

RESEARCH ARTICLE

# Dissecting Allele Architecture of Early Onset IBD Using High-Density Genotyping

David J. Cutler<sup>1☯‡</sup>, Michael E. Zwick<sup>1☯‡</sup>, David T. Okou<sup>2☯‡</sup>, Sampath Prahallad<sup>1,2</sup>, Thomas Walters<sup>3</sup>, Stephen L. Guthery<sup>4</sup>, Marla Dubinsky<sup>5</sup>, Robert Baldassano<sup>6</sup>, Wallace V. Crandall<sup>7</sup>, Joel Rosh<sup>8</sup>, James Markowitz<sup>9</sup>, Michael Stephens<sup>10</sup>, Richard Kellermayer<sup>11</sup>, Marian Pfefferkorn<sup>12</sup>, Melvin B. Heyman<sup>13</sup>, Neal LeLeiko<sup>14</sup>, David Mack<sup>15</sup>, Dedrick Moulton<sup>16</sup>, Michael D. Kappelman<sup>17</sup>, Archana Kumar<sup>2</sup>, Jarod Prince<sup>2</sup>, Promita Bose<sup>1</sup>, Kajari Mondal<sup>1</sup>, Dhanya Ramachandran<sup>1</sup>, John F. Bohnsack<sup>4</sup>, Anne M. Griffiths<sup>3</sup>, Yael Haberman<sup>18</sup>, Jonah Essers<sup>19</sup>, Susan D. Thompson<sup>18</sup>, Bruce Aronow<sup>18</sup>, David J. Keljo<sup>20</sup>, Jeffrey S. Hyams<sup>21</sup>, Lee A. Denson<sup>18</sup>, PRO-KIIDS Research Group<sup>22¶</sup>, Subra Kugathasan<sup>1,2\*</sup>

**1** Department of Human Genetics, Emory University, Atlanta, Georgia, United States of America, **2** Department of Pediatrics, Emory University, Atlanta, Georgia, United States of America, **3** Hospital for Sick Children, Toronto, Ontario, Canada, **4** Department of Pediatrics, University of Utah, Salt Lake City, Utah, United States of America, **5** Icahn School of Medicine, Mount Sinai Hospital, New York, New York, United States of America, **6** Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, **7** Nationwide Children's Hospital, Columbus, Ohio, United States of America, **8** Goryeb Children's Hospital, Morristown, New Jersey, United States of America, **9** Cohen Children's Medical Center, New Hyde Park, New York, United States of America, **10** Pediatric Gastroenterology, Mayo Clinic, Rochester, Minnesota, United States of America, **11** Baylor College School of Medicine, Texas Children's Hospital, Houston, Texas, United States of America, **12** Riley Children's Hospital, Indianapolis, Indianapolis, United States of America, **13** University of California, San Francisco, California, United States of America, **14** Hasbro Children's Hospital, Providence, Rhode Island, United States of America, **15** Children's Hospital of Eastern Ontario, Ottawa, Canada, **16** Vanderbilt Children's Hospital, Nashville, Tennessee, United States of America, **17** University of North Carolina, Chapel Hill, North Carolina, United States of America, **18** Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States of America, **19** Children's Hospital of Boston, Boston, Massachusetts, United States of America, **20** Children Hospital of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, **21** Connecticut Children's Medical Center, Hartford, Connecticut, United States of America, **22** PRO-KIIDS Research Group, New York, New York, United States of America

☯ These authors contributed equally to this work.

‡ These authors are co-first authors on this work.

¶ Membership of the PRO-KIIDS Research Group is listed in the Acknowledgments.

\* [skugath@emory.edu](mailto:skugath@emory.edu)



**OPEN ACCESS**

**Citation:** Cutler DJ, Zwick ME, Okou DT, Prahallad S, Walters T, Guthery SL, et al. (2015) Dissecting Allele Architecture of Early Onset IBD Using High-Density Genotyping. PLoS ONE 10(6): e0128074. doi:10.1371/journal.pone.0128074

**Editor:** Fabio Cominelli, CWRU/UH Digestive Health Institute, UNITED STATES

**Received:** October 2, 2014

**Accepted:** April 23, 2015

**Published:** June 22, 2015

**Copyright:** © 2015 Cutler et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data is available via Figshare (<http://dx.doi.org/10.6084/m9.figshare.1420590>).

**Funding:** This work was funded by Crohn's & Colitis Foundation of America; Grant Number: CCFA1878 | <http://www.cdfa.org/> to SK.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

### Background

The inflammatory bowel diseases (IBD) are common, complex disorders in which genetic and environmental factors are believed to interact leading to chronic inflammatory responses against the gut microbiota. Earlier genetic studies performed in mostly adult population of European descent identified 163 loci affecting IBD risk, but most have relatively modest effect sizes, and altogether explain only ~20% of the genetic susceptibility. Pediatric onset represents about 25% of overall incident cases in IBD, characterized by distinct disease physiology, course and risks. The goal of this study is to compare the allelic architecture of early onset IBD with adult onset in population of European descent.

## Methods

We performed a fine mapping association study of early onset IBD using high-density ImmunoChip genotyping on 1008 pediatric-onset IBD cases (801 Crohn's disease; 121 ulcerative colitis and 86 IBD undetermined) and 1633 healthy controls. Of the 158 SNP genotypes obtained (out of the 163 identified in adult onset), this study replicated 4% (5 SNPs out of 136) of the SNPs identified in the Crohn's disease (CD) cases and 0.8% (1 SNP out of 128) in the ulcerative colitis (UC) cases. Replicated SNPs implicated the well known *NOD2* and *IL23R*. The point estimate for the odds ratio (ORs) for *NOD2* was above and outside the confidence intervals reported in adult onset. A polygenic liability score weakly predicted the age of onset for a larger collection of CD cases ( $p < 0.03$ ,  $R^2 = 0.007$ ), but not for the smaller number of UC cases.

## Conclusions

The allelic architecture of common susceptibility variants for early onset IBD is similar to that of adult onset. This immunoChip genotyping study failed to identify additional common variants that may explain the distinct phenotype that characterize early onset IBD. A comprehensive dissection of genetic loci is necessary to further characterize the genetic architecture of early onset IBD.

## Introduction

The inflammatory bowel diseases (IBD) are complex, heritable disorders consisting of two main clinical forms: Crohn's disease (CD) and ulcerative colitis (UC). Both forms share several phenotypic features and genetic susceptibility and are mainly characterized by chronic and relapsing gut inflammation. CD can occur throughout the gastrointestinal tract, whereas UC affects only the colon. Inflammatory bowel diseases affect some 1.5 million Americans [1], and ~20% of cases occur in children [2]. There is no doubt that IBD is heritable, although the 400% rise in the incidence of IBD over the last 50 years argues for an important role of gene-environment or gene-diet interactions in its pathogenesis [3–6].

Recent genome-wide association studies (GWAS) and a recent meta analysis identified over 163 IBD-associated loci, mostly in adult population of European descent: two thirds of which are shared by CD and UC, while the remaining are unique to either CD or UC [7]. Variants in these loci correlate well with IBD risk but have relatively modest effect sizes and account for a minority of the total disease variance (13.6% CD, 7.5% UC [7]) suggesting much of the missing heritability still awaits discovery. The loci with the largest known effects, *NOD2* and *IL23R*, both have odds ratios of approximately 1.5, with *NOD2* affecting only CD and *IL23R* affecting both CD and UC [7]. With at least 11 exceptionally highly powered GWAS studies [7–17] there was compelling evidence that *NOD2* and *IL23R* were the only loci in the genome with odds ratios approaching 1.5 for IBD.

The natural history of IBD exhibits substantial heterogeneity. Pediatric onset IBD (defined as the onset under 18 years old) is characterized by a greater disease severity, a higher tendency for disease manifestation in the colon, more disease extension, a change in disease location over time, and a positive family history for IBD [18–23]. These differing pathologies of early onset IBD could arise from differences in genetically attributable risk compared to adult onset IBD. Early onset diseases including IBD are expected to be more influenced by genetic factors

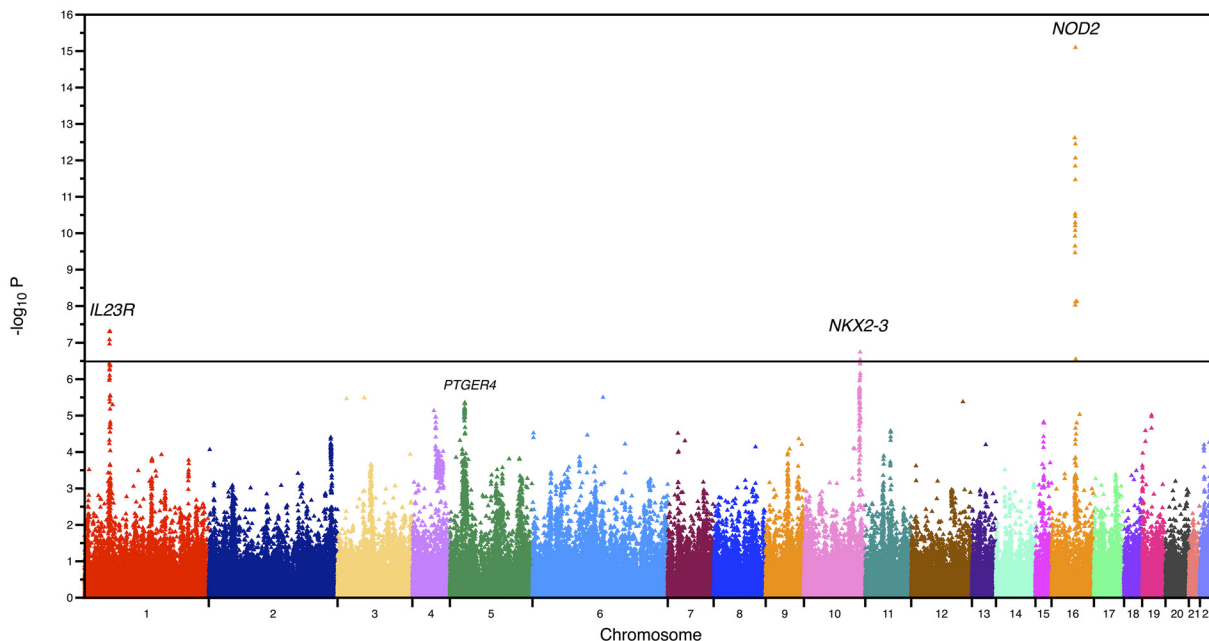
than environmental exposures [24]. Two previous pediatric GWAS studies, although underpowered, identified loci that have been replicated in adult onset IBD GWAS studies [12, 13]. Likewise, many established GWAS loci ascertained in adults did not distinguish early from later onset CD [25]. Here we used the ImmunoChip, a high-density custom array designed for fine mapping and deep replication of ~200,000 established genome-wide significant SNPs (within 186 loci) identified by GWAS for 12 autoimmune and inflammatory diseases, to determine how the allele architecture of early onset IBD compares to that of adult onset IBD.

## Results

We performed a genome-wide scan of our collection of early-onset CD cases vs controls (S1 Table) and early-onset UC vs controls (S2 Table). Fig 1 shows the results of CD cases vs controls for the ~140,000 SNPs that passed QC procedures. Highly conservative Bonferroni correction for multiple testing suggests any SNP p-value below  $0.05/140000 = 3.57 \times 10^{-7}$  is experiment-wide significant, which corresponds to  $-\log_{10}(p)$  of approximately 6.5. Some of the most significant loci are labeled for the CD vs controls analysis and are shown in Fig 1. None of the loci in the remaining cases (IBDminusCD) vs controls analysis exceeded our conservative experiment-wide threshold for statistical significance, which reflects the limited sample size of this portion of our study (S1 Fig).

After performing quality control, we successfully obtained genotypes for 158 of the top 163 SNPs from Jostins et al. [7] previously reported to be associated with CD and/or UC (S4 Table). For CD, 2 of 30 CD SNPs and 3 of 106 IBD SNPs successfully replicated at a Bonferroni corrected threshold of 0.0003. For UC, 0 of 22 UC SNPs and 1 of 106 IBD SNPs successfully replicated at a Bonferroni corrected threshold of 0.0003.

The limited number of Jostins et al. SNPs that replicated in our study is expected if reported effect sizes from a largely adult cohort are similar to those in our pediatric cohort. To better assess this pattern, we asked if our point estimate for the odds ratios (ORs) fell within the



**Fig 1. Manhattan plot of SNP association p-values result for CD vs controls within pediatric onset IBD.** The horizontal black line represent the thresholds of  $P = 3.57 \times 10^{-7}$  for Bonferroni significance

doi:10.1371/journal.pone.0128074.g001

confidence intervals reported in the more highly powered study reported by Jostins et al. For CD, the ORs for 128 SNPs were included within their respective Jostins' confidence intervals, while ORs for 8 SNPs fell outside the Jostins' estimated confidence intervals (S3 Table). The point estimates for the CD ORs from our study were greater for 3 of 8 SNPs as compared to those reported in Jostins et al. Of note, our OR point estimate for *NOD2*, the most strongly associated locus in both studies, is modestly larger and outside the CIs reported by Jostins (1.9 vs 1.6). A similar pattern was observed for UC, where the ORs for 118 SNPs were included within the CIs, while the ORs for 10 SNPs were found outside the Jostins' CIs (S4 Table). The point estimates for the UC ORs from our study were greater for only 1 of 10 SNPs as compared to those reported in Jostins et al. We next tested whether a polygenic liability score could predict the age of onset for either CD or UC. For CD, we observed a weak, but statistically significant relationship ( $p < 0.03$ ,  $R^2 = 0.007$ , S6 Fig). For UC, we did not observe a statistically significant relationship (S7 Fig).

We sought next to assess the genetic architecture differences between very early onset (VEO) cases (0 to < 10 years) as compared early onset (EO) cases (10 to < 17 years). The analysis of CD did not reveal any genome-wide significant findings (S2 Fig, S5 Table) between VEO and EO cases. Among the top findings, a SNP in the *IL-19* locus reached a nominal significance of  $7 \times 10^{-5}$ . The remaining cases analysis (IBDminusCD) also failed to reveal any statistically significant findings (S3 Fig and S6 Table).

## Discussion

There is strong evidence to support the genetic contribution and heritability to the pathogenesis of IBD [26–33]. Recent efforts to understand the genetics of IBD have identified over 163 IBD-associated loci, mostly in adult populations of European descent [7]. These results explain only 13.6% of CD and 7.5% of UC total disease variance, showing that these GWAS loci contribute only little towards the explained IBD heritability. The early onset IBD population presents with a distinct disease physiology, course and risks [18–23], suggesting the possibility of a different underlying molecular mechanism [34]. To date, GWAS has failed to identify loci that can distinguish early onset from adult IBD [12, 13, 25]. To determine how the allele architecture of early onset compares to that of adult onset, we performed fine mapping of established GWAS loci in a well characterized prospectively recruited pediatric onset IBD cases and control study, using a high-density custom array (ImmunoChip).

Our strongest associations for CD replicate alleles found at the *NOD2* and *IL23R* loci. This study suggests that common risk variants at both *NOD2* and *IL23R* loci act in a similar fashion in both early and later onset populations. This also confirms previous reported associations of *NOD2* and *IL23R* with both pediatric and adult onset CD [35–37]. Of interest, our OR point estimate for *NOD2*, the most strongly associated locus in both studies, is modestly larger and slightly outside the CIs reported by Jostins (1.9 vs 1.6). IBD is a complex disease whose heritability is influenced by only a small number of loci with large effect sizes (in this case *NOD2* and *IL23R*), but a relatively large number of loci whose effects are detectable in very large GWAS. This scenario has in fact been reported for common trait under biotic selection [38].

Only few in this study reached genome-wide significance threshold. Most of the loci previously reported by Jostins et al. did not replicate in our smaller pediatric cohort. For CD, we observed a weak, but statistically significant relationship between a polygenic liability score and age of onset ( $p < 0.03$ ,  $R^2 = 0.007$ , S6 Fig). For UC, we did not observe a statistically significant relationship (S7 Fig). At the vast majority of loci, our estimated effect sizes fell within the confidence intervals reported in adult onset [7]. Over 91% (128 SNPs) of the SNPs that explain slightly less than 20% of the CD heritability in the adult onset had similar estimated effect sizes

in early onset CD cases. Similarly, close to 89% (118 SNPs) of those UC SNPs had similar estimated effect sizes in early onset UC cases.

While our sample size was significantly smaller than that used by Jostins et al, this finding suggests an overall similar genetic architecture between adult and pediatric onset CD, IBD, and UC. Consistent estimates of SNP effect sizes in separate adult and early onset IBD cohorts suggests that there are likely to be great similarities in the overall pathogenesis and underlying molecular pathways of IBD. Furthermore, if early-onset susceptibility alleles exist, they will be difficult to discover in the absence of extremely large studies. As most common loci ascertained by GWAS, the loci in this study have small effect size, which are known to additively explain only a small proportion of the heritability of complex diseases [39]. One explanation for the “hidden heritability” is that additional or true causal variants could be present within known loci but they may be poorly tagged and cannot be revealed by the GWAS SNPs surveyed [40]. If true, this suggests that some of the reported clinical heterogeneity within IBD and differences in phenotypic characteristics between early and late onset could be driven by low frequency to rare variants, which will only be discovered through whole-exome sequencing, whole-genome sequencing, or family-based studies [41].

Pediatric onset IBD can be further separated into 2 distinct sub-groups based on the Paris classification; very early onset (VEO) 0 to < 10 years and early onset (EO) 10 to < 17 years [42]. Very early onset (VEO) IBD is characterized by a greater disease severity, a higher tendency for disease manifestation in the colon, more disease extension, a change in disease location over time, and a positive family history for IBD [18–23]. Our assessment of the genetic architecture differences between VEO and EO IBD cases revealed interesting differences in the allele architecture but no genome-wide significant findings (Figs S2, S3 and S5 and S6 Tables), which may reflect the limited analysis power for this sub-groups. One SNP in the *IL-19* locus reached a nominal significance for VEO CD analysis. *IL-19* belongs to *IL-10* superfamily and is known to suppress bacteria induced macrophage pro-inflammatory cytokine production in the host following injury to intestinal epithelial cells [43]. *IL10* is an established genetic risk for IBD, including a role of *IL10* receptor variants in VEO and EO IBD [44, 45]. It is not clear if the nominal significant *IL19* signal is independent or exhibited via *IL10* as both loci are linked together through genetic similarity and intron-exon gene structure. Owing to the limitation of GWAS and the expectation of higher genetic load and reduced environmental modifiers in VEO cases [24, 46], the allelic architecture of VEO must be characterized by low to rare frequency variants with greater penetrance not detectable by GWAS, or are simply located in portions of the genome not previously associated with IBD, and therefore not represented on the Immunochip. Rare variants are predicted to vastly outnumber common variants in the human genome [47].

This study showed that the allelic architecture of common susceptibility variants seen in adult onset is also shared with early onset cases, suggesting similarities in the overall pathogenesis of IBD. GWAS have limited utility in identifying low frequency to rare variants and true causal genetic variants associated with distinct phenotypes within the inflammatory bowel complex disease. It is therefore necessary that a more comprehensive dissection of the human genome or known loci be done to further understand the genetic architecture of early onset IBD. Sequencing-based methods are better at systematically ascertaining both common and rare SNPs.

## Materials and Methods

### Samples

We present a multi-center collaborative study of pediatric inflammatory bowel disease (the RISK cohort). IBD subjects were obtained from the RISK Study. The RISK Study is an ongoing,



prospective observational inception IBD cohort funded by the Crohn's and Colitis Foundation of America (CCFA) and currently includes 28 pediatric gastroenterology centers in North America. Children and adolescents younger than 17 years newly diagnosed with inflammatory bowel disease (IBD) were eligible for enrollment in RISK between November 2008 and June 2012. All subjects presented are of European ancestry.

All patients were required to undergo baseline colonoscopy and confirmation of characteristic chronic active colitis/ileitis by histology prior to diagnosis and treatment, with the recording of findings in standardized fashion. Once standard and published guidelines were met, patients were diagnosed with CD and UC and inflammatory bowel disease-undetermined (IBD-U). A firm and consistent diagnosis of IBD was required during the one-year follow-up for inclusion into this study. At enrollment and during ongoing prospective follow-up, clinical and laboratory data were obtained for each enrolled patient and submitted to a centralized data management center. All patients were managed according to the dictates of their physicians, not by standardized protocols.

The study populations were derived from three sources, including 1) 1401 cases with IBD (after QC and removal of non-Europeans: CD = 801; UC = 128; IBD-U = 86) recruited from the RISK cohort, (2) 1,663 healthy controls consisting of 831 pediatric healthy children recruited from Cincinnati Children's Hospital Medical Center (CCHMC), 647 healthy adults from Utah, and 185 non-IBD subjects from the RISK Study. Control subjects did not have any chronic autoimmune or inflammatory disorders. Most of the control subjects were part of other GWAS and Immunochip studies [48]. The institutional review board at Emory University approved the protocol. Written informed consent was provided by all parents/caregivers, and written assent was obtained from children as appropriate.

### High-density Genotyping with the Illumina Immunochip

The Immunochip, a custom Illumina Infinium High-Density array, contains 196,524 polymorphisms (718 small insertion/deletions, 195,806 SNPs). It was initiated by the Wellcome Trust Case-Control Consortium and designed for deep replication of established autoimmune and inflammatory disease loci identified by GWAS of common variants using data from the 1000 Genomes Project and any other available disease-specific resequencing data. The Immunochip Consortium selected 186 distinct loci containing markers reaching genome-wide significance ( $P < 5 \times 10^{-8}$ ) from 12 diseases (autoimmune thyroid disease, ankylosing spondylitis, Crohn's disease, celiac disease, IgA deficiency, multiple sclerosis, primary biliary cirrhosis, psoriasis, rheumatoid arthritis, systemic lupus erythematosus, type 1 diabetes, and ulcerative colitis). For each disease, ~3,000 SNPs were selected from available GWAS data for deep replication, as well as to cover strong candidate genes. Samples were genotyped using the Immunochip according to Illumina's protocols at laboratories at Emory University and The Feinstein Institute for Medical Research and Cincinnati Children's Hospital Medical Center (Utah samples and Cincinnati controls).

### Genotype Determination, Data Cleaning, and Statistical Analyses

All chip images were merged into a single batch for simultaneous genotype calling with BeadStudio. Included in these samples were 133 replicates (the same sample run twice), and 33 parent-offspring pairs. These replicates and family members were used to directly test for genotyping error [49]. Samples were tested for cryptic (unexpected) relatedness, incorrect genders, overall data completeness, and overall heterozygosity; samples were excluded if they had less than 90% data completeness, differed by more than three standard deviations from the mean heterozygosity for the study, had the wrong gender, or were unexpectedly the first-degree

relative of any other sample in the study. Approximately 15% of our samples were dropped for one or more of these reasons.

BeadStudio reported genotypes for 189,012 autosomal SNPs. Any SNP with more than 2% missing data, only one allele present (i.e. not segregating), any detectable genotype error (either through duplicates or parent-offspring), or a Hardy-Weinberg  $p$ -value less than  $10^{-5}$  in controls was dropped. This resulted in a final dataset of approximately 140,000 autosomal SNPs. Over 40% of the dropped SNPs were dropped simply because there was only one allele present (~20,600 SNPs).

For this study, we performed three main association analyses. We first evaluated patients with a diagnosis of CD with matched controls. We then evaluated all remaining cases (IBDminusCD), which consisted of cases with UC or other indeterminate IBD diagnosis (but not CD). Finally, we performed association analyses using very early onset (VEO) CD or UC / IBD-U labeled as “cases” contrasted with matched early onset (EO) CD or UC / IBD-U labeled as “controls”. Matching of cases and controls was done by determining principal components (PC) with Eigenstrat [50, 51], plotting PC1 against PC2, followed by visual inspection and elimination of outlier samples. Four successive rounds were necessary until a satisfactory matching of CD cases/controls (S4 Fig) and UC (IBDminusCD) cases/controls (S5 Fig) was obtained.

After all SNP and sample removal, 1,633 controls, 801 CD, and 207 UC / IBD-U individuals remained. From the 801 CD cases, we performed a second association with 267 VEO CD “cases” contrasted with 525 EO CD “controls”. From the 207 UC cases, there were 62 VEO UC / IBD-U “cases” contrasted with 143 EO UC / IBD-U “controls”. We excluded 7 CD and 2 UC samples because of indeterminate age of diagnosis. All samples were unrelated to one another. All association analysis was performed with PLINK 1.0.7 via a logistic regression [52], additive model, adjusting for the first five principal components of ancestry as determined by Eigenstrat. Replication of the top SNPs from Jostins et al. successfully genotyped in this study was assessed by performing a Bonferroni correction for 158 tests (0.05/158), resulting in a threshold of 0.0003. The complete summary for all SNP associations for CD (S1 Table) and UC (S2 Table) are contained in the supplemental materials. The summary for the 158 top SNPs for CD (S3 Table) and UC (S4 Table) are also included in the supplemental materials.

Polygenic liability scores were calculated assuming an underlying normal distribution of liability with disease (CD or UC) state representing a threshold on the continuous liability scale [53]. To do so, we first assumed the odds ratio estimated by Jostins' et al [7] is the true odds ratio for the identified allele. Using the observed allele frequency in controls from this study, and an assumed prevalence for CD of 5 in 10,000, and for UC of 1 in 10,000, independent of sex, the additive effect on liability of each of the (163) Jostins' identified SNPs was calculated [53]. Final polygenic liability for each sample was calculated by summing the additive effects for both alleles at all 163 loci. Thus we assume both additive dominance and additive epistasis on the liability scale. This polygenic liability score was regressed against age of onset for all CD and UC cases separately. While our cohort is purely pediatric onset (less than 18 years) only about 10% of the Jostins et al cohort were under 18 years of age at onset (personal communication with Dr Judy Cho).

The data of this study have been deposited into the Odum Institute Dataverse Network hosted at UNC (<http://arc.irss.unc.edu/dvn/>) and is accessible through the accession number doi:10.15139/S3/11991.

## Supporting Information

**S1 Fig. Manhattan plot of SNP association  $p$ -values result for IBD-minusCD vs Controls within pediatric onset IBD.** The horizontal black line represent the thresholds of  $P = 3.57 \times$

$10^{-7}$  for Bonferroni significance.  
(TIFF)

**S2 Fig. Manhattan plot of the association analysis for CD between VEO taken as cases and EO taken as “control”.** The horizontal black line represent the thresholds of  $P = 3.57 \times 10^{-7}$  for Bonferroni significance.

(TIFF)

**S3 Fig. Manhattan plot of the association analysis for IBD-minusCD between VEO taken as cases and EO taken as “control”.** The horizontal black line represent the thresholds of  $P = 3.57 \times 10^{-7}$  for Bonferroni significance.

(TIFF)

**S4 Fig. Principal component analysis (PCA) plot of the pediatric CD and control group.** Yellow and blue dots represent CD and control respectively.

(TIFF)

**S5 Fig. Principal component analysis (PCA) plot of the pediatric IBD-minusCD and control group.** Yellow and blue dots represent IBD-minusCD and control respectively.

(TIFF)

**S6 Fig. Relationship Between Polygenic Liability Score and Age of Onset of CD.**

(TIFF)

**S7 Fig. Relationship Between Polygenic Liability Score and Age of Onset of UC.**

(TIFF)

**S1 Table. Summary of SNP association findings for CD analysis.**

(CSV)

**S2 Table. Summary of SNP association findings for UC analysis.**

(CSV)

**S3 Table. Summary of CD association findings from this study for top SNPs reported in Jostins et al. 2012**

(XLSX)

**S4 Table. Summary of UC association findings from this study for top SNPs reported in Jostins et al. 2012**

(XLSX)

**S5 Table. Summary of association findings from this study for CD comparing very early onset (VEO) with early onset (EO) patients.**

(CSV)

**S6 Table. Summary of association findings from this study for UC comparing very early onset (VEO) with early onset (EO) patients.**

(CSV)

## Acknowledgments

PRO-KIIDS Research Group

Kugathasan, Subra <sup>§</sup> [skugath@emory.edu](mailto:skugath@emory.edu)

Rabizadeh, Shervin [Shervin.Rabizadeh@cshs.org](mailto:Shervin.Rabizadeh@cshs.org)

Baldassano, Robert [baldassano@email.chop.edu](mailto:baldassano@email.chop.edu)



Denson, A Lee [lee.denson@cchmc.org](mailto:lee.denson@cchmc.org)  
Hyams, S. Jeffrey [jhyams@ccmckids.org](mailto:jhyams@ccmckids.org)  
Noe, Joshua [jnoe@mcw.edu](mailto:jnoe@mcw.edu)  
Markowitz, James [jmarkowi@nshs.edu](mailto:jmarkowi@nshs.edu)  
Heyman, B Melvin [mheyman@peds.ucsf.edu](mailto:mheyman@peds.ucsf.edu)  
Wallace, V. Crandall [wallace.crandall@nationwidechildrens.org](mailto:wallace.crandall@nationwidechildrens.org)  
Snapper, Scott [Scott.snapper@childrens.harvard.edu](mailto:Scott.snapper@childrens.harvard.edu)  
Kellermayer, Richard [kellerma@bcm.edu](mailto:kellerma@bcm.edu)  
Kappelman, D. Michael [michael\\_kappelman@med.unc.edu](mailto:michael_kappelman@med.unc.edu)  
Griffiths, M. Anne [anne.griffiths@sickkids.ca](mailto:anne.griffiths@sickkids.ca)  
Walters, Thomas [thomas.walters@sickkids.ca](mailto:thomas.walters@sickkids.ca)  
Rosh, Joel [joel.rosh@atlantichhealth.org](mailto:joel.rosh@atlantichhealth.org)  
Otley, Anthony [arotley@dal.ca](mailto:arotley@dal.ca)  
Pfefferkorn, Marian [mdelrosa@iu.edu](mailto:mdelrosa@iu.edu)  
Cohen, Stanley [stancohen@aol.com](mailto:stancohen@aol.com)  
Guthery, L. Stephen [stephen.guthery@hsc.utah.edu](mailto:stephen.guthery@hsc.utah.edu)  
LeLeiko, Neal [nleleiko@lifespan.org](mailto:nleleiko@lifespan.org)  
Oliva-Hemker, Maria [moliva@jhmi.edu](mailto:moliva@jhmi.edu)  
Keljo, J. David [david.keljo@chp.edu](mailto:david.keljo@chp.edu)  
Moulton, Dedrick [dedrick.moulton@vanderbilt.edu](mailto:dedrick.moulton@vanderbilt.edu)  
Kirschner, Barbara [bkirschn@peds.bsd.uchicago.edu](mailto:bkirschn@peds.bsd.uchicago.edu)  
Ashish, Patel [ashish.patel@childrens.com](mailto:ashish.patel@childrens.com)  
Ziring, David [dziring@mednet.ucla.edu](mailto:dziring@mednet.ucla.edu)  
Evans, Jonathan [jevans@nemours.org](mailto:jevans@nemours.org)  
Baker, Susan [sbaker@upa.chob.edu](mailto:sbaker@upa.chob.edu)  
Mack, David [dmack@cheo.on.ca](mailto:dmack@cheo.on.ca)  
Dubinsky, Marla [marla.dubinsky@mssm.edu](mailto:marla.dubinsky@mssm.edu)  
§Lead author for the "PRO-KIIDS Research Group

## Author Contributions

Conceived and designed the experiments: SK MEZ DJC. Analyzed the data: DJC MEZ DTO SK. Contributed reagents/materials/analysis tools: SK MEZ DJC. Wrote the paper: DJC MEZ DTO LAD SK. Contributed samples and made critical revision of manuscript: SK LAD SP TW SLG MD RB WVC JR JM MS RK MP MBH NL D. Mack D. Moulton MDK JFB AMG JE SDT BA DJK JSH PKRG. Acquired and processed data: AK JP PB KM DR YH.

## References

1. Loftus EV. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology*. 2004; 126(6):1504–17. doi: [10.1053/j.gastro.2004.01.063](https://doi.org/10.1053/j.gastro.2004.01.063) PMID: [15168363](https://pubmed.ncbi.nlm.nih.gov/15168363/)
2. Beattie R, Croft N, Fell J, Afzal N, Heuschkel R. Inflammatory bowel disease. *Archives of disease in childhood*. 2006; 91(5):426–32. doi: [10.1136/adc.2005.080481](https://doi.org/10.1136/adc.2005.080481) PMID: [16632672](https://pubmed.ncbi.nlm.nih.gov/16632672/)
3. Bach JF. The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med*. 2002; 347(12):911–20. doi: [10.1056/NEJMra020100](https://doi.org/10.1056/NEJMra020100) PMID: [12239261](https://pubmed.ncbi.nlm.nih.gov/12239261/)
4. Cosnes J, Gower-Rousseau C, Seksik P, Cortot A. Epidemiology and natural history of inflammatory bowel diseases. *Gastroenterology*. 2011; 140(6):1785–94. doi: [10.1053/j.gastro.2011.01.055](https://doi.org/10.1053/j.gastro.2011.01.055) PMID: [21530745](https://pubmed.ncbi.nlm.nih.gov/21530745/)
5. Khor B, Gardet A, Xavier R. Genetics and pathogenesis of inflammatory bowel disease. *Nature*. 2011; 474(7351):307–17. doi: [10.1038/nature10209](https://doi.org/10.1038/nature10209) PMID: [21677747](https://pubmed.ncbi.nlm.nih.gov/21677747/)

6. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. 2012; 142(1):46–54 e42; quiz e30. doi: [10.1053/j.gastro.2011.10.001](https://doi.org/10.1053/j.gastro.2011.10.001) PMID: [22001864](https://pubmed.ncbi.nlm.nih.gov/22001864/)
7. 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(7422):119–24. doi: [10.1038/nature11582](https://doi.org/10.1038/nature11582) PMID: [WOS:000310434500042](https://pubmed.ncbi.nlm.nih.gov/WOS:000310434500042/)
8. Duerr R, Taylor K, Brant S, Rioux J, Silverberg M, Daly M, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006; 314(5804):1461–3. doi: [10.1126/science.1135245](https://doi.org/10.1126/science.1135245) PMID: [17068223](https://pubmed.ncbi.nlm.nih.gov/17068223/)
9. 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(5):596–604. doi: [10.1038/ng2032](https://doi.org/10.1038/ng2032) PMID: [17435756](https://pubmed.ncbi.nlm.nih.gov/17435756/); PubMed Central PMCID: [PMC2757939](https://pubmed.ncbi.nlm.nih.gov/PMC2757939/)
10. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, 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(4):e58. doi: [10.1371/journal.pgen.0030058](https://doi.org/10.1371/journal.pgen.0030058) PMID: [17447842](https://pubmed.ncbi.nlm.nih.gov/17447842/); PubMed Central PMCID: [PMC1853118](https://pubmed.ncbi.nlm.nih.gov/PMC1853118/)
11. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008; 40(8):955–62. doi: [10.1038/ng.175](https://doi.org/10.1038/ng.175) PMID: [18587394](https://pubmed.ncbi.nlm.nih.gov/18587394/); PubMed Central PMCID: [PMC2574810](https://pubmed.ncbi.nlm.nih.gov/PMC2574810/)
12. Kugathasan S, Baldassano RN, Bradfield JP, Sleiman PM, Imielinski M, Guthery SL, et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nat Genet*. 2008; 40(10):1211–5. doi: [10.1038/ng.203](https://doi.org/10.1038/ng.203) PMID: [18758464](https://pubmed.ncbi.nlm.nih.gov/18758464/); PubMed Central PMCID: [PMC2770437](https://pubmed.ncbi.nlm.nih.gov/PMC2770437/)
13. Imielinski M, Baldassano RN, Griffiths A, Russell RK, Annese V, Dubinsky M, et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat Genet*. 2009; 41(12):1335–40. doi: [10.1038/ng.489](https://doi.org/10.1038/ng.489) PMID: [19915574](https://pubmed.ncbi.nlm.nih.gov/19915574/); PubMed Central PMCID: [PMC3267927](https://pubmed.ncbi.nlm.nih.gov/PMC3267927/)
14. McGovern D, Gardet A, Torkvist L, Goyette P, Essers J, Taylor K, et al. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat Genet*. 2010; 42(4):332–7. doi: [10.1038/ng.549](https://doi.org/10.1038/ng.549) PMID: [20228799](https://pubmed.ncbi.nlm.nih.gov/20228799/)
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(12):1118–25. doi: [10.1038/ng.717](https://doi.org/10.1038/ng.717) PMID: [21102463](https://pubmed.ncbi.nlm.nih.gov/21102463/); PubMed Central PMCID: [PMC3299551](https://pubmed.ncbi.nlm.nih.gov/PMC3299551/)
16. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet*. 2011; 43(3):246–52. doi: [10.1038/ng.764](https://doi.org/10.1038/ng.764) PMID: [21297633](https://pubmed.ncbi.nlm.nih.gov/21297633/); PubMed Central PMCID: [PMC3084597](https://pubmed.ncbi.nlm.nih.gov/PMC3084597/)
17. Kenny E, Pe'er I, Karban A, Ozelius L, Mitchell A, Ng S, et al. A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS Genet*. 2012; 8(3):e1002559. doi: [10.1371/journal.pgen.1002559](https://doi.org/10.1371/journal.pgen.1002559) PMID: [22412388](https://pubmed.ncbi.nlm.nih.gov/22412388/)
18. Gryboski JD. Crohn's disease in children 10 years old and younger: comparison with ulcerative colitis. *J Pediatr Gastroenterol Nutr*. 1994; 18(2):174–82. PMID: [8014764](https://pubmed.ncbi.nlm.nih.gov/8014764/)
19. Seidman E. Chronic Inflammatory Bowel Disease. In: Roy OC ASaDA, editor. *Pediatric Clinical Gastroenterology*. 4th ed: Mosby; 1995. p. 427–93.
20. Polito JM 2nd, Childs B, Mellits ED, Tokayer AZ, Harris ML, Bayless TM. Crohn's disease: influence of age at diagnosis on site and clinical type of disease. *Gastroenterology*. 1996; 111(3):580–6. PMID: [8780560](https://pubmed.ncbi.nlm.nih.gov/8780560/).
21. Bianco AM, Zanin V, Girardelli M, Magnolato A, Martelossi S, Tommasini A, et al. A common genetic background could explain early-onset Crohn's disease. *Medical hypotheses*. 2012; 78(4):520–2. doi: [10.1016/j.mehy.2012.01.023](https://doi.org/10.1016/j.mehy.2012.01.023) PMID: [22309886](https://pubmed.ncbi.nlm.nih.gov/22309886/)
22. Cannioto Z, Berti I, Martelossi S, Bruno I, Giurici N, Crovella S, et al. IBD and IBD mimicking enterocolitis in children younger than 2 years of age. *European journal of pediatrics*. 2009; 168(2):149–55. doi: [10.1007/s00431-008-0721-2](https://doi.org/10.1007/s00431-008-0721-2) PMID: [18546019](https://pubmed.ncbi.nlm.nih.gov/18546019/)
23. Ruemmele FM, El Khoury MG, Talbotec C, Maurage C, Mougenot JF, Schmitz J, et al. Characteristics of inflammatory bowel disease with onset during the first year of life. *J Pediatr Gastroenterol Nutr*. 2006; 43(5):603–9. PMID: [17130735](https://pubmed.ncbi.nlm.nih.gov/17130735/)
24. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science (New York, NY)*. 1994; 265(5181):2037–48.

25. Essers JB, Lee JJ, Kugathasan S, Stevens CR, Grand RJ, Daly MJ, et al. Established genetic risk factors do not distinguish early and later onset Crohn's disease. *Inflamm Bowel Dis*. 2009; 15(10):1508–14. doi: [10.1002/ibd.20922](https://doi.org/10.1002/ibd.20922) PMID: [19322901](https://pubmed.ncbi.nlm.nih.gov/19322901/); PubMed Central PMCID: PMC2768775
26. Tysk C, Lindberg E, Jarnerot G, Floderus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut*. 1988; 29(7):990–6. PMID: [3396969](https://pubmed.ncbi.nlm.nih.gov/3396969/); PubMed Central PMCID: PMC1433769
27. Thompson NP, Driscoll R, Pounder RE, Wakefield AJ. Genetics versus environment in inflammatory bowel disease: results of a British twin study. *Bmj*. 1996; 312(7023):95–6. PMID: [8555939](https://pubmed.ncbi.nlm.nih.gov/8555939/); PubMed Central PMCID: PMC2349773
28. Orholm M, Binder V, Sorensen TI, Rasmussen LP, Kyvik KO. Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand J Gastroenterol*. 2000; 35(10):1075–81. PMID: [11099061](https://pubmed.ncbi.nlm.nih.gov/11099061/)
29. Orholm M, Munkholm P, Langholz E, Nielsen OH, Sorensen TI, Binder V. Familial occurrence of inflammatory bowel disease. *N Engl J Med*. 1991; 324(2):84–8. doi: [10.1056/NEJM199101103240203](https://doi.org/10.1056/NEJM199101103240203) PMID: [1984188](https://pubmed.ncbi.nlm.nih.gov/1984188/)
30. Peeters M, Nevens H, Baert F, Hiele M, de Meyer AM, Vlietinck R, et al. Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. *Gastroenterology*. 1996; 111(3):597–603. PMID: [8780562](https://pubmed.ncbi.nlm.nih.gov/8780562/)
31. Halme L, Paavola-Sakki P, Turunen U, Lappalainen M, Farkkila M, Kontula K. Family and twin studies in inflammatory bowel disease. *World J Gastroenterol*. 2006; 12(23):3668–72. PMID: [16773682](https://pubmed.ncbi.nlm.nih.gov/16773682/)
32. Yang H, Rotter JI. Genetic aspects of idiopathic inflammatory bowel disease. In: Kirsner JB, Shorter RG, editors. *Inflammatory Bowel Disease*. 4th ed: Williams & Wilkins; 1995. p. 301–31.
33. Satsangi J, Jewell DP, Bell JI. The genetics of inflammatory bowel disease. *Gut*. 1997; 40(5):572–4. PMID: [9203931](https://pubmed.ncbi.nlm.nih.gov/9203931/); PubMed Central PMCID: PMC1027155
34. Biank V, Broeckel U, Kugathasan S. Pediatric inflammatory bowel disease: clinical and molecular genetics. *Inflamm Bowel Dis*. 2007; 13(11):1430–8. doi: [10.1002/ibd.20213](https://doi.org/10.1002/ibd.20213). [17600381](https://pubmed.ncbi.nlm.nih.gov/17600381/) PMID: [17600381](https://pubmed.ncbi.nlm.nih.gov/17600381/)
35. de Ridder L, Weersma RK, Dijkstra G, van der Steege G, Benninga MA, Nolte IM, et al. Genetic susceptibility has a more important role in pediatric-onset Crohn's disease than in adult-onset Crohn's disease. *Inflamm Bowel Dis*. 2007; 13(9):1083–92. doi: [10.1002/ibd.20171](https://doi.org/10.1002/ibd.20171) PMID: [17476680](https://pubmed.ncbi.nlm.nih.gov/17476680/)
36. Amre DK, Mack D, Israel D, Morgan K, Lambrette P, Law L, et al. Association between genetic variants in the IL-23R gene and early-onset Crohn's disease: results from a case-control and family-based study among Canadian children. *The American journal of gastroenterology*. 2008; 103(3):615–20. Epub 2007/12/01. doi: [10.1111/j.1572-0241.2007.01661.x](https://doi.org/10.1111/j.1572-0241.2007.01661.x) PMID: [18047539](https://pubmed.ncbi.nlm.nih.gov/18047539/)
37. Latiano A, Palmieri O, Valvano MR, D'Inca R, Cucchiara S, Riegler G, et al. Replication of interleukin 23 receptor and autophagy-related 16-like 1 association in adult- and pediatric-onset inflammatory bowel disease in Italy. *World J Gastroenterol*. 2008; 14(29):4643–51. PMID: [18698678](https://pubmed.ncbi.nlm.nih.gov/18698678/); PubMed Central PMCID: PMC2738788
38. Louthan AM, Kay KM. Comparing the adaptive landscape across trait types: larger QTL effect size in traits under biotic selection. *BMC Evol Biol*. 2011; 11:60. doi: [10.1186/1471-2148-11-60](https://doi.org/10.1186/1471-2148-11-60) PMID: [21385379](https://pubmed.ncbi.nlm.nih.gov/21385379/); PubMed Central PMCID: PMC3061918
39. Manolio T, Collins F, Cox N, Goldstein D, Hindorf L, Hunter D, et al. Finding the missing heritability of complex diseases. *Nature*. 2009; 461(7265):747–53. doi: [10.1038/nature08494](https://doi.org/10.1038/nature08494) PMID: [19812666](https://pubmed.ncbi.nlm.nih.gov/19812666/)
40. Spencer C, Hechter E, Vukcevic D, Donnelly P. Quantifying the underestimation of relative risks from genome-wide association studies. *PLoS Genet*. 2011; 7(3):e1001337. doi: [10.1371/journal.pgen.1001337](https://doi.org/10.1371/journal.pgen.1001337) PMID: [21437273](https://pubmed.ncbi.nlm.nih.gov/21437273/); PubMed Central PMCID: PMC3060077
41. Cirulli E, Goldstein D. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet*. 2010; 11(6):415–25. doi: [10.1038/nrg2779](https://doi.org/10.1038/nrg2779) PMID: [20479773](https://pubmed.ncbi.nlm.nih.gov/20479773/)
42. Levine A, Griffiths A, Markowitz J, Wilson DC, Turner D, Russell RK, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflammatory bowel diseases*. 2011; 17(6):1314–21. Epub 2011/05/12. doi: [10.1002/ibd.21493](https://doi.org/10.1002/ibd.21493) PMID: [21560194](https://pubmed.ncbi.nlm.nih.gov/21560194/)
43. Azuma YT, Matsuo Y, Kuwamura M, Yancopoulos GD, Valenzuela DM, Murphy AJ, et al. Interleukin-19 protects mice from innate-mediated colonic inflammation. *Inflamm Bowel Dis*. 2010; 16(6):1017–28. Epub 2009/10/17. doi: [10.1002/ibd.21151](https://doi.org/10.1002/ibd.21151) PMID: [19834971](https://pubmed.ncbi.nlm.nih.gov/19834971/)
44. Glocker E-O, Kotlarz D, Boztug K, Gertz EM, Schäffer AA, Noyan F, et al. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *The New England journal of medicine*. 2009; 361(21):2033–45. doi: [10.1056/NEJMoa0907206](https://doi.org/10.1056/NEJMoa0907206) PMID: [19890111](https://pubmed.ncbi.nlm.nih.gov/19890111/)

45. Kotlarz D, Beier R, Murugan D, Diestelhorst J, Jensen O, Boztug K, et al. Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy. *Gastroenterology*. 2012; 143(2):347–55. doi: [10.1053/j.gastro.2012.04.045](https://doi.org/10.1053/j.gastro.2012.04.045) PMID: [22549091](https://pubmed.ncbi.nlm.nih.gov/22549091/)
46. Cosnes J, Beaugerie L, Carbonnel F, Gendre JP. Smoking cessation and the course of Crohn's disease: an intervention study. *Gastroenterology*. 2001; 120(5):1093–9. doi: [10.1053/gast.2001.23231](https://doi.org/10.1053/gast.2001.23231) PMID: [11266373](https://pubmed.ncbi.nlm.nih.gov/11266373/)
47. Marth GT, Yu F, Indap AR, Garimella K, Gravel S, Leong WF, et al. The functional spectrum of low-frequency coding variation. *Genome Biol*. 2011; 12(9):R84. Epub 2011/09/16. doi: [10.1186/gb-2011-12-9-r84](https://doi.org/10.1186/gb-2011-12-9-r84) PMID: [21917140](https://pubmed.ncbi.nlm.nih.gov/21917140/); PubMed Central PMCID: [PMC3308047](https://pubmed.ncbi.nlm.nih.gov/PMC3308047/)
48. Hinks A, Cobb J, Marion M, Prahalad S, Sudman M, Bowes J, et al. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat Genet*. 2013. doi: [10.1038/ng.2614](https://doi.org/10.1038/ng.2614) PMID: [23603761](https://pubmed.ncbi.nlm.nih.gov/23603761/)
49. Consortium IH. A haplotype map of the human genome. *Nature*. 2005; 437(7063):1299–320. doi: [10.1038/nature04226](https://doi.org/10.1038/nature04226) PMID: [16255080](https://pubmed.ncbi.nlm.nih.gov/16255080/)
50. Price A, Patterson N, Plenge R, Weinblatt M, Shadick N, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38(8):904–9. doi: [10.1038/ng1847](https://doi.org/10.1038/ng1847) PMID: [16862161](https://pubmed.ncbi.nlm.nih.gov/16862161/)
51. Patterson N, Price A, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006; 2(12):e190. doi: [10.1371/journal.pgen.0020190](https://doi.org/10.1371/journal.pgen.0020190) PMID: [17194218](https://pubmed.ncbi.nlm.nih.gov/17194218/)
52. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75. doi: [10.1086/519795](https://doi.org/10.1086/519795) PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/)
53. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014; 515(7526):209–15. doi: [10.1038/nature13772](https://doi.org/10.1038/nature13772) PMID: [25363760](https://pubmed.ncbi.nlm.nih.gov/25363760/); PubMed Central PMCID: [PMC4402723](https://pubmed.ncbi.nlm.nih.gov/PMC4402723/)