Exome Analysis of Patients with Concurrent Pediatric Inflammatory Bowel Disease and Autoimmune Disease

Gaia Andreoletti, MSc, James J. Ashton, BMBS, BMedSci,*,† Tracy Coelho, MRCPCH,*,† Claire Willis,‡ Rachel Haggarty,‡ Jane Gibson, PhD,§ John Holloway, PhD,* Akshay Batra, MRCPCH,† Nadeem A. Afzal, MD, MRCPCH,† Robert Mark Beattie,† and Sarah Ennis, PhD**

Background: Pediatric Inflammatory Bowel Disease (PIBD) is a chronic condition seen in genetically predisposed individuals. Genome-wide association studies have implicated >160 genomic loci in IBD with many genes coding for proteins in key immune pathways. This study looks at autoimmune disease burden in patients diagnosed with PIBD and interrogates exome data of a subset of patients.

Methods: Patients were recruited from the Southampton Genetics of PIBD cohort. Clinical diagnosis of autoimmune disease in these individuals was ascertained from medical records. For a subset of patients with PIBD and concurrent asthma, exome data was interrogated to ascertain the burden of pathogenic variants within genes implicated in asthma. Association testing was conducted between cases and population controls using the SKAT-O test.

Results: Forty-nine (28.3%) PIBD children (18.49% CD, 8.6% UC, and 21.15% IBDU patients) had a concurrent clinical diagnosis of at least one other autoimmune disorder; asthma was the most prevalent, affecting 16.2% of the PIBD cohort. Rare and common variant association testing revealed 6 significant genes ($P < 0.05$) before Bonferroni adjustment. Three of these genes were previously implicated in both asthma and IBD (*ZPBP2 IL1R1*, and *IL18R1*) and 3 in asthma only (*PYHIN1, IL2RB*, and *GSTP1*).

Conclusions: One-third of our cohort had a concurrent autoimmune condition. We observed higher incidence of asthma compared with the overall pediatric prevalence. Despite a small sample size, SKAT-O evaluated a significant burden of rare and common mutations in 6 genes. Variant burden suggests that a systemic immune dysregulation rather than organ-specific could underpin immune dysfunction for a subset of patients.

(Inflamm Bowel Dis 2015;21:1229–1236)

Key Words: pediatric inflammatory bowel disease, comorbidity, exome sequencing, autoimmune disorders, asthma, genetics

Pediatric inflammatory bowel disease (PIBD) encompasses Crohn's disease (CD), ulcerative colitis (UC), and inflammatory bowel disease unclassified, a group of complex and multifactorial illnesses. The etiology is complex and likely to comprise 4 key concepts: immune dysregulation, barrier dysfunction, microbial flora, and a genetic predisposition[1,2]; how specific interaction between these factors leads to development of disease is poorly understood.

The incidence of PIBD is increasing in Europe and North America. Recent studies from England,[3] Scotland,[4] and Scandanvia[5] have shown increasing incidence over the last 20 years with incidence as high as 12.8/100,000 person-years in Sweden.[6] This increase may be driven by lifestyle changes. The hygiene hypothesis relating to autoimmune conditions is well established in the literature, and it is conceivable that altered microbial exposure in these children may have a role in development of disease.[7,8]

Over the past decade, genome-wide association studies (GWAS) have substantially advanced the understanding of many complex diseases.[9] Since the discovery in 2001 of *NOD2*, the first genetic susceptibility gene for IBD[10,11] more than 160 distinct loci have been shown to have a robust association with IBD.[12–14] However, common variation in these genes account for only approximately one-fourth of the disease heritability.[15] It is now assumed that rare variants in pathways implicated across various autoimmune conditions may account for some of the missing heritability in IBD.[16] Various studies in adult populations have looked at the incidence of other autoimmune-mediated disorders in patients suffering from IBD.[17–19] Previous studies into the coexistence of PIBD

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.ibdjournal.org).

Received for publication January 14, 2015; Accepted February 3, 2015.

From the *Human Genetics and Genomic Medicine, University of Southampton, Southampton General Hospital, Southampton, United Kingdom; †Department of Paediatric Gastroenterology, University Hospital Southampton NHS Foundation Trust, Southampton General Hospital, Southampton, United Kingdom; ‡NIHR Nutrition Biomedical Research Centre, Southampton Centre for Biomedical Research, University Hospital Southampton NHS Foundation Trust, Southampton General Hospital, Southampton, United Kingdom; and §Centre for Biological Sciences, Faculty of Natural and Environmental Studies, University of Southampton, Southampton, United Kingdom.

Supported by the Crohn's in Childhood Research Association (CIRCA) and the Gerald Kerkut Charitable Trust. J. J. Ashton is supported by a University of Southampton National Institute of Health Research Academic Clinical Fellowship.

The authors have no conflicts of interest to disclose.

G. Andreoletti and J. J. Ashton have contributed equally to this study.

Reprints: Sarah Ennis, PhD, Human Genetics and Genomic Medicine, University of Southampton, Duthie Building (Mailpoint 808), Southampton General Hospital, Southampton SO16 6YD, United Kingdom (e-mail: s.ennis@southampton.ac.uk).

DOI 10.1097/MIB.0000000000000381

Published online 17 April 2015.

and other autoimmune disease have reported strong association between PIBD (both CD and UC) and rheumatoid arthritis, systemic lupus erythematosus, and hypothyroidism with a trend towards increased prevalence of other autoimmune conditions, including asthma and eczema.[20]

Predisposition to IBD and other autoimmune disease has a strong genetic component, and analyses of exomes of these patients may yield variations associated with both groups of disease. This study examines the autoimmune disease burden in patients diagnosed with PIBD and interrogates exome data of a subset of patients.

MATERIALS AND METHODS

Recruitment

Recruitment of children diagnosed with PIBD was through services at University Hospital Southampton. All children younger than 18 years at the point of diagnosis were eligible. Diagnosis was established according to the Porto criteria.[21] Clinical data were recorded for each patient including family history of IBD and any history of autoimmune disease.[22]

Patient Data Extraction

Data from 173 patients prospectively collected at recruitment were interrogated to identify patients with PIBD with: (1) comorbidity of other autoimmune diseases (clinician diagnosed) and (2) positive family history of autoimmune diseases other than IBD. Medical notes were consulted for any ambiguous diagnosis and review allowed for exclusion of any unconfirmed diagnoses.

DNA Extraction

Genomic DNA was extracted from peripheral venous blood samples collected in EDTA, using the salting out method.[23] DNA concentration was estimated using the Qubit 2.0 Fluorometer (Life Technologies Ltd). and 260:280 ratio calculated using a nanodrop spectrophotometer. The average DNA yield obtained was 150 μg/mL and approximately 20 μg of DNA was used for next generation sequencing for each patient.

Patient Selection

Of the 28 patients with IBD who had a concurrent diagnosis of asthma, we selected the 18 youngest of these for exome analysis. All 18 were of white British ancestry.

Exome Data Generation and Processing

Whole exome sequencing was performed using the Agilent SureSelect Human all Exon 51 Mb version 4 capture kit. The fastq raw data generated from Illumina paired-end sequencing were aligned against the human reference genome (hg19) using Novoalign (novoalign/2.08.02). Sequence coverage for each sample was calculated using the BedTools package (v2.13.2) (see Table, Supplemental Digital Content 1, http://links.lww.com/IBD/A809).

SAMtools[24] MPileup tool (SAMtools/0.1.18) was used to detect variation from the mapping information to call SNPs and short INDELs from the alignment file. Variations with read depth <4 were excluded. The Phred software[25] reads DNA sequencing trace files, calls bases, and assigns a quality value to each called base and is powered to discriminate between correct and incorrect base calls. Only good quality bases with a Phred score >20 were retained for analysis (99% base call accuracy). ANNOVAR (annovar/February 21, 2013)[26] was applied for variant annotation against a database of RefSeq transcripts. Resultant variants files for each subject were subjected to further in-house quality control tests to detect DNA sample contamination and ensure sex concordance by assessing autosomal and X chromosome heterozygosity. Variant sharing between all pairs of individuals was assessed to confirm sample relationships. Sample provenance was confirmed by independent genotyping of a validated SNP panel, developed specifically for exome data.[27]

Gene Selection

The latest genome-wide meta-analysis of IBD reported 193 genes across 163 loci with statistically independent signals of association at genome-wide significance ($P < 5 \times 10^{-8}$).[13] These genes were cross-referenced with 49 genes associated with asthma identified by linkage studies and GWAS.[28] Sixteen of these genes have been associated with both asthma and IBD (Fig. 1). Gene names were cross-referenced with the HUGO web server to confirm the approved gene symbol.

Variant Association Testing

Our findings and those of others[29,30] indicate asthma is the most common concurrent autoimmune disease in patients with IBD. For this reason, we wanted to further investigate if a subset of patients with a concurrent diagnosis of both IBD and asthma present with a significant burden of mutation within known genes associated with asthma. Our modest sample size was underpowered to extend this analysis to all 193 IBD genes.

To detect association between the genetic component and disease status, first, a single variant test and then a gene-based test (SKAT-O) were performed. To run these tests, genotype information (homozygous alternative, homozygous reference, or heterozygous status) were retrieved using customized scripts applying samtools,[24] vcftools,[31] and bedtools[32] packages. All variant sites across 49 genes (comprising 33 genes specific to asthma and 16 genes common to both diseases) were used to generate the variant call file for each of the 18 exome analyzed patients and 56 unrelated germline controls. Our genomics bioinformatics group has a rolling database of non-IBD clinical exomes. Controls without any clinical diagnosis of autoimmune disease were selected from this in-house database.

Variations were further excluded based on the Hardy–Weinberg equilibrium status ($P < 0.001$) in the control group, by using vcftools.[31] Variant call file files containing genotype information for all cases and controls were merged together and annotated.[33] Both single and joint analysis were carried out using the EPACTS software.[34]

Single Variant Association Testing

The single variant logistic score test[34] was performed to detect differences in variant frequency between cases and control

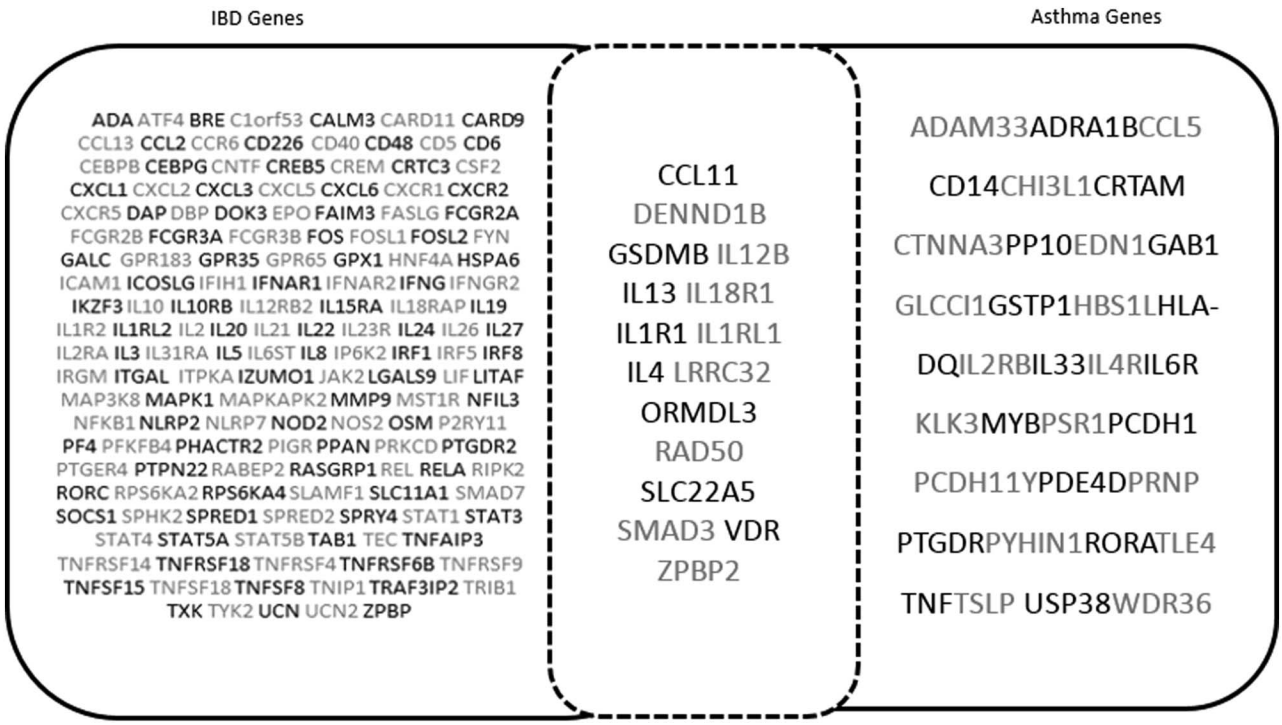

FIGURE 1. Overlap of GWAS significant gene loci in IBD (left) and asthma (right).

group. The test was not performed on mutations occurring in one individual in either case or control group.

Rare Variant Profile Filtering

The burden of rare and novel damaging variation was described for each of the 18 patients across 49 asthma genes. Synonymous variations were excluded from the analysis on the assumption of their low impact on protein function. All novel to individual, novel to Southampton PIBD cohort, and clinical variants as well as frameshift insertion, frameshift deletion, stop gain, and stop loss mutations were retained for further analysis. Novel to individual denotes variants not previously reported in dbSNP137 database, 1000 Genomes Project, Exome Variants Server (EVS) of European Americans in the NHLI-ESP project with 6500 exomes (http://evs.gs.washington.edu/EVS/) in 46 unrelated human subjects sequenced by Complete Genomics,[35] in other individuals of the Soton IBD cohort, or in the Southampton reference exome database. Novel to Soton cohort denotes variants not previously reported in dbSNP137 database,[36] 1000 Genomes Project,[37] Exome Variants Server of European Americans in the NHLI-ESP project with 6500 exomes (http://evs.gs.washington.edu/EVS/) in 46 unrelated human subjects sequenced by Complete Genomics[35] but has been seen in other individuals of the Soton IBD cohort.

To refine this list to variations most likely to have a biological impact, common variants occurring in ≥5% of individuals from 1000 genomes project[37] were excluded and variants less likely to impact on protein function as expressed by the logit categorical score[38] were excluded (logit = N). Pathways were determined using DAVID (Database for Annotation, Visualization and Integrated Discovery)[39] and KEGG pathway.[40]

Joint Variant Association Testing

The sequence kernel association testing optimal unified test (SKAT-O)[41] is a gene-based test for assessing the contribution of rare and common variations within a genomic loci with trait.[41] Specifically, SKAT-O encompasses both a burden test and a SKAT[41] test to offer a powerful way of conducting association analysis on combined rare and common variation as single variant tests are often underpowered because of the large sample size needed to detect a significant association.

SKAT-O was executed with the small sample adjustment and by applying an MAF threshold of 0.05 to define rare variations and using default weights. To conduct the test, a group file with mutations of interests (missense, nonsense, splice-site variants, and coding indels) was created for each of the 49 genes.

Ethical Considerations

The study has ethics approval from Southampton and South West Hampshire Research Ethics Committee (09/H0504/125).

RESULTS

Southampton PIBD Cohort

At the time of analysis, the Southampton PIBD study cohort comprised 173 children (98 CD, 55 UC, and 20

TABLE 1. PIBD Cohort Demographics

	CD	UC	IBDU	Total IBD
No. patients	98	55	20	173
Female, %	44	44	60	46
Median age of onset (25th/75th percentile)	12.28 (9.16/14.27)	11.5 (9.61/13.48)	15.33 (8.72/14.10)	12.28
Mean age of onset (SD)	11.58 (3.13)	10.87 (3.76)	11.50 (3.31)	11.35 (3.35)

inflammatory bowel disease unclassified); demographic data are shown in Table 1.

Prevalence of Comorbidity

Analysis of the cohort revealed concurrent diagnosis of PIBD with 12 distinct autoimmune-mediated conditions (Table 2).

Asthma (n = 28) and atopic dermatitis (n = 24) represented the conditions with the highest frequency. Additional cases of sclerosing cholangitis (n = 4), coeliac disease (n = 2), and vitiligo (n = 2) were present.

Forty-nine children (28.3%) presented with a second autoimmune condition. Across the cohort, there was a family history of asthma, atopic dermatitis, coeliac disease, sclerosing cholangitis, and vitiligo although no probands had diagnosis of these conditions at the time of analysis.

Single Variant Association Test for Variants in Asthma and Dual Susceptibility Genes

Among the 28 patients affected by asthma, 18 youngest patients were selected for exome sequencing (9 CD and 9 UC). Characteristics for each of the patients that underwent exome sequencing are presented in Table 3. Thirty six of the 49 genes either specific to asthma and common to both asthma and IBD were analyzed, as no coding variants were called in *ADRA1B, CCL5, CD14, HLA-DQ, IL12B, IL13, IL4, ORMDL3, PCDH1, RAD50, TNF, TSLP*, and *SLC22A5* across cases and controls, and these genes were excluded from single variant and joint testing.

A total of 175 different variants were identified across 36 genes in the cases and controls exomes. A total of 73 occurred only in 1 individual (case or control), and these were not analyzed in the single variant test. The single variant test was applied to 102 variants across 33 genes. Three of these variants showed significant association with disease status (association $P < 0.05$; *PYHIN1, ZPBP2*, and *LRRC32*). However, none of these variants would withstand multiple testing corrections (see Table, Supplemental Digital Content 2, http://links.lww.com/IBD/A810). *ZPBP2* and *LRRC32* are known to be involved in both asthma and IBD. Within these genes, *ZPBP2* nonsynonymous variant at position 38027030 bp and *LRRC32* synonymous variant at position 76372052 bp ($P = 0.011$ and $P = 0.043$, respectively) were found with higher frequency in cases compared with controls. *PYHIN1* is known to be involved in asthma pathogenesis. In this gene, the nonsynonymous variant at position 158943483bp ($P = 0.008$) was observed at higher frequency in cases compared with control group. The frequency of these mutations suggests their possible deleterious effect in increasing disease risk in genetically susceptible individuals.

Individual Profiles of Rare and Deleterious Variants

Individual burden of variation revealed 24 variants (see Table, Supplemental Digital Content 3, http://links.lww.com/IBD/A811). Several dual susceptibility genes (for PIBD and asthma) were identified as harboring one or more variants. Mutations fall within 3 pathways consistently reported in KEGG and DAVID. *ZPBP2* (zona pellucida binding protein 2[12]) and *SMAD3* (involved in the adherens junction pathway[48]) have variants observed across both CD and UC but these variants also occur in 1% and 2% of the 1000 genomes reference population.

TABLE 2. Prevalence of Autoimmune Disease in the PIBD Cohort (173 Patients)

Autoimmune Disease	CD (n = 98) (%)	UC (n = 55) (%)	IBDU (n = 20) (%)	Overall PIBD Cohort Prevalence (n = 173) (%)	Overall Population Pediatric Prevalence, %
Asthma	19 (19.40)	9 (16.40)	0	28 (16.18)	15.00[30,42,43]
Atopic dermatitis	18 (18.40)	6 (8.11)	0	24 (13.87)	16.50[44]
Coeliac disease	1 (1.020)	1 (1.35)	0	2 (1.15)	0.99[45]
Sclerosing cholangitis	1 (1.02)	2 (3.64)	1 (50)	4 (2.31)	0.01[a,46]
Vitiligo	1 (1.02)	0	1 (50)	2 (1.15)	1.00[47]

[a]Data for general population prevalence only and not specific to general IBD.

TABLE 3. Clinical Profile of the 18 Patients with Concurrent IBD and Asthma Selected for Exome Analysis

Study ID	Sex	Age at Diagnosis of IBD	Diagnosis	Paris Classification of IBD	Other Autoimmune Disease Status Other than Asthma
PR0007	M	11.41	CD	A1L34B1	
PR0011	M	15.51	CD	A1L1B2	
PR0031	M	11.43	CD	A1L3B1p	
PR0032	M	7.29	CD	A1L24B1p	Atopic dermatitis
PR0036	F	9.67	CD	A1L3B1	
PR0039	M	10.30	UC	E3-	
PR0068	F	11.23	UC	E2-	
PR0083	F	9.68	UC	E3S3	
PR0085	M	13.12	UC	E3S3	
PR0107	M	9.22	CD	A1L1-	
PR0110	F	2.98	UC	E3-	
PR0146	M	14.52	CD	A1L3-	
PR0148	M	9.13	CD	A1L34B3p	Atopic dermatitis
PR0151	F	13.30	CD	A1L24B1	
PR0158	F	15.25	UC	E3S3	Atopic dermatitis
PR0160	M	12.55	UC	E3S1	Atopic dermatitis
PR0167	M	13.30	UC	E2S2	Atopic dermatitis
PR0188	M	11.03	CD	A1L1-	

- denotes missing classification data.

Patient PR0085 not only carries the nonsynonymous *ZPBP2* mutation but also carries a novel frameshift deletion in *DENND1B* gene expressed by natural killer cells and dendritic cells.[14,49] The same patient harbors a nonsynonymous mutation at position 67353579 bp within *GSTP1*, which is reported to be involved in asthma pathogenesis only. Also of interest among the genes previously implicated in both diseases are 2 distinct and very rare mutations in the *RAD50* gene located within the IBD5 cytokine cluster on chromosome 5q31.[50] This gene contains the locus control region required for the Th2 cytokine gene expression.[51] In asthma specific genes, variations were found in 8 genes. A more common variant (rs3918396) is seen to recur within the *ADAM33* gene, a second variant in the same gene has been identified in PR0158 and other patients within the Southampton PIBD cohort.

PR0110 is a patient diagnosed aged 2 years with severe UC. She is seen to harbor a mutation that could impact splicing at position 8009439 bp in *GLCCI1* and a novel frameshift insertion at position 69407255 within *CTNNA3*. This gene encodes the α–T-catenin protein; a key component of the adherens junctional complex in epithelial cells necessary for cellular adherence.[52]

Joint Rare Variant Association Test for Variants in IBD and Dual Susceptibility Genes

The joint test for assessing the contribution of private, rare, and common mutation between disease status and genes highlighted 6 genes with a $P < 0.05$ before Bonferroni correction (Table 4).

Of these 6 genes, 3 are known susceptibility genes for both IBD and asthma (*ZPBP2*, $P = 0.009$; *IL1R1*, $P = 0.036$; and *IL18R1*, $P = 0.038$); the remaining genes were asthma specific (*PYHIN*, $P = 0.025$; *IL2RB*, $P = 0.036$; *GSTP1*, $P = 0.040$). These genes are all key determinants of the immune response and have variants observed across both CD and UC.

DISCUSSION

Our cohort of 173 children with PIBD revealed forty-nine children (28.3%) with a concurrent diagnosis of an autoimmune disease. Asthma and atopic dermatitis occurred with the highest frequency; the prevalence of clinically diagnosed asthma was 19.4% in children with CD and 16.4% in patients with UC, exceeding UK disease estimates (15.3%[53,54]).

Although this study is not powered to demonstrate a statistically significant increase in autoimmune disease burden in children with PIBD, our observations indicate prevalence estimates approaching the upper limit recorded in the literature.[20] Our findings are consistent with literature indicating that children with PIBD are more likely to have other autoimmune conditions, and that a common genetic components etiology may predispose individuals to multiple autoimmune manifestations.[20,29]

Even in a very modest cohort, SKAT-O association analysis revealed 6 genes with significant burden of mutation. Although significant levels would not withstand a Bonferroni correction for 36 genes tested, the strong prior hypothesis to the

TABLE 4. Joint Variant Test (SKAT-O) Result for the 36 Known Asthma Genes in Which Variations was Found Across the Entire Cohort

Gene set	Gene	Chr	Bp Position (hg19)	Total No. Samples, (18 Cases; 56 controls)	Fraction of Individuals Who Carry Rare Variants Under the MAF Thresholds (MAF <0.05)[a]	No. All Variants Defined in the Group File	No. Variant Defined as Rare (MAF <0.05)[a]	Unadjusted *P*
Asthma/IBD	ZPBP2	17	38024626-38032996	74	0.027	4	1	0.009
Asthma	PYHIN1	1	158906777-158943483	74	0.108	5	5	0.025
Asthma/IBD	IL1R1	2	102781629-158943483	74	0.014	5	2	0.037
Asthma	IL2RB	22	37524329-37539651	74	0.014	4	1	0.037
Asthma/IBD	IL18R1	2	102984279-103001402	74	0.014	3	1	0.039
Asthma	GSTP1	11	67352183-67353970	74	0.014	4	1	0.041
Asthma	TLE4	9	82187750-82336794	74	0.108	4	4	0.056
Asthma	NPSR1	7	34698177-34917768	74	0.135	12	4	0.066
Asthma	CTNNA3	10	67680203-69407255	74	0.162	6	3	0.081

Only genes with a $P < 0.1$ are shown.
No. variants were found in ADRA1B, CCL5, CD14, HLA-DQ, IL12B, IL13, IL4, ORMDL3, PCDH1, RAD50, TNF, TSLP, and SLC22A5 across cases and controls.
[a]These variants received different weights in the SKAT-O joint test.

analysis of these genes might suggest that such a multiple testing correction would be inappropriate.

ZPBP2 is located on the chr17q12-q21 region, which has been associated with early-onset asthma, and variants in the same linkage disequilibrium block have been associated with Crohn's disease, type 1 diabetes, and primary biliary cirrhosis.[55]

IL1R1 encodes for a cytokine receptor that belongs to the interleukin-1 receptor family. The gene was found associated with asthma in a GWAS on 933 European ancestry individuals with severe asthma based on Global Initiative for Asthma criteria.[56] At the same genomic region of *IL1R1*, *IL18R1* was also identified as associated with asthma. Specifically, the gene was evaluated in a GWAS conducted on Mexican pediatric patients.[57] The association was further replicated on a family-based study on Denmark, United Kingdom, and Norway families.[58]

GSTP1 is involved in the detoxification of a wide variety of exogenous and endogenous compounds, including reactive oxygen species. This gene was discovered by a GWAS conducted on early-onset asthma.[59] *IL2RB* is involved in lymphoid cell differentiation, and it was first discovered by GWAS conducted by the GABRIEL consortium in 2012.[60] *PYHIN1* (Pyrin And HIN Domain Family, Member 1) encodes a protein that belongs to the HIN-200 family of interferon-inducible proteins, important in controlling cell cycle, differentiation, and apoptosis.[61] It has been noted to be an asthma susceptibility locus, specifically in those of African descent.[62] *PYHIN1* was identified as associated with asthma in 2011, through a meta-analysis conducted on 5416 European American, African American, or African Caribbean, and Latino ancestry individuals with asthma. The *PYHIN1* association was specific to the African descent groups.[62]

PYHIN1 and *ZPBP2* were significantly associated with asthma in both single variant testing and after SKAT-O testing. Variants within these genes were found with higher frequency in cases compared with controls suggesting a deleterious role of the mutations in the pathogenesis of disease. Susceptibility genes for both IBD and asthma are most commonly involved with immune regulation raising the possibility of an overall immune dysregulation underlying both diseases. These genes may be implicated in the same pathways as found in other probands but may not yet have been associated with IBD/asthma or did not hold enough significance power to be included in the GWAS meta-analyses.

This study demonstrates robust data collection, all PIBD diagnoses are made using strict criteria,[21] and autoimmune comorbidity was validated through integration of the medical notes (paper and electronic). This study looked only at genes identified through GWAS (of asthma and IBD); this increased the probability of finding causal, rare, and private mutation within known implicated genes. By design, GWAS are powered only to implicate genes in which common variant alleles are overrepresented in the disease population. It is highly likely that pathogenic coding changes that are either very rare or even private to individuals in other genes have gone undetected by these methods.

Exome sequencing allows capture of extremely large and useful amounts of data. Limitations of this sequencing technique still exist and can have an impact on research data; inefficiencies in the exon targeting process can lead to uneven capture and result in exons with low sequence coverage and off-target hybridizations. Alongside, this unknown or yet-to-be-annotated exons, evolutionary conserved noncoding regions and regulatory sequences (such as enhancers or promoters) involved in IBD and

asthma will not be captured. Exome sequencing is not designed to capture information regarding the methylation state of DNA, and therefore, epigenetic factors in disease are not investigated. Necessary filtering of vast data sets intrinsic to next generation sequencing may lead to missed calling of valid variants.

In this study, we identified the prevalence of concurrent autoimmune diagnoses in a cohort of children with childhood onset IBD. We observe a frequency of asthma and atopic dermatitis at the highest end of the normal range. In children with asthma, we demonstrate that patient-specific mutations in known disease-related genes are extensive and varied, even when restricted to mutations predicted to be pathogenic. Next generation sequencing may be set to become a key routine diagnostic tool of the future, and it is important that we begin to elucidate the role of key genes and pathways already known to us. Improved assessment of true functional significance of mutations will require substantial improvements to in silico annotation informed by rigorous and extensive functional validation of rare variants. However, perfect annotation of single variants in isolation cannot predict outcome in patients who harbor a profile of variants and across genes and pathways. This bottleneck to the interpretation of genomic data may be aided by the assessment of highly selected patient groups.[63] Our study uncovers the patient-specific burden of pathogenic mutations in known disease genes. We find evidence to support causality of key genes such as *ZPBP2* and *PYHIN1* and further postulate that for a subset of patients, the relationship between concurrent PIBD and autoimmune disease lies in systemic immune dysregulation rather than organ-specific immune dysfunction.

ACKNOWLEDGMENTS

The authors are very grateful to all participants and their families. They thank Matthew Smith for helping with the demographic data and reviewing the clinical notes, Liz Blake for assisting pediatric recruitment, Nikki J Graham for technical assistance in DNA laboratory in Human Genetics & Genomic Medicine, University of Southampton, David Buck & Lorna Gregory from the Wellcome Trust Centre for Human Genetics, the NIHR & the Southampton Centre for Biomedical Research (SCBR).

REFERENCES

1. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. *Lancet*. 2007;369:1627–1640.
2. Baumgart DC, Sandborn WJ. Crohn's disease. *Lancet*. 2012;380: 1590–1605.
3. Ashton JJ, Wiskin AE, Ennis S, et al. Rising incidence of paediatric inflammatory bowel disease (PIBD) in Wessex, Southern England. *Arch Dis Child.* 2014;99:659–664.
4. Henderson P, Hansen R, Cameron FL, et al. Rising incidence of pediatric inflammatory bowel disease in Scotland. *Inflamm Bowel Dis.* 2012;18: 999–1005.
5. Hildebrand H, Finkel Y, Grahnquist L, et al. Changing pattern of paediatric inflammatory bowel disease in northern Stockholm 1990-2001. *Gut*. 2003;52:1432–1434.
6. Malmborg P, Grahnquist L, Lindholm J, et al. Increasing incidence of paediatric inflammatory bowel disease in northern Stockholm County, 2002-2007. *J Pediatr Gastroenterol Nutr.* 2013;57:29–34.
7. Gearry RB, Dodgshun AJ. The "hygiene hypothesis" in IBD. *J Crohns Colitis.* 2012;6:869; author reply 870.
8. Aujnarain A, Mack DR, Benchimol EI. The role of the environment in the development of pediatric inflammatory bowel disease. *Curr Gastroenterol Rep.* 2013;15:326.
9. Kilpinen H, Barrett JC. How next-generation sequencing is transforming complex disease genetics. *Trends Genet.* 2013;29:23–30.
10. Lesage S, Zouali H, Cézard JP, et al. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet.* 2002;70:845–857.
11. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*. 2001; 411:599–603.
12. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet.* 2011;43:246–252.
13. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–124.
14. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet.* 2010;42:1118–1125.
15. Rivas MA, Beaudoin M, Gardet A, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet.* 2011;43:1066–1073.
16. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A.* 2012;109:1193–1198.
17. Gupta G, Gelfand JM, Lewis JD. Increased risk for demyelinating diseases in patients with inflammatory bowel disease. *Gastroenterology*. 2005;129:819–826.
18. Weng X, Liu L, Barcellos LF, et al. Clustering of inflammatory bowel disease with immune mediated diseases among members of a northern California-managed care organization. *Am J Gastroenterol.* 2007;102: 1429–1435.
19. Bardella MT, Elli L, De Matteis S, et al. Autoimmune disorders in patients affected by celiac sprue and inflammatory bowel disease. *Ann Med.* 2009; 41:139–143.
20. Kappelman MD, Galanko JA, Porter CQ, Sandler RS. Association of paediatric inflammatory bowel disease with other immune-mediated diseases. *Arch Dis Child.* 2011;96:1042–1046.
21. IBD Working Group of the European Society for Paediatric Gastroenterology, Hepatology and Nutrition. Inflammatory bowel disease in children and adolescents: recommendations for diagnosis—the Porto criteria. *J Pediatr Gastroenterol Nutr.* 2005;41:1–7.
22. Levine A, Griffiths A, Markowitz J, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflamm Bowel Dis.* 2011;17:1314–1321.
23. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16: 1215.
24. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079.
25. Illumina. *Quality Scores for Next-generation Sequencing*. Available at: http://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf. Accessed April 22, 2013.
26. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38:e164.
27. Pengelly RJ, Gibson J, Andreoletti G, et al. A SNP profiling panel for sample tracking in whole-exome sequencing studies. *Genome Med.* 2013; 5:89.
28. Lockett GA, Holloway JW. Genome-wide association studies in asthma; perhaps, the end of the beginning. *Curr Opin Allergy Clin Immunol.* 2013; 13:463–469.
29. Van Limbergen J, Russell RK, Nimmo ER, et al. Filaggrin loss-of-function variants are associated with atopic comorbidity in pediatric inflammatory bowel disease. *Inflamm Bowel Dis.* 2009;15:1492–1498.
30. Andersen ABT, Ehrenstein V, Erichsen R, et al. Parental inflammatory bowel disease and risk of asthma in offspring: a nationwide cohort study in Denmark. *Clin Transl Gastroenterol.* 2013;4:e41.

31. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–2158.
32. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–842.
33. Zhan X, Liu DJ. TaSer (TabAnno and SeqMiner): a toolset for annotating and querying next-generation sequence data.2013. Available at: http://arxiv.org/abs/1306.5715v1. Accessed January 22, 2014.
34. Kang HM, Zhan X, Sim X, et al. Biostatistics Dept, Univ Michigan, Ann Arbor, Ann Arbor, M. EPACTS (Efficient and Parallelizable Association Container Toolbox). Available at: http://genome.sph.umich.edu/wiki/EPACTS. Accessed January 22, 2014.
35. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*. 2010;327:78–81.
36. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29:308–311.
37. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.
38. Li MX, Kwan JS, Bao SY, et al. Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet.* 2013;9:e1003143.
39. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4:44–57.
40. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27–30.
41. Ionita-Laza I, Lee S, Makarov V, et al. Sequence kernel association tests for the combined effect of rare and common variants. *Am J Hum Genet.* 2013;92:841–853.
42. Anderson HR, Gupta R, Strachan DP, Limb ES. 50 years of asthma: UK trends from 1955 to 2004. *Thorax*. 2007;62:85–90.
43. Bernstein CN, Wajda A, Blanchard JF. The clustering of other chronic inflammatory diseases in inflammatory bowel disease: a population-based study. *Gastroenterology*. 2005;129:827–836.
44. Emerson RM, Williams HC, Allen BR. Severity distribution of atopic dermatitis in the community and its relationship to secondary referral. *Br J Dermatol.* 1998;139:73–76.
45. Van Heel DA, West J. Recent advances in coeliac disease. *Gut*. 2006;55:1037–1046.
46. Lindkvist B, Benito de Valle M, Gullberg B, Björnsson E. Incidence and prevalence of primary sclerosing cholangitis in a defined adult population in Sweden. *Hepatology*. 2010;52:571–577.
47. Alkhateeb A, Fain PR, Thody A, et al. Epidemiology of vitiligo and associated autoimmune diseases in Caucasian probands and their families. *Pigment Cell Res.* 2003;16:208–214.
48. Lu L, Wang J, Zhang F, et al. Role of SMAD and non-SMAD signals in the development of Th17 and regulatory T cells. *J Immunol.* 2010;184:4295–4306.
49. Sleiman PMA, Flory J, Imielinski M, et al. Variants of DENND1B associated with asthma in children. *N Engl J Med.* 2010;362:36–44.
50. Rioux JD, Daly MJ, Silverberg MS, et al. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet.* 2001;29:223–228.
51. Lee GR, Fields PE, Griffin TJ, Flavell RA. Regulation of the Th2 cytokine locus by a locus control region. *Immunity*. 2003;19:145–153.
52. Kim SH, Cho BY, Park CS, et al. Alpha-T-catenin (CTNNA3) gene was identified as a risk variant for toluene diisocyanate-induced asthma by genome-wide association analysis. *Clin Exp Allergy* 2009;39:203–212.
53. The asthma epidemic. *N Engl J Med.* Available at: http://www.nejm.org/doi/full/10.1056/NEJMra054308. Accessed December 2, 2014.
54. Asher MI, Montefort S, Björkstén B, et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *Lancet*. 2006;368:733–743.
55. Verlaan DJ, Berlivet S, Hunninghake GM, et al. Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. *Am J Hum Genet.* 2009;85:377–393.
56. Wan YI, Shrine NR, Soler Artigas M, et al. Genome-wide association study to identify genetic determinants of severe asthma. *Thorax*. 2012;67:762–768.
57. Wu H, Romieu I, Shi M, et al. Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans. *J Allergy Clin Immunol.* 2010;125:321–327.e13.
58. Zhu G, Whyte MK, Vestbo J, et al. Interleukin 18 receptor 1 gene polymorphisms are associated with asthma. *Eur J Hum Genet.* 2008;16:1083–1090.
59. Michel S, Liang L, Depner M, et al. Unifying candidate gene and GWAS Approaches in Asthma. *PLoS One.* 2010;5:e13894.
60. Zhang Y, Moffatt MF, Cookson WOC. Genetic and genomic approaches to asthma: new insights for the origins. *Curr Opin Pulm Med.* 2012;18:6–13.
61. PYHIN1 pyrin and HIN domain family, member 1 [Homo sapiens (human)]. Available at: http://www.ncbi.nlm.nih.gov/gene/149628. Accessed January 5, 2015.
62. Torgerson DG, Ampleford EJ, Chiu GY, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet.* 2011;43:887–892.
63. Chisholm J, Caulfield M, Parker M, et al. Briefing genomics England and the 100K genome project. *Genomics Engl.* 2013. Available at: http://www.genomicsengland.co.uk/briefing/. Accessed January 5, 2015.