Published in final edited form as:
Clin Gastroenterol Hepatol. 2007 August ; 5(8): 972–976. doi:10.1016/j.cgh.2007.04.024.

Association of variants of the interleukin-23 receptor (*IL23R*) gene with susceptibility to pediatric Crohn's disease

Robert N. Baldassano[1], **Jonathan P. Bradfield**[2], **Dimitri S. Monos**[3], **Cecilia E. Kim**[2], **Joseph T. Glessner**[2], **Tracy Casalunovo**[2], **Edward C. Frackelton**[2], **F. George Otieno**[2], **Stathis Kanterakis**[3], **Julie L. Shaner**[2], **Ryan M. Smith**[2], **Andrew W. Eckert**[2], **Luke J. Robinson**[2], **Chioma C. Onyiah**[2], **Debra J. Abrams**[1], **Rosetta M. Chiavacci**[2], **Robert Skraban**[2], **Marcella Devoto**[4,5], **Struan F.A. Grant**[2,4,*], and **Hakon Hakonarson**[2,4,*]

[1] Division of Gastroenterology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104

[2] Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104

[3] Department of Pediatrics University of Pennsylvania, School of Medicine and Department of Pathology and Laboratory Medicine, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104

[4] Division of Human Genetics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA

[5] CCEB, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Abstract

Background & Aims—Recently an association was demonstrated between the single nucleotide polymorphism (SNP), rs11209026, within the interleukin-23 receptor (*IL23R*) locus and Crohn's disease (CD) as a consequence of a genome wide association study of this disease in adults. We examined the effects of this and other previously reported SNPs at this locus with respect to CD in children.

Methods—Utilizing data from our ongoing genome-wide association study in our cohort of 142 pediatric CD cases and 281 matched controls, we investigated the association of the previously reported SNPs at the *IL23R* locus with the childhood form of this disease.

Results—Using a Fisher's exact test, the minor allele frequency (MAF) of rs1120902 in the cases was 1.75% while it was 6.61% in controls, yielding a protective odds ratio (OR) of 0.25 (95% CI 0.10 – 0.65; one-sided $P = 9.2\times10^{-4}$). Furthermore, of all the SNPs previously reported, rs11209026 was the most strongly associated. A subsequent family-based association test (which is more resistant to population stratification) with 65 sets of trios derived from our initial patient

*To whom correspondence should be addressed. E-mail: grants@chop.edu or hakonarson@chop.edu.

cohort yielded significant association with rs11209026 in a transmission disequilibrium test (one-sided P=0.0017). In contrast, no association was detected to the *CARD15* gene for the IBD phenotype.

Conclusions—The OR of the *IL23R* variant in our pediatric study is highly comparable with that reported previously in a non-Jewish adult IBD case-control cohort (OR=0.26). As such, variants in *IL23R* gene confer a similar magnitude of risk of CD to children as for their adult counterparts.

Keywords

IL23R; gene; association; Crohn's Disease

The inflammatory bowel diseases (IBDs), Crohn's disease (CD) and ulcerative colitis (UC) are complex genetic disorders of the gastrointestinal tract that are recurring and chronic in nature. The exact etiology of the disease is unknown, but complex interplay between environmental risk factors and immunological components have been implicated in the triggering of the disease in those who are genetically susceptible to the disease. The hunt for genetic variants conferring risk to IBD has been ongoing for more than 10 years, with a number of susceptibility genes being implicated through linkage and candidate gene studies1–4, the most prominent and most consistently replicated being the caspase-recruitment domain 15 (*CARD15*) gene (formerly *NOD2*) 1, 2, 5 which encodes a protein involved in bacterial recognition by cells involved in innate immunity. However, *CARD15* does not explain all cases of IBD, suggesting that other genetic factors may be involved.

Recently, a number of studies have reported remarkably strong, replicable associations with complex disease, including the complement factor H (*CFH*) gene in acute macular degeneration6–8 and the transcription factor 7-like-2 (*TCF7L2*) gene in type 2 diabetes9–11, with other such signals waiting for replication, such as insulin-induced gene 2 (*INSIG2*) in obesity12. In October this year, through the utilization of the Illumina HumanHap300 Genotyping BeadChip, Duerr *et al* 13 added to this repertoire by reporting a highly significant association between CD and the *IL23R* gene on chromosome 1p31. Specifically, an uncommon coding variant, rs11209026 (c.1142G>A, p.Arg381Gln), was shown to confer a strong protective effect against the disease and was then replicated in the same study in separate cohorts of patients with CD or ulcerative colitis.

Although these *IL23R* findings are compelling, there are continuing concerns regarding the performance of association studies in complex traits; as such independent replication efforts are now considered mandatory14. With the many errors and biases that can blight any individual study, replication by others can ensure that the original findings are robust and can also provide a more accurate estimate of the likely effect size15, 16.

In this study we demonstrate that the variant, rs11209026, in the *IL23R* gene confers similar magnitude of protection of CD in children as reported in adults with IBD, further suggesting that the IL23 pathway may be causally linked to CD in children.

Materials and Methods

Study Subjects

The case-control study consisted of 142 pediatric patients with CD and 281 matched controls. All subjects were biologically unrelated ethnic Caucasian. Patients were consecutively recruited from the Greater Philadelphia area from 2000 to 2006 at the Children's Hospital of Philadelphia. 65 cases had both parents also recruited for purposes of family based association analysis.

The diagnosis of CD was based on standard criteria17: 1) symptoms of abdominal pain, diarrhea, rectal bleeding, perirectal disease or growth failure; 2) two or more episodes of symptoms separated > 8 weeks apart or ongoing symptoms of >6 weeks duration, and; 3) signs of inflammation from endoscopic, histologic or radiologic evaluations. Involvement of the following intestinal regions was noted in each patient: esophagus, stomach, small bowel, terminal ileum, colon and their combinations based on objective evidence of mucosal ulcerations, cobblestoning, intestinal strictures or bowel wall thickening. All subjects had ileal or ileal-colonic disease and over 90 % of patients had involvement of the terminal ileum. Children with ulcerative colitis were excluded.

This study was approved by the Institutional Review Board of the Children's Hospital of Philadelphia.

Genotyping

We performed high throughput genome-wide SNP genotyping using the Illumina Infinium™ II HumanHap500 BeadChip technology18, 19. The Infinium™ II Assay protocol enables effective multiplexing and genome-wide SNP access through a single-base extension (SBE) method with enzymatic SNP scoring. The minimal hands-on three-day Infinium™ II protocol began with 750ng of input DNA. In day one, genomic DNA was amplified 1000–1500-fold. Day two, amplified DNA was fragmented ~300–600bp, then precipitated and resuspended followed by hybridization onto a BeadChip. SBE utilizes a single probe sequence ~50bp long designed to hybridize immediately adjacent to the SNP query site. Following targeted hybridization to the bead array, the arrayed SNP locus-specific primers (attached to beads) were extended with a single hapten-labeled dideoxynucleotide in the SBE reaction. The haptens were subsequently detected by a multi-layer immunohistochemical sandwich assay, as recently described18, 19. Our technicians followed the standard operating procedures issued by Illumina from the step that DNA is added to the single-tube whole genome amplification up to the processing of the BeadChip genotyping files. The Illumina BeadArray Reader scanned each BeadChip at two wavelengths and created an image file. As BeadChip images were collected, intensity values were determined for all instances of each bead type, and data files were created that summarized intensity values for each bead type. These files consisted of intensity data that was loaded directly into Illumina's genotype analysis software, BeadStudio. A bead pool manifest created from the LIMS database containing all the BeadChip data was loaded into BeadStudio along with the intensity data for the samples. BeadStudio used a normalization algorithm to minimize BeadChip to BeadChip variability. Once the normalization was

complete, the clustering algorithm was run to evaluate cluster positions for each locus and assign individual genotypes. Each locus was given an overall score based on the quality of the clustering and each individual genotype call was given a GenCall score. GenCall scores provided a quality metric that ranges from 0 to 1 assigned to every genotype called. GenCall scores were then calculated using information from the clustering of the samples. The location of each genotype relative to its assigned cluster determined its GenCall score.

The resources available for this project included the Illumina technology platform itself plus nine Tecan pipetting robotic systems, eight scanners, a laboratory information management system (LIMS) and automated allele-calling software. The workflow was robotic-based for automatic sample processing and included algorithms for quality control of genotypes. The facility infrastructure had sufficient computational power and servers for data processing and storing, including a series of computers that were integrated (warehouse setting) to perform continuous datamining of all gathered and generated datasets.

Analysis

Both the genetic matching of the cases and controls and the statistical tests for association were carried out using the software package *plink* (http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml). Controls were genetically matched to the Caucasian CD cases by clustering of the pair-wise identity-by-state distances. Complete linkage agglomerative clustering was used with restrictions based on a pair-wise population concordance test so that one case was matched with two controls. Controls were matched with cases based on their pair-wise identity-by-state as long as a pair-wise population concordance test had a p-value > 0.001. In some limited cases, only one genetically matched control could be found based on the restrictions. A one-sided Fisher's exact test was used to calculate p-values on allele count differences between 142 cases and 281 controls. Odds ratios and the corresponding 95% confidence intervals were calculated for the association analysis. A binomial exact transmission disequilibrium test was used to calculate one-sided p-values on differences between transmitted and untransmitted allele counts in 65 trios derived from the study cohort.

Results

Association between rs11209026 and pediatric CD risk

In this replication attempt, we genotyped 142 Caucasian pediatric CD cases (71 male, 71 female) and 281 matched controls (140 male, 141 female) recruited from the greater Philadelphia area with the Illumina HumanHap550 Genotyping BeadChip as part of our ongoing genome-wide association study of the disease. The diagnosis of CD was based on standard criteria (see Study Subjects section).

The probe for rs11209026 is present on our 550K BeadChip, so as a first pass we queried the data with a single test for this SNP to investigate if this marker is associated with CD in our pediatric cohort as reported by Duerr et al13. Using a Fisher's exact test, we observed a significant protective effect of the minor A allele on the risk of CD. The minor allele frequency (MAF) in the cases was 1.75% while it was 6.61% in controls, yielding an odds

ratio (OR) of 0.25 (95% CI 0.10 – 0.65; one-sided P = 9.2×10^{-4}) – see Table 1. This OR is very much in line with that reported previously in the non-Jewish adult case-control cohort (OR=0.26) and is also clearly lower than that reported in the adult Jewish case-control cohort (OR=0.45)13.

Duerr *et al* 13 also summarized all other SNPs in the locus that yielded a P<0.0001. As we had definitively replicated the association with rs11209026, we subsequently addressed these same SNPs to investigate their association with CD. The results in Table 1 indicate that the results are comparable to the original publication and that rs11209026 is indeed the most strongly associated SNP among those previously reported. In line with the original study13, many of these SNPs represent independent association signals, with six of the seven additional nominally significant SNPs having an r^2 <0.2 with rs11209026 (based on HapMap data20), thus they cannot be fully explained by the most significant association with rs11209026 (Table 1).

Family-based association between individual IL23R polymorphisms and pediatric CD risk

We then looked at the data using a family-based association test; this test is more resistant to population stratification and presents us with the opportunity to confirm that the observations from our case-control analysis are accurate. Utilizing 65 sets of trios, derived from our initial patient cohort, we also observed significant association with rs11209026 in a transmission disequilibrium test (one-sided P=0.0017). In addition, investigating the other SNPs outlined in Table 1 with this family-based approach again revealed no other association stronger than that of this key SNP – see Table 2.

Assessment of association to the CARD15 gene

Duerr *et al* 13 pointed out that the *CARD15* variants on the BeadChip, rs2066843 and rs2076756, in fact produced the best association in their analysis. In our cohort, rs2066843 had a MAF of 30.4% in cases and 28.2% in controls OR=1.11, [95% CI (0.82 – 1.52; one-sided P = 0.27)] and rs2076756 had a MAF of 30.3% in cases and 25.9% in controls [OR=1.24 (95% CI 0.91 – 1.70; one-sided P = 0.10)].

Discussion

From an interim analysis of our ongoing genome-wide association study of pediatric CD, we have investigated variation in the *IL23R* locus previously reported to be associated with adult IBD13. Consequently, we have replicated association of this gene with CD by demonstrating its effect in the childhood form of the disorder. More specifically, the uncommon coding variant, rs11209026 (c.1142G>A, p.Arg381Gln), was shown to confer a strong protective effect against the disease with a highly comparable odds ratio (0.25) to that previously observed in an adult non-Jewish cohort (0.26). In addition, the association seen within the family-based association analysis discounts the possibility that the original or our findings had arisen as a result of population substructure.

Although the size of the cohort in the original genome wide association study was larger, consisting of 567 non-Jewish, European ancestry patients and 571 non-Jewish controls13, the amount of testing in our cohort was very restricted to a focused effort of specifically

investigating if SNPs at a single locus also yield association in the same direction as previously reported. Our cohort is sufficiently powered to ask this straight-forward validation question, as the original finding impressively suggested that those individuals carrying the A allele of this SNP were approximately four-fold more protected from IBD that those carrying the G allele.

As the association we observe is indeed of a very similar magnitude to that of the original report, this independent replication confirms *IL23R* as a genuine IBD susceptibility gene. As such, the "winner's curse" that is often seen for other complex trait susceptibility genes21 is not observed in this instance. What is of particular note is that the association observed in children is almost identical to that of adults. With the gene-environment interaction22 models in mind, we have been motivated to look at the genetics of childhood disease in order to more readily distill the genetic component in these phenotypes due to the fact that environmental exposure and impact has been for a relatively short period of their lifetime. However, with the magnitude of the association being so comparable between children and adults, this particular research outcome suggests that the environmental interaction with this variant over time is negligible and in fact this variant may be primarily associated with early onset IBD. This may be further supported by the fact that the *CARD15* variants failed to show association in our pediatric cohort (and also debated by others23) as opposed to significant association in the original genome-wide study in adults13. While the lack of association with the *CARD15* variants may be related to power issues, it is clear that the *IL23R* association is readily detected in our pediatric cohort.

Our results lend further support for the protective role of the *IL23R* gene in IBD, suggesting that interventions at the IL23 pathway level may be of value in patients who suffer from this disease. The *IL23R* gene encodes a subunit of the receptor for the pro-inflammatory cytokine, interleukin-23. There are at least six alternatively spliced forms of the mRNA that in turn lead to diverse isoforms of the receptor protein24. The variants that we observe association to may directly dictate splicing or some other regulatory mechanism but more likely are in linkage disequilibrium with the causative variant(s).

Once our genome-wide association study is complete, we will have the opportunity to look for other variants in the genome that are associated with IBD, both CD and ulcerative colitis, as a consequence of our use of a higher resolution BeadChip. In addition, we will explore the *IL23R* gene further to elucidate other potential variants that may confer genetic susceptibility to this debilitating disorder in our cohort.

Acknowledgments

We would like to thank Alejandrina Estevez, Kenya Fain, Pamela Kline, Erin Santa, Alexandria Thomas and LaShea Williams for their expert assistance with the data collection and management. We would also like to thank Smari Kristinsson, Larus Arni Hermannsson and Asbjörn Krisbjörnsson of Raförninn ehf for their extensive software design and contribution. This research was financially supported by the Children's Hospital of Philadelphia, the Edmunds Fund, the Heineman Foundation, the IBD Family Research Council and NIH grant for the General Clinical Research Center (GCRC) 5-MO1-RR-000240.

Abbreviations used in this paper

CD	Crohn's Disease
IBD	Inflammatory Bowel Disease
SNP	single nucleotide polymorphism
IL23R	interleukin-23 receptor
OR	odds ratio
CI	confidence interval

References

1. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature. 2001; 411(6837):599–603. [PubMed: 11385576]
2. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature. 2001; 411(6837):603–6. [PubMed: 11385577]
3. Peltekova VD, Wintle RF, Rubin LA, et al. Functional variants of OCTN cation transporter genes are associated with Crohn disease. Nature Genetics. 2004; 36(5):471–5. [PubMed: 15107849]
4. Stoll M, Corneliussen B, Costello CM, et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nature genetics. 2004; 36(5):476–80. [PubMed: 15107852]
5. Radford-Smith G, Pandeya N. Associations between NOD2/CARD15 genotype and phenotype in Crohn's disease-Are we there yet? World J Gastroenterol. 2006; 12(44):7097–103. [PubMed: 17131470]
6. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005; 308(5720):385–9. [PubMed: 15761122]
7. Edwards AO, Ritter R 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. Science. 2005; 308(5720):421–4. [PubMed: 15761121]
8. Haines JL, Hauser MA, Schmidt S, et al. Complement factor H variant increases the risk of age-related macular degeneration. Science. 2005; 308(5720):419–21. [PubMed: 15761120]
9. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nature Genetics. 2006; 38(3):320–3. [PubMed: 16415884]
10. Florez JC, Jablonski KA, Bayley N, et al. TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. The New England Journal of Medicine. 2006; 355(3):241–50. [PubMed: 16855264]
11. Zeggini E, McCarthy MI. TCF7L2: the biggest story in diabetes genetics since HLA? Diabetologia. 2006
12. Herbert A, Gerry NP, McQueen MB, et al. A common genetic variant is associated with adult and childhood obesity. Science. 2006; 312(5771):279–83. [PubMed: 16614226]
13. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science. 2006; 314(5804):1461–3. [PubMed: 17068223]
14. Patterson M, Cardon L. Replication publication. PLoS Biology. 2005; 3(9):e327. [PubMed: 16149852]
15. Page GP, George V, Go RC, Page PZ, Allison DB. Are we there yet? Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. American Journal of Human Genetics. 2003; 73(4):711–9. [PubMed: 13680525]
16. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genet Med. 2002; 4(2):45–61. [PubMed: 11882781]
17. Walker-Smith, JA.; Sanderson, IR. Diagnostic Criteria for Chronic Inflammatory Bowel Disease in Childhood. In: Bistrian, BR.; Walker-Smith, JA., editors. Inflammatory Bowel Diseases. Vol. 2.

Nestlé Nutrition Workshop Series Clinical & Performance Programme, Nestec Ltd; Vevey/S Karger AG, Basel: 1999. p. 107-20.

18. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide scalable SNP genotyping assay using microarray technology. Nature genetics. 2005; 37(5):549–54. [PubMed: 15838508]
19. Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL. Whole-genome genotyping with the single-base extension assay. Nat Methods. 2006; 3(1):31–3. [PubMed: 16369550]
20. A haplotype map of the human genome. Nature. 2005; 437(7063):1299–320. [PubMed: 16255080]
21. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature genetics. 2003; 33(2):177–82. [PubMed: 12524541]
22. Hemminki K, Lorenzo Bermejo J, Forsti A. The balance between heritable and environmental aetiology of human disease. Nature Reviews. 2006; 7(12):958–65.
23. Ferraris A, Knafelz D, Torres B, Fortina P, Castro M, Dallapiccola B. Analysis of CARD15 gene variants in Italian pediatric patients with inflammatory bowel diseases. The Journal of Pediatrics. 2005; 147(2):272–3. [PubMed: 16126067]
24. Zhang XY, Zhang HJ, Zhang Y, et al. Identification and expression analysis of alternatively spliced isoforms of human interleukin-23 receptor gene in normal lymphoid cells and selected tumor cells. Immunogenetics. 2006; 57(12):934–43. [PubMed: 16372191]

Table 1

Crohn's disease (CD) case-control association study results for *IL23R* region markers with P-values < 0.0001 in Duerr *et al*13. Minor allele frequencies (MAF), allelic test one-sided P-values, and odds ratios (OR) with 95% confidence intervals (CI) are shown for each SNP. The ORs shown are for the minor allele. r^2 values, derived from HapMap20, are shown with respect to each SNP's linkage disequilibrium with rs11209026.

Marker	Location	MAF (Cases n= 142)	MAF (Controls n= 281)	One-sided P-value	OR [95% CI]	r^2 to rs11209026 (HapMap)
rs1004819	intron	0.385	0.296	0.0060	1.49 [1.10,2.00]	0.01
rs7517847	intron	0.339	0.441	0.0026	0.65 [0.48,0.87]	0.013
rs10489629	intron	0.399	0.465	0.039	0.76 [0.57,1.02]	0.067
rs2201841	intron	0.378	0.326	0.077	1.26 [0.93,1.69]	0.027
rs11465804	intron	0.021	0.067	0.0024	0.30 [0.13,0.72]	0.881
rs11209026	Arg381Gln	0.017	0.066	9.2×10^{-4}	0.25 [0.09,0.68]	1
rs1343151	intron	0.268	0.353	0.0073	0.67 [0.49,0.92]	0.111
rs10889677	exon-3′UTR	0.378	0.326	0.079	1.25 [0.93,1.69]	0.027
rs11209032	inter-genic	0.402	0.323	0.014	1.41 [1.05,1.90]	0.032
rs1495965	inter-genic	0.539	0.454	0.012	1.40 [1.05,1.87]	0.046

Table 2

Family-based association results using 65 trios of SNPs previously investigated by Duerr *et al*13. Family-based association one-sided P-values were computed using a binomial exact transmission disequilibrium test. r^2 values, derived from HapMap20, are shown with respect to each SNP's linkage disequilibrium with rs11209026.

Marker	Location	One-sided P-value	r^2 to rs11209026 (HapMap)
rs1004819	intron	0.13	0.01
rs7517847	intron	0.11	0.013
rs10489629	intron	0.5	0.067
rs2201841	intron	0.39	0.027
rs11465804	intron	0.011	0.881
rs11209026	Arg381Gln	0.0032	1
rs1343151	intron	0.097	0.111
rs10889677	exon-3′UTR	0.39	0.027
rs11209032	inter-genic	0.21	0.032
rs1495965	inter-genic	0.22	0.046