Genetic Variants Assessing Crohn's Disease Pattern in Pediatric Inflammatory Bowel Disease Patients by a Clinical Exome Survey

Dago Dougba Noel[1,2,3] (iD), Pinelli Marinella[1], Giacomelli Mauro[1], Serena Ilaria Tripodi[1], Alessia Pin[4], Arrigo Serena[5], Bramuzzo Matteo[4], Fuoti Maurizio Giuseppe[6], Alvisi Patrizia[7], Calza Stefano[2], Alberto Tommasini[4] and Badolato Raffaele[1]

[1]Department of Clinical and Experimental Sciences, Institute of Molecular Medicine Angelo Nocivelli, University of Brescia and Children's Hospital, ASST Spedali Civili, Brescia, Italy. [2]Unit of Biostatistics and Biomathematics & Unit of Bioinformatics, Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy. [3]Department of Genetic and Biochemistry, Genetic Research Unit, Peleforo Gon Coulibaly University of Korhogo, Korhogo, Ivory Coast. [4]IRCCS Materno Infantile Burlo Garofolo, Trieste, Italy. [5]Pediatric Hospital Giannina Gaslini, Genoa, Italy. [6]ASST Spedali Civili, Brescia, Italy. [7]Pediatric Gastroenterology Unit, Maggiore Hospital, Bologna, Italy.

ABSTRACT

BACKGROUND: Inflammatory bowel diseases (IBDs) are complex, multifactorial disorders that comprise Crohn's disease (CD) and ulcerative colitis (UC). Recent discoveries have brought much attention to the genetic predisposition of patients with IBD. Here we evaluate the interaction between IBD genetic risk factors susceptibility and CD occurrence in an IBD pediatric patient population, performing a clinical exome survey.

METHODS: From February 2018 to April 2019, we collected blood samples from 7 pediatric patients with IBD concerns from several collaborating health centers and/or hospitals. Blood samples were processed by extracting and sequencing DNA for a clinical exome survey. Shophia-DDM-v3-4 platform allowed sequenced reads alignment on hg19 genome as well as genetic variant calling. Both IBD risk and pathogenic genetic variants covered by at least 20 reads were selected for subjacent analysis.

RESULTS: Normality and Bartlett tests of both risk and pathogenic genetic variants suggested random and heterogeneous distribution of these variants in this group of IBD pediatric patients. P value clustering analysis by processing 157 IBD risk factors revealed genetic heterogeneity in IBD population and suggested two pathways influencing IBD development. In particular, (1) genetic variants associated with autoimmune and (2) metabolic diseases and CD risk factors (rs2066844 and rs2241880 single nucleotide polymorphism variants, respectively, of genes NOD2 and ATG16L) were identified in distinct clusters of IBD patients ($P < .05$). Moreover, the heterogeneous distribution of the following variants rs10065172 (IRGM), rs1805010 (IL4R), rs5030737 (MBL2), and rs33995883 (LRRK2) in this group of IBD patients was consistent with their random distribution in that population.

CONCLUSION: Our study revealed specific genetic variants linked to CD susceptibility, autoimmune and/or innate immunodeficiency as well as to metabolic defects, as favoring factors of IBD, suggesting the valuable role of next generation sequencing (NGS) approaches in IBD molecular diagnostic procedures.

KEYWORDS: Inflammatory bowel disease, Crohn's disease, genetic variants, NGS-next generation sequencing, pediatric patients

Introduction

The inflammatory bowel disease (IBD) is a group of gastrointestinal disorders including Crohn's disease (CD) and ulcerative colitis (UC), which are distinguished by the distribution of chronic inflammatory changes. The inflammation in UC is continuous and restricted to the mucosal layer of the colon, whereas CD is characterized by segmental transmural lesions that can affect any part of the gastrointestinal tract. The gastrointestinal tract is always solicited in regulating the gut microbial flora, controlling immune responses to food chemical substances antigens as well as maintaining homeostasis. It is well known that homeostasis disturbances contribute to risk of occurrence of IBD. Of note, the innate immune response is indispensable for maintaining homeostasis, and abnormal innate immune activity is deeply involved in the pathogenesis of IBD. Research in this field has made substantial advancements in recent years.[1-3] Genome-wide association studies (GWAS) have identified over 200 loci associated with IBD,[4,5] providing many potential therapeutic targets. The large majority of these loci are common to CD and UC, implicating

abundant pathways, notably the pro-inflammatory interleukin (IL)-23 pathway.[6] CD predominant loci include nucleotide-binding oligomerization domain-containing protein 2 (*NOD2*), and a number of autophagy genes (eg, *ATG16L1*, *IRGM*). Numerous studies revealed the central role played by autophagy in human inflammatory disorder by direct elimination of intracellular bacteria and activation of pattern recognition receptor signaling, which is involved in gut homeostasis and CD pathogenesis.[7,8] Furthermore, *NOD2* is an intracellular receptor for bacterial peptidoglycan and is expressed in a wide variety of cells including plasma-cells, macrophages, and other leukocytes which are located at the base of the small intestinal producing potent antimicrobial peptides. Of note, even if several studies justify the crucial role of the gut microbiota in the development of gut inflammation, however, it is still unclear whether a specific individual bacterial species and/or strain might be causative of IBD or only contribute in exacerbation of IBD pathogenesis. Recently, genome-wide association studies displayed the linkage of susceptibility genes with certain bacteria prevalence in IBD. Indeed, these studies have identified more than 160 genetic loci which confer protection from IBD or increased risk to IBD development.[4] Moreover, there is evidence of numerous diverse genetic factors involved in IBD pathophysiology supporting the hypothesis that IBD is a polygenic disease that can develop due to inefficient handling of the gut microbiota. Variation and/or defects detected in these genes can be associated to microbial recognition and killing functions, mucosa barrier integrity, and immune regulation.[9-11] IBD might develop in patients with specific genetic defects of the host in the regulation of commensal bacteria, or regulation of the immune response to these challenges. However, extra-intestinal manifestations affecting nearly every organ are frequent in IBD patients, occurring in 20% to 40% of patients.[12,13] Skin, musculoskeletal system, and eyes are often involved, but the hepatic-pancreatic-biliary, nervous, cardiovascular, renal and respiratory systems may also be affected. Although factors associated with insulin resistance are commonly augmented in IBD patients, low-density lipoprotein (LDL) cholesterol and total cholesterol levels are generally low in that patients typology.[14] These changes are more pronounced in CD patients as compared with UC and they are present independently to disease activity, possibly mediated by cytokine production. Plasma LDL cholesterol levels are lower in IBD patients as compared with healthy individuals. This is probably related to bile acids malabsorption leading to stimulation of cholesterol synthesis, cholesterol degradation, and LDL receptor expression in the liver, which secondarily result in reducing LDL cholesterol.[15] Although the number of studies evaluating the prevalence of metabolic syndrome in patients with IBD is limited, some of them showed that type 2 diabetes mellitus and cardiovascular disease may affect the prognosis in patients with IBD. Indeed, obesity increases the risk of developing colorectal cancer, which is a serious concern for IBD patients.[16,17] Abnormal fat metabolism may also influence disease activity in

IBD patients.[18] A recent study suggests that hypertrophied mesenteric adipose tissue in CD patients contributes to increased disease activity and development of complications; making inflammatory bowel disease (IBD), a complex multifactorial and/or polygenic pathology. There are many hypotheses on the cause of CD ranging from possible association with environmental factors including microorganisms to imbalance in the intestinal normal flora of the patients. Regardless of the environmental trigger, there is strong evidence that a genetic disposition is a major key in acquiring CD. Many studies have proven the link between several single nucleotide polymorphism (SNP) risk factors involved in the IBD CD pathways. Indeed, whole genome association studies, in which SNPs across the genome is genotyped, are novel approach to assess the role of genetic variation in disease. Because CD is multifactorial as well as polygenic pathology, we evaluated the correlation of CD genetic risk factors, identified by clinical exome analysis, to CD phenotype features in a small IBD pediatric patient population.

Materials and Methods

IBD patient population

Seven pediatric patients with diagnosis of IBD were included in the present clinical exome diagnostic analysis. Patients diagnosed with different types of inflammatory bowel disease were enrolled from February 2018 to August 2018; samples were collected in the reference center and shipped to our laboratory for a clinical exome sequencing analysis. The overall research included 4 research institutes (IRCCS) and 8 Italian universities. Here, we reported and summarized general and clinical features regarding processed IBD pediatric patients (Table 1).

DNA library preparation and exome sequencing

Whole blood samples from 7 unrelated patients were collected in EDTA. Genomic DNA (gDNA) was extracted by QIAamp DNA Mini Kit Qiagen following manufacturer's instructions. Genomic DNA was loaded into the spectrophotometer UV/Vis Infinite M200 TECAN to verify purity of the samples. Then, we used 200 ng of DNA for library preparation with SOPHiA GENETICS Clinical Exome Solution. Briefly, gDNA is fragmented and ligated to Dual Index Adapters, and then fragments with the correct size (300-700 bp) are selected using the AMPure XP beads. We assessed libraries quality by using Quant-iT PicoGreen dsDNA Reagent (Life Technologies) consisting of a calibration curve by quantifying serial dilution of standard DNA. For this purpose, we mixed 1 μL of library sample with 99 μL of TE Buffer and 100 μL of PicoGreen working solution (1:200 in TE Buffer) and quantified. Four individual libraries (300 g each) from 4 patients were pooled and lyophilized in the Eppendorf Concentrator 5301 at 45°C, then the pellet is hybridized to the xGen Lockdown Probes for 4 hours; the hybridization product is then bound to streptavidin beads prior to polymerase chain reaction (PCR)

Table 1. Summary of generic and clinical data features of processed inflammatory bowel disease pediatric patients.

	PATIENT A	PATIENT B	PATIENT C	PATIENT D	PATIENT E	PATIENT F	PATIENT G
Age (years)	14	11	7	18	16	11	18
Age at onset (years)	11	8	4	15	14	4	12
Gender	M	F	F	M	M	F	M
IBD type	Crohn's disease	Ulcerative colitis	Crohn's disease	Ulcerative rectocolitis	Crohn's disease	Pancolitis	Gastritis, lympho-monocytic colitis
Other pathologies	–	Recurrent infections	–	Autism	–	–	Diabetes type 1, IgA deficiency, vitiligo, psoriasis, dactylitis

Abbreviation: IBD, inflammatory bowel diseases.

Table 2. Descriptive statistic and quality control analysis of aligned read sequences.

	PATIENT A	PATIENT B	PATIENT C	PATIENT D	PATIENT E	PATIENT F	PATIENT G
Total reads	2011960	2128442	1830550	8826962	12284164	10791312	11846172
Mapped reads (%)	1975967 (98.18%)	2096306 (98.48%)	1787906 (97.67%)	8665542 (98.17%)	12039520 (98.01%)	10584652 (98.08%)	11599547 (97.92%)
Duplicate fraction	1.03%	0.96%	0.96%	3.64%	4.93%	4.17%	4.9%
Selected reads length	300 bp	300 bp	300 bp	300 bp	300 bp	300 bp	300 bp
Low coverage regions	57067	54883	63479	1690	705	723	629
Target regions at 25× coverage	41.49%	45.82%	33.91%	98.63%	99.55%	99.40%	99.55%

amplification with KAPA HiFi HotStart ReadyMix (Roche). The PCR product is cleaned up through binding with AMPure XP magnetic beads and finally resuspended in IDTE buffer to be stored at –20°C. Before sequencing, the sample was quantified with Qubit 4 Fluorometer; the pool molarity was determined, and sample was diluted to reach the concentration of 10 pM. Pair end sequencing was performed on MiSeq (Illumina) platform.

Reads quality control and aligned sequences statistic

In this section, we summarized read quality control item for each processed inflammatory bowel disease (IBD) pediatric patients. Exome sequencing were executed by Miseq machine by using Illumina technology, performing DNA pair end sequencing. Read quality control as well as base calling statistic were automatically generated by sequencer machine. We summarized reads and base calling statistical survey (Table 2).

Clinical genomic data processing by Sophia platform

SOPHiA DDM is the platform of choice for clinicians performing routine diagnostic testing. This platform performs sequence data analysis by including alignment up to calling annotated facilitated variants. So, after login to SOPHiA

DDM platform, we processed in loading raw DNA sequence FastQ file. Loaded raw DNA sequence are automatically and immediately processed and aligned on GRCh37/hg19 reference genome, using patented advanced algorithms and machine learning approaches assuring the highest accuracy of subjacent genomic analysis. After this step, processed data are visualized on SOPHiA DDM platform. SOPHiA AI an algorithm of SOPHiA DDM platform allows processed data interpretation by pre-classifying variants in different pathogenic classes from highly pathogenic to benign. We directly interpret the variants and generate a variant report all that within one tool. Next, we exported variant VCF files in an external environment (R software) for subjacent variant filtering as well as statistical analysis. Risk factor and/or pathogenic clinical variants for IBD were filtered for target regions covered at least 20 reads, as well as by checking SNP variant on online mendelian inheritance in man (OMIM) inerrant status (heterozygous and/or homozygous status) concordance. In addition, we checked and compared for each selected SNP variants, associated biological functions by using several genomic databases (ie, Gene Card and Ensembl).

Statistical analysis

Data (selected IBD risk factor as well as pathogenic genetic variants) were statistically processed in R programming

environment. We checked for IBD SNP risk factor variants heterogeneity as well as normality distribution by applying respectively Bartlett and Shapiro normality tests. Indeed, risk factors under ClinVar parameter as reported in genomic variation as it relates to human health with sequencing depth at least to 20 reads were processed for above-mentioned variance homogeneity as well as normality tests. We analyzed IBD pediatric patient's interaction by achieving Z-score and P value clustering analysis, handling selected IBD risk factors as well as pathogenic SNP variants. A test was statistically significant at $P < .05$.

Results

Assessment of IBD risk factor SNP variants variance homogeneity in processed IBD pediatric patients

Because of IBD multifactorial features, we assessed variance homology with selected IBD-related SNP variants by performing a Bartlett test. In total, 157 IBD genetic variants, under ClinVar parameter as reported in genomic variation as it relates to human health, covered by at least to 20 read sequences were selected for present variance homogeneity test. Of note, Bartlett test assessing homogeneity of risk factor variance in analyzed IBD pediatric patient population revealed statistically significant Bartlett's K-squared = 100.71 at a freedom degree $(df) = 6$ $(P < .05)$. But, high and significant Bartlett's K-squared coefficient suggested variance heterogeneity of IBD-related SNP variants among IBD pediatric patient population. Partial Bartlett's test showed weak and nonsignificant $(P > .05)$ Bartlett's K-squared coefficient (Bartlett's K-squared coefficient: 0.00-0.06) by internal comparison of IBD pediatric patients A, E, F, and G. In addition, partial Bartlett's test internal comparison of IBD pediatric patients B, C, and D showed significant $(P < .05)$ Bartlett's K-squared coefficient (Table 3). Taken together, these results suggested that analysis of IBD pediatric patients showed significant variance heterogeneity. By contrast, Bartlett test supported partial variance homogeneity of IBD-related variants in patients A, E, F, and G, suggesting that this group of IBD subjects shares common susceptibility patterns. In contrast, IBD pediatric patients B, C, and D group constitute a distinct cluster of subjects sharing common genetic variants (Table 3).

Inflammatory bowel disease pediatric patients clustering analysis by processing IBD risk factor SNPs variants

We investigated 91 non-redundant risk factor variants in 7 children with IBD. Indeed, analysis of selected genetic variants by Z-score clustering suggested a high heterogeneity in the analyzed IBD pediatric patient population (Figure 1A). Z-score analysis of genetic variants showed that IBD pediatric patients' heterogeneity behavior clustered in two groups, suggesting a relative resemblance of IBD pediatric patients A,

Table 3. Bartlett variance homogeneity test out came by comparing IBD patient population each to other by processing genetic risk factor variants.

BARTLETT'S TEST PARTIAL CONTRAST	DEGREE OF FREEDOM (*DF*)	BARTLETT'S K-SQUARED	*P*
Patient A vs patient B	1	8.11	0.00
Patient A vs patient C	1	56.64	0.00
Patient A vs patient D	1	22.76	0.00
Patient A vs patient E	**1**	**0.06**	**0.80**[NS]
Patient A vs patient F	**1**	**0.00**	**0.94**[NS]
Patient A vs patient G	**1**	**0.04**	**0.85**[NS]
Patient B vs patient C	1	24.46	0.00
Patient B vs patient D	1	3.98	0.04
Patient B vs patient E	1	9.56	0.00
Patient B vs patient F	1	8.51	0.00
Patient B vs patient G	1	9.23	0.00
Patient C vs patient D	1	9.18	0.00
Patient C vs patient E	1	59.88	0.00
Patient C vs patient F	1	57.56	0.00
Patient C vs patient G	1	59.18	0.00
Patient D vs patient E	1	25.03	0.00
Patient D vs patient F	1	23.40	0.00
Patient D vs patient G	1	24.53	0.00
Patient E vs patient F	**1**	**0.03**	**0.86**[NS]
Patient E vs patient G	**1**	**0.00**	**0.96**[NS]
Patient F vs patient G	**1**	**0.02**	**0.92**[NS]

Abbreviation: IBD, inflammatory bowel diseases.
NS: non sigificant variance difference; $P > 0.05$.

E, F, and G as compared with patients B, C, and D (Figure 1). Furthermore, P values clustering survey confirmed the high heterogeneity between all analyzed IBD pediatric patients, and revealed relatively low variability between patients A and F as compared with the other patients (Figure 1B). Indeed, both patients A and F shared common risk factor variants linked to (1) metabolic disorders such as diabetes mellitus and/or insulin and/or diabetes mellitus type II susceptibility, (2) obesity and metabolic syndrome, or (3) autoimmune disease susceptibility. Interestingly, well-noted IBD risk factors variants (ie, rs2066844 and rs2241880 IBD risk factor variants associated to NOD2 and ATG16L1 genes respectively) were identified in both patients A and F (Table 4). In addition, the same analysis suggested alternate detection of NOD2 (rs2066844) versus ATG16L1 (rs2241880) genetic variants in IBD pediatric patients E and G.
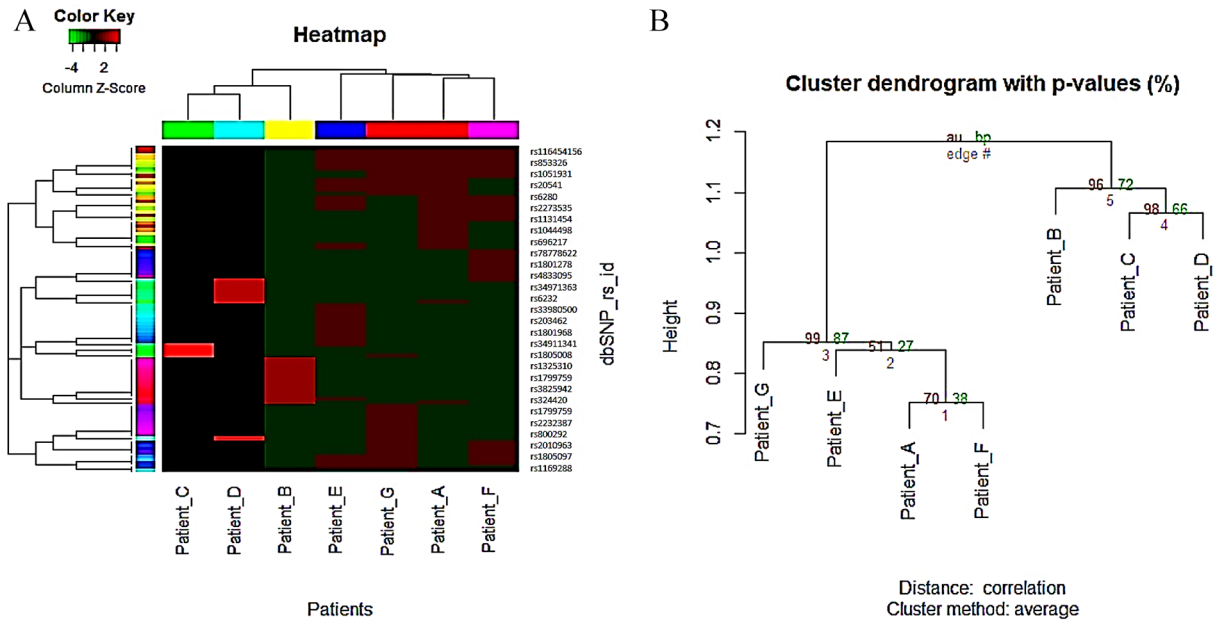
Figure 1. IBD pediatric patient population clustering analysis by processing recurrent risk factor genetic variants (SNPs) by Z-score (A) and *P* value clustering analysis (B).
IBD indicates inflammatory bowel diseases; SNP, single nucleotide polymorphism.

Evaluation of risk factor variant features distribution in the IBD pediatric patient population

Venn diagram analyses of genetic variants in IBD patients revealed high heterogeneity and random distribution of selected risk factor variants in IBD pediatric patient population (Figure 2). Indeed, normality test (Shapiro normality test) by exhibiting Shapiro coefficient; W = 0.54 at *P* < .05, and risk factor variants density plot analysis supported these hypotheses (Figure 2B). These observations showed a relatively higher frequency of risk factor variants involved in autoimmune and metabolic disorders (Supplementary Table 1). Of note, more than 25% of selected risk factor variants was associated with autoimmune and/or metabolic diseases disorder susceptibility. Indeed, rs237025, rs5219, rs13266634, rs1805097, rs854560, and rs1799945 SNP variants revealed (1) insulin resistance susceptibility, (2) diabetes mellitus, and (3) obesity and metabolic syndrome susceptibility in IBD pediatric patients A, E, F, and G, but not in patients B, C, and D. In patients A, E, F, and G, we detected the susceptibility risk factor variants for autoimmune thyroid disease rs180223 and rs853326. In the same patients, we identified the genetic risk factor variant sepsis susceptibility rs497116 (CASP12), which is a homozygous UTR variation. In patients A, E, and F, we identified the homozygous risk factor variant autoimmune rs180223 in TG gene which is linked to thyroid disease. In term of autoimmune susceptibility risk factors, we identified in patients B and F heterozygous risk factor variant rs78778622 in SIAE gene which is linked to autoimmune disease susceptibility, while in patient D we detected in NLRP1 gene the variant rs12150220, which is associated to multiple autoimmune diseases including vitiligo.

However, as previously showed both IBD patients A and F carried the variants rs2066844 (NOD2) and rs2241880 (ATG16L1) which are well known IBD CD susceptibility risk factors (Table 2). In addition, IBD patients C and G carried rs2066844 (NOD2) and rs2241880 (ATG16L1) IBD CD susceptibility risk factor variants, respectively (Figure 2, Table 4). These two patients carried as well a genetic variant rs5744168 in TLR5 resulting in a stop codon which can affect response of TLR5 to flagellin, thereby influencing intestinal microbiota composition. Furthermore, the genetic variants rs231775 (CTLA4), rs1053874 (DNASE1), and rs2070197 (IRF5), which are missense heterozygous genetic variants associated to systemic lupus erythematosus susceptibility and to metabolic disease susceptibility, were shared by patients A, D, E, F, and G. Considering as a whole, patients A, E, F, and G seem to clustered for susceptibility risk factors associated to (1) autoimmune, (2) IBD CDs, as well as (3) metabolic disease disorder, while patients B, C, and D clustered only for autoimmune risk factor variants (*P* < .05). This is in accord with the clustering survey, with the heterogeneity and random selection analyses of SNP variants in IBD patients which were reported above (Figures 1 and 2).

Genetic pathogenic variant in evaluating IBD pediatric patient population diagnosis

Next, we assessed the occurrence of genetic pathogenic variants in children with IBD. The study showed in patient A the synonymous rs10065172 heterozygous (c.313C > T/ p.(Leu105=)) variant in IRGM gene and in patient C the missense homozygous variant rs1805010 (c.223A > G / p.(Ile75Val)) in IL4R gene, which are associated with inflammatory bowel disease

Table 4. List of the top risk factors and/or pathogenic genetic variants recurrently recorded in analyzed inflammatory bowel disease (IBD) pediatric patients cover by at least 20 reads.

REF SNP ID (CLINICAL SIGN)	CHR	COORDINATE	GENE	SUBSTITUTION	IBD PEDIATRIC PATIENTS ID
rs2066844 (risk factor)	16	50745926	NOD2	R765W	3 (patients A, C, and E)
ᵃrs2241880 (risk factor)	2	234183368	ATG16L1	T216A	4 (patients A, E, F, and G)
ᵃrs497116 (risk factor)	11	104763117	CASP12	–	4 (patients A, E, F, and G)
rs5744168 (risk factor)	1	223285200	TLR5	R392*	2 (patients C and G)
rs231775 (risk factor)	2	204732714	CTLA4	T17A	3 (patients A, D, and G)
rs5743289 (pathogenic/risk factor)	16	50756774	NOD2	–	2 (patients E and F)
rs10065172 (pathogenic)	5	150227998	IRGM	L105=	3 (patients A, C, and D)
ᵃrs1805010 (pathogenic, protective)	16	27356203	IL4R	I75V	2 (patients A and C)
rs5030737 (pathogenic)	10	54531242	MBL2	R52C	2 (patients D and F)

Abbreviations: IBD, inflammatory bowel diseases; SNP, single nucleotide polymorphism.
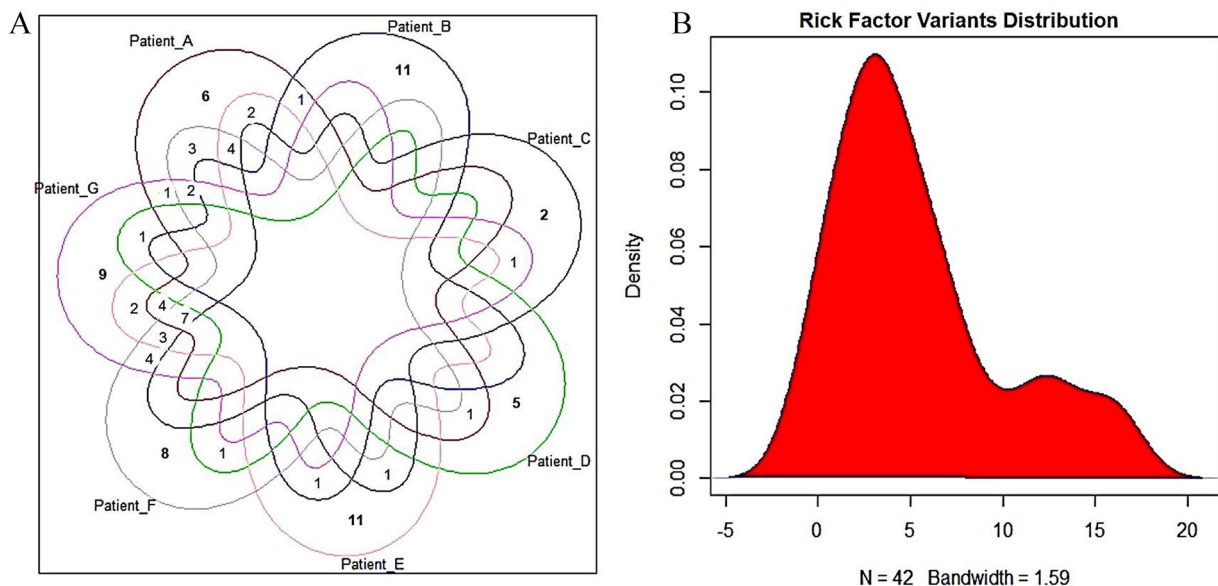ᵃHomozygote variant.

Figure 2. (A) Venn diagram assessing risk factor variants (SNP) data distribution in processed inflammatory bowel disease (IBD) pediatric patients. (B) Density plot measuring risk factor SNP variants normality in processed IBD pediatric population.
IBD indicates inflammatory bowel diseases; SNP, single nucleotide polymorphism.

and immunodeficiency syndrome, respectively (Table 2). Interestingly, in patient D we identified the above-mentioned pathogenic variants (rs10065172 and rs1805010), together with the MBL2-related rs5030737 heterozygous variant (c.154C > T/p.(Arg52Cys)), which is associated with partial mannose-binding protein deficiency. Of note, both patients A and F carried the pathogenic variant rs33995883 (p.(Asn2081Asp)) of LRRK2 gene, which confer increased risk to develop CD. In addition, we identified both the heterozygous variant of IL4R gene (rs1805010 variant) and the mannose binding protein deficiency variant (rs5030737) in patients F and G. In patients E and F, we identified an intronic NOD2

variant (rs5743289: c.2717 + 158C > T), although with low coverage average, which was linked to both Yao syndrome and CD. Besides gene variants associated with immune response, we have also identified in children with IBD pathogenic variants associated with metabolic disorders. In patients D and F, we identified several metabolism disorder disease pathogenic genetic variant such as, rs696217 (c.211C > A/p.(Leu71Met)) in GHRL gene, rs3733402 (c.428G > A/p.(Ser143Asn)) in KLKB1 gene and rs351855 in FGFR4 gene. All of them are involved in metabolic syndrome, obesity age at onset, body fat distribution, cancer progression and tumor cell motility regulation. In addition, genetic variants in PQBP1 (rs606231195:

INDEL mutation resulting in frameshift) and in SLC6A20 gene (rs17279437) causing metabolic dysfunction were identified in patient F (Supplementary Tables 1 and 2).

Discussion

Our study suggests that risk factor variants associated with metabolic disorders, autoimmune or CD susceptibility can be linked to distinct clusters of IBD patients. Numerous studies have shown that multiple factors, including genetic susceptibility, microbiota, and immune system functions can influence IBD pathogenesis.[19-22] Our study revealed rs2066844 (NOD2) and rs2241880 (ATG16L1) IBD susceptibility risk factor variants in A and F patients. In addition, our study revealed a CTLA4 variant in patient A with IBD. These results are in agreement with previous articles which reported the link between genetic variants in ATG16L, NOD2, CARD15, IBD5, CTLA4, TNFSF15, IL2R genes, and CD.[23-27] Indeed, CTLA4 gene is a member of the immunoglobulin superfamily which is expressed on the surface of helper T cells and is essential in the control process of intestinal inflammation by CD25+/CD4+ regulatory T cells.[28] We have also observed that the genetic variant of ATG16L1 gene (rs2241880) was heterozygous in 3 patients while homozygous in an additional one, which is associated with highly increased risk for CD. Some of the genetic variants identified in patients with IBD are related to common biological pathways. In particular, the IBD-associated variants in NOD2, ATG16L1, and IRGM genes affect cellular autophagy processes and bacterial clearance in innate immune cells, and may influence bacterial composition of the gut in patients with IBD. Analysis of the functional effects of the exonic synonymous rs10065172 variant in *IRGM* gene showed that the mutation affected the binding site for a specific microRNAs (miR-196) by causing deregulation of IRGM-dependent xenophagy of bacteria in CD patients,[29] suggesting that rs10065172 in IRGM gene can influence the risk to develop IBD. Moreover, the rs2241880 variant in the ATG16L1 gene, which leads to amino acid substitution of the polar residue threonine with a nonpolar amino acid such as alanine (Thr300Ala), has been identified as a CD risk allele.[7,30] Moreover, as autophagy is an innate defense mechanism that acts as a cell-autonomous process to eliminate intracellular pathogens, a number of studies have linked the CD-associated mutation Thr300Ala in ATG16L1 to autophagy activity in handling and clearing intracellular microorganisms.[31] In addition, Kuballa et al[32] have shown CD-associated ATG16L1 gene rs2241880 variant is associated with impaired efficiency of autophagy-mediated clearance of the intracellular enteric pathogen *Salmonella typhimurium* in human epithelial cells. Of note, we cannot hypothesized an interaction between, ATG16L1 rs2241880 variant and the IRGM exonic synonymous SNP rs10065172 variant in IBD pediatric patient A.[33] However, Glas et al[33] confirmed *IRGM* as susceptibility gene for CD in the German population, highlighting strongest CD susceptibility association signals for rs13361189 (20-kb deletion polymorphism upstream of *IRGM*) and the exonic synonymous SNP rs10065172=p.Leu105Leu, supporting that these two SNPs may be the causal variants. Therefore, our study has highlighted the potential role of the autophagy pathway in CD onset with the identification of risk polymorphisms in various autophagy-related genes and those, which regulate autophagy. In addition to NOD2, ATG16L1, and IRGM genes, other autophagy-related genes were identified as CD susceptibility genes, encoding upstream regulators of autophagy pathways PTPN2[34,35] and LRRK2,[36] as well as for proteins essential for autophagy initiation such as ULK1[37] or proteins involved in bacterial sensing and their subsequent uptake by specific selective autophagy, Toll-like receptor 4 (TLR4).[38] Interestingly, our study reported LRRK2 gene rs33995883 and ATG16L1 gene rs2241880 variants in two IBD patients (A and F) suggesting an interaction between these CD risk factor variants, which is in accord with the hypothesized defect in autophagy function in IBD patients.[39,40] Other studies had also shown that carrying the N2081D allele (LRRK2 gene rs33995883 variant) was significantly associated with a younger age of disease onset.[40] Variant alleles of genes which are known to put individuals at risk for CD can often be linked to other forms of IBD including pancolitis.[41] This can account for the high genetic concordance between IBD pediatric patients A and F as suggested by Z-score and *P* value clustering analysis. NOD2 was the first locus identified as a risk factor for CD,[20,42] which is a member of the NLR (NOD-like receptors) family of intracellular sensors of pathogen/microbe-associated molecular patterns playing crucial roles in regulation of innate immunity.[43] NOD2 gene is highly associated with CD,[44] as CD-associated NOD2 polymorphisms exhibit a reduced capacity to activate NF-κB following muramyl dipeptide stimulation, suggesting that impairment of NOD2 activation promotes CD.[45] NOD2 polymorphisms, rs2066844 (R702W), rs2066845 (G908R), and rs2066847 (l1007fs), are the most common genetic variants which are associated with an increased risk of CD.[46] Indeed, we observed that patients C and E are carriers of rs2066844 (R702W) NOD2 polymorphism.[46] We have also detected in two IBD patients a pathogenic genetic variant of TLR5-(rs2241880) which results in a stop codon associated with reduced response to flagellin and possibly changes in intestinal microbiota composition.[47] Hawn et al[48] showed that TLR5-stop mutation (rs2241880) (Table 4 and Supplementary Table 1) increases human susceptibility to infection through an unusual dominant mechanism that compromises TLR5's essential role as a regulator of the lung epithelial innate immune response. Several studies have shown that mannose-binding lectin is another critical protein with a major role in the innate immune defense against pathogens and involved in intestinal homeostasis.[49] Original studies by Choteau et al[50] have shown that the MBL2 variant

rs5030737 is associated with a low level of mannose-binding lectin and impaired mannose binding lectin mannose associated serine protease functional activity in CD patients. Interestingly, 4 IBD patients out of 7 carried at least one genetic variant in MBL2 gene. Among these SNP variants, two were previously reported as pathogenic variant in two IBD patients. In this study, there is evidence that the NOD2 variant rs2066844, which is associated with susceptibility to CD, results in impairment of mannose binding lectin and impaired mannose binding lectin mannose associated serine protease functional activity.[51]

In conclusion, we report a correlation of IBD CD occurrence and autoimmune with metabolic disorder susceptibility in a small pediatric IBD patient population by performing a clinical exome analysis. Our observations suggest that both (1) autoimmune and/or innate autoimmune autophagy vulnerability and (2) metabolic disorders can influence the risk to develop IBD in children. Despite the limited number of patients analyzed, our study promotes the use of next generation sequencing as a reasonable and reliable tool to identify genetic risk factors which can help IBD diagnosis in children.

Author Contributions

Prof. BR set up the experiment. Dr DDN wrote the paper. Prof. BR revised the manuscript. Dr MP prepared the DNA library for the exome sequencing. Dr DDN performed bioinformatics and biostatistical analysis. Dr GM give a contribution interpreting genetic and genomic results. Dr SIT provided an assistant in understanding patients clinical features. All co-authors read and approved final version of the paper.

ORCID iD

Dago Dougba Noel https://orcid.org/0000-0003-1517-054X

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Ebbo M, Crinier A, Vély F, Vivier E. Innate lymphoid cells: major players in inflammatory diseases. *Nat Rev Immunol*. 2017;17:665-678. doi:10.1038/nri.2017.86.
2. Choy MC, Visvanathan K, De Cruz P. An overview of the innate and adaptive immune system in inflammatory bowel disease. *Inflamm Bowel Dis*. 2017;23:2-13. doi:10.1097/MIB.0000000000000955.
3. Lee SH, Kwon JE, Cho ML. Immunological pathogenesis of inflammatory bowel disease. *Intest Res*. 2018;16:26-42. doi:10.5217/ir.2018.16.1.26.
4. 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. doi:10.1038/nature11582.
5. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47:979-986. doi:10.1038/ng.3359.
6. Saleh AN, Melissa A, Anam K, et al. Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. *World J Gastroenterol*. 2012;18:412-424. doi:10.3748/wjg.v18.i5.412.
7. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet*. 2007;39:596-604.
8. Singh SB, Davis AS, Taylor GA, Deretic V. Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science*. 2006;313:1438-1441.
9. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006;314:1461-1463. doi:10.1126/science.1135245.
10. Fisher S, Tremelling M, Anderson CA, et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat Genet*. 2008;40:710-712. doi:10.1038/ng.145.
11. Huttenhower C, Kostic AD, Xavier RJ. Inflammatory bowel disease as a model for translating the microbiome. *Immunity*. 2014;40:843-854. doi:10.1016/j.immuni.2014.05.013.
12. Bernstein CN, Blanchard JF, Rawsthorne P, Yu N. The prevalence of extraintestinal diseases in inflammatory bowel disease: a population-based study. *Am J Gastroenterol*. 2001;96:1116-1122.
13. Levine JS, Burakoff R. Extra-intestinal manifestations of inflammatory bowel disease. *Gastroenterol Hepatol (NY)*. 2011;7:235-241.
14. Agouridis AP, Elisaf M, Milionis HJ. An overview of lipid abnormalities in patients with inflammatory bowel disease. *Ann Gastroenterol*. 2011;24:181-187.
15. Akerlund JE, Reihnér E, Angelin B, et al. Hepatic metabolism of cholesterol in Crohn's disease. Effect of partial resection of ileum. *Gastroenterology*. 1991;100:1046-1053. doi:10.1016/0016-5085(91)90281-o.
16. Gillen CD, Walmsley RS, Prior P, Andrews HA, Allan RN. Ulcerative colitis and Crohn's disease: a comparison of the colorectal cancer risk in extensive colitis. *Gut*. 1994;35:1590-1592.
17. Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk for colon cancer and adenoma in men. *Ann Intern Med*. 1995;122:327-334.
18. Hass DJ, Brensinger CM, Lewis JD, Lichtenstein GR. The impact of increased body mass index on the clinical course of Crohn's disease. *Clin Gastroenterol Hepatol*. 2006;4:482-488.
19. Darfeuille-Michaud A, Neut C, Barnich N, et al. Presence of adherent *Escherichia coli* strains in ileal mucosa of patients with Crohn's disease. *Gastroenterology*. 1998;115:1405-1413. doi:10.1016/s0016-5085(98)70019-8.
20. Hugot JP, Chamaillard M, Zaouli H, Lesage S, Cézard JP, Belaiche J. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*. 2001;411:599-603.
21. Hart AL, Al-Hassi HO, Rigby RJ, et al. Characteristics of intestinal dendritic cells in inflammatory bowel diseases. *Gastroenterology*. 2005;129:50-65.
22. Wagner J, Sim WH, Ellis JA, et al. Interaction of Crohn's disease susceptibility gene in an Australian paediatric cohort. *PLoS ONE*. 2010;5:e1537. doi:10.1371/journal.pone.0015376.
23. 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. doi:10.1086/339432.
24. Xia B, Crusius JB, Wu J, Zwiers A, van Bodegraven AA, Peña AS. CTLA4 gene polymorphisms in Dutch and Chinese patients with inflammatory bowel disease. *Scand J Gastroenterol*. 2002;37:1296-1300.
25. Brand S, Staudinger T, Schnitzler F, Pfennig S, et al. The role of Toll-like receptor 4 Asp299Gly and Thr399Ile polymorphisms and CARD15/NOD2 mutations in the susceptibility and phenotype of Crohn's disease. *Inflamm Bowel Dis*. 2005;11:645-652.
26. Cummings JR, Ahmad T, Geremia A, et al. Contribution of the novel inflammatory bowel disease gene IL23R to disease susceptibility and phenotype. *Inflamm Bowel Dis*. 2007;13:1063-1068.
27. Cho JH. Inflammatory bowel disease: genetic and epidemiologic considerations. *World J Gastroenterol*. 2008;14:338-347.
28. Read S, Malmström V, Powrie F. Cytotoxic T lymphocyte associated antigen 4 plays an essential role in the function of CD25 (+) CD4 (+) regulatory cells that control intestinal inflammation. *J Exp Med*. 2000;192:295-302.
29. Brest P, Lapaquette P, Souidi M, et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet*. 2011;43:242-245.
30. Hampe J, Franke A, Rosenstiel P, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet*. 2007;39:207-211.
31. Cadwell K, Liu JY, Brown SL, et al. A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. *Nature*. 2008;456:259-263.
32. Kuballa P, Huett A, Rioux JD, Daly MJ, Xavier RJ. Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated ATG16L1 variant. *PLoS ONE*. 2008;3:e3391. doi:10.1371/journal.pone.0003391.
33. Glas J, Seiderer J, Bues S, et al. IRGM variants and susceptibility to inflammatory bowel disease in the German population. *PLoS ONE*. 2013;8:e54338. doi:10.1371/journal.pone.0054338.
34. Glas J, Wagner J, Seiderer J, et al. PTPN2 gene variants are associated with susceptibility to both Crohn's disease and ulcerative colitis supporting a common genetic disease background. *PLoS ONE*. 2012;7:e33682. doi:10.1371/journal.pone.0033682.

35. Scharl M, Mwinyi J, Fischbeck A, et al. Crohn's disease-associated polymorphism within the PTPN2 gene affects muramyl-dipeptide-induced cytokine secretion and autophagy. *Inflamm Bowel Dis*. 2012;18:900-912.

36. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008;40:955-962.

37. Henckaerts L, Cleynen I, Brinar M, et al. Genetic variation in the autophagy gene ULK1 and risk of Crohn's disease. *Inflamm Bowel Dis*. 2011;17:1392-1397.

38. Franchimont D, Vermeire S, El Housni H, et al. Deficient host-bacteria interactions in inflammatory bowel disease? The toll-like receptor (TLR)-4 Asp299gly polymorphism is associated with Crohn's disease and ulcerative colitis. *Gut*. 2004;53:987-992.

39. Michail S, Bultron G, DePaolo RW. Genetic variants associated with Crohn's disease. *Appl Clin Genet*. 2013;6:25-32.

40. Ken YH, Heriberto FH, Jianzhong H, et al. Functional variants in the LRRK2 gene confer shared effect on risk for Crohn's disease and Parkinson's disease. *Sci Transl. Med*. 2018;10:1-31.

41. Franke A, Balschun T, Sina C, Ellinghaus D, et al. Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL). *Nat Genet*. 2010;42:292-294. doi:10.1038/ng.553.

42. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*. 2001;411:603-606.

43. Strober W, Watanabe T. NOD2, an intracellular innate immune sensor involved in host defense and Crohn's disease. *Mucosal Immunol*. 2011;4:484-495.

44. Kullberg BJ, Ferwerda G, de Jong DJ, et al. Crohn's disease patients homozygous for the 3020insC NOD2 mutation have a defective NOD2/TLR4 cross-tolerance to intestinal stimuli. *Immunology*. 2008;123:600-605.

45. Lapaquette P, Bringer MA, Darfeuille-Michaud A. Defects in autophagy favour adherent-invasive *Escherichia coli* persistence within macrophages leading to increased pro-inflammatory response. *Cell Microbiol*. 2012;14:791-807. doi:10.1111/j.1462-5822.2012.01768.x.

46. Hoefkens E, Nys K, John JM, et al. Genetic association and functional role of Crohn disease risk alleles involved in microbial sensing, autophagy, and endoplasmic reticulum (ER) stress. *Autophagy*. 2013;9:2046-2055.

47. Vijay-Kumar M, Aitken JD, Carvalho FA, et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science*. 2010;328:228-231.

48. Hawn TR, Verbon A, Lettinga KD, et al. A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to Legionnaires disease. *J Exp Med*. 2003;198:1563-1572. doi:10.1084/jem.20031220.

49. Seibold F, Konrad A, Flogerzi B, et al. Genetic variants of the mannan-binding lectin are associated with immune reactivity to mannans in Crohn's disease. *Gastroenterology*. 2004;127:1076-1084.

50. Choteau L, Parny M, François N, et al. Role of mannose-binding lectin in intestinal homeostasis and fungal elimination. *Mucosal Immunol*. 2016;9:767-776. doi:10.1038/mi.2015.100.

51. Choteau L, Vasseur F, Lepretre F, et al. Polymorphisms in the mannose-binding lectin gene are associated with defective mannose-binding lectin functional activity in Crohn's disease patients. *Sci Rep*. 2016;6:29636. doi:10.1038/srep29636.