhen-2.

Fine mapping of chromosome 16 haplotypes

To more precisely define the nature of the various association signals in the NOD2 haplotype, we performed single-point and logistic regression analyses on the region’s key markers in the 1167 Ashkenazi cases and 1220 Ashkenazi controls described above (Table 2). Consistent with prior reports, the Gly908Arg variant is the most common NOD2 risk allele (10.4% minor allele frequency in cases) in the Ashkenazi Jewish cohort; this is in contrast to non-Jewish European ancestry CD cohorts, where the frameshift mutation, Leu1007fsinsC, is the most common.19,20 The Arg642Ser variant in HEATR3 (3.2% minor allele frequency in cases) is highly associated with CD (P-value = 3.53 \times 10^{-7}). We genotyped the Arg642Ser variant using Sanger sequencing in 384 non-Jewish European ancestry CD cases and identified no carriers, indicating an Ashkenazi Jewish predominance for this novel variant. Because the HEATR3 Arg642Ser variant is present in Ashkenazi Jews on an extended risk haplotype also containing the more common NOD2 Gly908Arg allele (‘D’ between Arg642Ser in HEATR3 and Gly908Arg in NOD2 = 0.828), we next performed conditional logistic regression analysis to test for additional conditionally dependent relationships. Several models were compared using the Akaike Information Criterion to determine the best subset of SNPs. Additive, dominant and factor coding for the SNP genotypes were examined during the modeling stage. The final model with minimum Akaike Information Criterion contained the associated marker in BRD7, rs9922362, as well as all four SNPs in NOD2, but not Arg642Ser in HEATR3. Under this regression model, Gly908Arg in NOD2 demonstrated the most significant evidence for association (P-value = 1.66 \times 10^{-12}). In addition, in a full model containing all six variants across the three genes, a conditional relationship to NOD2 accounted for much of the association evidence observed at HEATR3 Arg642Ser variant (P-value = 0.333).

**HEATR3 knockdown and overexpression**

Although the genetic association signal at Arg642Ser in HEATR3 on chromosome 16 did not appear to be independent, we noted that this gene is highly expressed in innate immune cells,30 and we thus investigated its function in regulating NOD2 signaling. To test this, we transiently knocked down HEATR3 using small interfering RNA (siRNA) in cells stably expressing NOD2 and stimulated them with MDP in order to activate NOD2 signaling. We monitored NOD2-dependent signaling using both NF-κB activation, as measured by a luciferase reporter assay (Promega, Madison, WI, USA), and interleukin-8 (IL-8) secretion, as measured by enzyme-linked immunosorbent assay. Cells transfected with siRNA-targeting HEATR3 showed no significant loss in cell viability but resulted in reduced MDP-induced NF-κB activation and >50% reduction in IL-8 secretion compared with non-targeting siRNA controls. These results suggest that HEATR3 is a positive regulator of NOD2 signaling. To further establish this relationship, we also overexpressed HEATR3 in HEK293 cells stably expressing NOD2. Cells overexpressing HEATR3 exhibited increased luciferase activity and increased IL-8 secretion, thereby establishing a role for HEATR3 in positively regulating NOD2 signaling (Figure 3).

**DISCUSSION**

The chromosome 16 associations at BRD7 and NOD2 represent an empirical paradigm of synthetic association that illustrates and extends the proposals of Dickson et al.10 Using computer simulation, the researchers demonstrated that a cluster of low frequency, highly penetrant mutations often occurs stochastically more frequently with one allele than the other at a common SNP, resulting in ‘synthetic’ association signals at these common variants. Although it has been argued that synthetic associations are not common in general, the interactions between and within BRD7 and NOD2 provide an important illustration of the complexities inherent in regression analyses to identify independent alleles. The minor allele (minor allele frequency 23.5% in healthy controls) at the common NOD2 polymorphism rs2076756 tags a shared haplotype on which four independent,

### Table 2. SNP association and logistic regression analysis in Ashkenazi Jewish samples

| SNP          | Positiona | Gene       | Alleles | MAF case (N = 1167) | MAF control (N = 1220) | P-value   | P-value with model selectionb | P-value, full modelc |
|--------------|-----------|------------|---------|---------------------|------------------------|-----------|-------------------------------|---------------------|
| Chr16 (16q12.1)—HEATR3, BRD7 and NOD2 | 16425     | HEATR3    | C/A     | 0.032               | 0.100                  | 3.53 × 10^{-7} | 0.33                           |                     |
|                | R9922362  | BRD7      | G/A     | 0.39                | 0.32                   | 2.36 × 10^{-7} | 0.092                           | 0.041               |
|                | R7032     | NOD2      | C/T     | 0.040               | 0.022                  | 5.00 × 10^{-4} | 0.55 × 10^{-5}                 | 5.27 × 10^{-5}      |
|                | N8525     | NOD2      | A/G     | 0.024               | 0.012                  | 0.00140         | 0.0027                          | 0.0028              |
|                | G908R     | NOD2      | G/C     | 0.10                 | 0.040                  | 1.93 × 10^{-17} | 1.66 × 10^{-13}                 | 3.57 × 10^{-10}     |
|                | L1007fsinsC | NOD2     | G/C     | 0.087               | 0.028                  | 1.65 × 10^{-18} | 3.10 × 10^{-8}                 | 3.90 × 10^{-8}      |
| Chr21 (21q22.3)—TRPM2 | 7595     | TRPM2     | C/T     | 0.043               | 0.026                  | 0.0015           |                                |                     |
|                | Q953X     | TRPM2     | C/T     | 0.0080              | 0.0080                 | 0.76              |                                |                     |
|                | T1347M    | TRPM2     | C/T     | 0.030               | 0.025                  | 0.40              |                                |                     |

**Abbreviations:** MAF, minor allele frequency; SNP, single-nucleotide polymorphism. *Chromosomal coordinates in hg18. bLogistic regression model with minimum Akaike Information Criterion, comprising all four variants on chromosome 16. cFull logistic regression model comprising all six variants on chromosome 16.
causal and uncommon NOD2 variants (Arg702Trp, Asn852Ser, Gly908Arg and Leu1007fsinsC) completely reside. This has been widely acknowledged as an example of synthetic association, as the NOD2 GWAS signal is created by this cluster of rare variants with high effect, and the signal is mappable using linkage, which was originally predicted by Dickson et al.\(^\text{10,31}\)

In our analysis, common variants at \(\text{BRD7}\) (rs9922362) and NOD2 (rs2076756) appeared to confer independent evidence for association, with \(P\)-values \(\approx 1.42 \times 10^{-13}\) and \(3.25 \times 10^{-13}\), respectively, in a logistic regression model consisting of the two SNPs. After conditioning the \(\text{BRD7}\) SNP on four uncommon causal NOD2 variants, however, the \(\text{BRD7}\) association signal largely diminishes (conditional \(P\)-value \(\approx 0.0292\)). Taken together, these statistical results imply that the \(\text{BRD7}\) SNP functions as a partial background for the directly causal NOD2 mutations, but not for the entire NOD2 CD-associated haplotype. This is visualized more directly in Figure 4, which shows that the \(\text{BRD7}\) SNP indeed does not appear to simply be associated through linkage disequilibrium with the NOD2 background SNP, but rather that its carriers are enriched for the rare causal variants within the NOD2 background haplotype.

In our analysis, common variants at \(\text{BRD7}\) (rs9922362) and NOD2 (rs2076756) appeared to confer independent evidence for association, with \(P\)-values \(\approx 1.42 \times 10^{-13}\) and \(3.25 \times 10^{-13}\), respectively, in a logistic regression model consisting of the two SNPs. After conditioning the \(\text{BRD7}\) SNP on four uncommon causal NOD2 variants, however, the \(\text{BRD7}\) association signal largely diminishes (conditional \(P\)-value \(\approx 0.0292\)). Taken together, these statistical results imply that the \(\text{BRD7}\) SNP functions as a partial background for the directly causal NOD2 mutations, but not for the entire NOD2 CD-associated haplotype. This is visualized more directly in Figure 4, which shows that the \(\text{BRD7}\) SNP indeed does not appear to simply be associated through linkage disequilibrium with the NOD2 background SNP, but rather that its carriers are enriched for the rare causal variants within the NOD2 background haplotype. Importantly, the 0.4-Mb distance between the \(\text{BRD7}\) variant and NOD2 is consistent with the findings by Dickson et al.,\(^\text{10}\) which showed that causal relationships can extend over long regions surrounding synthetic association, in the range of several megabases. However, we note that the absence of conditional dependence between the two common background SNPs in \(\text{BRD7}\) and NOD2 conflicts with the hierarchical genealogy model supported by Dickson et al. Overall, the example of CD illustrates the variability of conditioning on highly associated, non-causal background variants such as the NOD2 common haplotype background variant, rs2076756, as opposed to functional uncommon variants.

In this study, we report associations of uncommon extended haplotypes in Ashkenazi Jewish, but not non-Jewish European ancestry populations, thereby illustrating the greater power of this approach in endogamous populations. The haplotype associations on chromosome 16 tag the Gly908Arg variant within NOD2, the most prevalent causal mutation in Ashkenazi Jewish populations. The extended haplotype carriers in this region tag a less common, Ashkenazi Jewish-specific variant, Arg642Ser in HEATR3, which we demonstrated to function in the NOD2 pathway. Given the extensive linkage disequilibrium between Gly908Arg in NOD2 and Arg642Ser in HEATR3, however, we did not observe independent

\[ \text{HEATR3} \] is a positive component of the NOD2 signaling pathway. siRNA-mediated knockdown of \(\text{HEATR3}\) diminished NOD2-dependent NF-\(\kappa\)B induction (a) and IL-8 secretion (b) relative to non-targeting siRNA controls. Furthermore, overexpression of \(\text{HEATR3}\) increased NF-\(\kappa\)B signaling (c) and IL-8 production (d). Cells were left unstimulated (black bars), or stimulated overnight with 20 ng ml\(^{-1}\) of MDP (gray bars). Together, these results support a positive role for \(\text{HEATR3}\) in NOD2 signaling. \(* P < 0.05, ** P < 0.01\) and \(*** P < 0.001\).
evidence for association for the HEATR3 SNP. This marker, along with others in TRPM2 on chromosome 21, demonstrated the capability of our haplotype-based analysis to direct the discovery of new disease-associated and population-specific polymorphisms. Furthermore, the extended length of the Jewish-specific haplotypes may indicate the presence of long-distance conditional relationships, as in the case of rs9922362 of BRD7 and the NOD2 causal variants on chromosome 16, an example of synthetic association. As additional fine-mapping studies of GWAS loci are reported, an increasing number of population-specific polymorphisms and haplotypes will be identified. The etiology for the higher prevalence of CD in the Ashkenazi Jewish population remains largely undefined; however, our study indicates that extended haplotypes containing multiple functional polymorphisms inherited in cis may contribute in a population-specific manner.

**MATERIALS AND METHODS**

**Subjects**

Three hundred and ninety-seven Ashkenazi CD cases and 431 controls, and 547 non-Ashkenazi CD cases and 549 controls (Supplementary Table 1), were ascertained through Genetics Research Centers in Baltimore, Chicago, Montreal, Pittsburgh, Los Angeles and Toronto, and through the New York Health Project. In all cases, informed consent was obtained following protocols approved by each local institutional review board. Diagnostic inclusion and exclusion criteria are described elsewhere.34 Ashkenazi Jewish ancestry was validated using principal components analysis.

**Genotyping and haplotype association analysis**

Genomic DNA was isolated from whole blood from the case–control cohort. Genotyping was performed on the Illumina (San Diego, CA, USA) HumanHap300 platform, with an average inter-marker distance <10 Kb. All samples had genotype yields >94%, and missing genotypes were imputed using BEAGLE 3.1.0 (http://faculty.washington.edu/browning/beagle/beagle.html) with default parameters.32 Analysis was performed using GERMLINE 1.5.0 (Genetic Error-tolerant Regional Matching with LINear-time Extension), a computationally efficient program for identifying shared identical-by-descent segments between pairs of individuals in a large population, and DASH 1.1.0 (DASH Associates Shared Haplotypes), a program for clustering these segments into haplotype regions that are shared by multiple individuals.33,34 Our analysis utilized default parameters, with a minimum haplotype length of 3 Mb and a maximum of four homoygous mismatches between shared segments. In the Ashkenazi samples, 473621 haplotype clusters across the autosomes were tested for association, compared with 8986 clusters in the non-Jewish samples. Because of correlation between the haplotype clusters, precise multiple testing corrections could not be calculated, but given that all clusters in the Jewish and non-Jewish analyses had P-values >10−5 and 10−3, respectively, it was clear that none were genome-wide significant.

**Exome sequencing and variant screening**

Genomic DNA was extracted from whole blood and whole blood captured with the NimbleGen 2.1M human exome array following the manufacturer’s protocol (Roche/NimbleGen, Madison, WI, USA). Captured libraries were sequenced on the Illumina genome analyzer as paired-end 75-bp reads, following the manufacturer’s protocol. Sequence reads were mapped to the reference genome (hg18) using the BWA program (http://bio-bwa.sourceforge.net) using default parameters.35 The Genome Analysis Toolkit v2 (http://www.broadinstitute.org/gatk/) was used to call alleles at variant sites.36 Sample-level realignment and multi-sample SNP calling were performed using default parameters and the Genome Analysis Toolkit standard variant filter.

**HEATR3 knockdown and overexpression**

HEK293 cells stably expressing NOD2 were cultured in Dulbecco’s modified Eagle’s medium with 10% fetal bovine serum, 2 mM glutamine, 1 mM sodium pyruvate and 1 x penicillin-streptomycin (Gibco, Grand Island, NY, USA). All experiments were performed in duplicate. In the silencing experiments, these cells were then reverse-transfected with siGENOME siRNA pools (Thermo Scientific, Hudson, NH, USA) targeting HEATR3 (or non-targeting siRNA as a negative control) at a concentration of 20 nM with the lipid transfection agent Lipofectamine 2000 (Invitrogen, Grand Island, NY, USA) diluted in Opti-MEM (Invitrogen) for 48 h. Cells were either left unstimulated or subsequently stimulated with MDP (20 ng/ml−1, Bachem, Torrance, CA, USA) for 18 h. In the overexpression experiment, HEK293 cells were transiently transfected using Lipofectamine 2000 with a pcMV-SPORT6-based expression plasmid encoding murine HEATR3 (Thermo Scientific), followed by MDP stimulation as described above. Transfection with the empty vector was used as a negative control.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**ACKNOWLEDGEMENTS**

This work was supported by the National Institutes of Health (NIDDK U01 DK062429, U01 DK062422, R01 DK092235 and RC1 DK086800 to JHC, NIDDK F30 DK98927 and NIGMS T32 GM007205 to KYH and R01 DK77905 to CA), the New York Crohn’s Disease Foundation (LM, RJD and IP) and the Crohn’s and Colitis Foundation of America (K12RR0024138 to JHC).

**REFERENCES**

1. Abraham C, Cho JH. Inflammatory bowel disease. N Engl J Med 2009; 361: 2066–2078.
2. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn’s disease. Nature 2001; 411: 599–603.
3. 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: 2393–2404.
4. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012; 491: 119–124.
5. Mayberry JF, Judd D, Smart H, Rhodes J, Calcott B, Morris JS. Crohn’s disease in Jewish people—an epidemiological study in south-east Wales. Digestion 1986; 35: 237–240.
6. Bernstein CN, Rawsthorne P, Cheung M, Blanchard JF. A population-based case control study of potential risk factors for IBD. Am J Gastroenterol 2006; 101: 993–1002.
7. Ostrer H. A genetic profile of contemporary Jewish populations. Nat Rev Genet 2001; 2: 891–898.
8. Atzmon G, Hao L, Pe’er I, Velez C, Pearlman A, Palamara PF et al. Abraham’s children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. Am J Hum Genet 2010; 86: 850–859.
9. Bray SM, Mulle JG, Dodd AF, Pulver AE, Wooding S, Warren ST. Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. Proc Natl Acad Sci USA 2010; 107: 16222–16227.
10. Dickson SP, Wang K, Krantz I, Jakobsson M, Goldstein DB. Rare variants create synthetic genome-wide associations. PLoS Biol 2010; 8: e1000294.
11. Ogura Y, Bonen DK, Inohara N, Nikolai DL, Chen FF, Ramos R et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn’s disease. Nature 2001; 411: 603–606.
12. Kang K, Kugathasan S, Georges M, Zhao H, Cho JH. Improved risk prediction for Crohn’s disease with a multi-focus approach. Hum Mol Genet 2011; 20: 2433–2442.
13. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ et al. Genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006; 314: 1461–1463.
14. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A 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.
15. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn’s disease susceptibility loci. Nat Genet 2010; 42: 1118–1125.
16. Karban A, Atia O, Leitersdorf E, Shahbari A, Sbeit W, Ackerman Z et al. The relation between NOD2/CARD15 mutations and the prevalence and phenotypic heterogeneity of Crohn’s disease: lessons from the Israeli Arab Crohn’s disease cohort. Dis Colon Rectum 2005; 50: 1692–1697.
17. Peter I, Mitchell AA, Oezelius L, Erazo M, Hu J, Doherty D et al. Evaluation of 22 genetic variants with Crohn’s disease risk in the Ashkenazi Jewish population: a case-control study. BMC Med Genet 2011; 12: 63.
18. Newman B, Silverberg MS, Gu X, Zhang Q, Lazaro A, Steinhart AH et al. CARD15 and HLA DRB1 alleles influence susceptibility and disease localization in Crohn’s disease. Am J Gastroenterol 2004; 99: 306–315.

© 2013 Macmillan Publishers Limited Genes and Immunity (2013) 310 – 316 Extended haplotype association study in Crohn’s disease

W Zhang et al
Supplementary Information accompanies this paper on Genes and Immunity website (http://www.nature.com/gene)