ORIGINAL ARTICLE

Gut and Liver, Vol. 9, No. 6, November 2015, pp. 767-775

Is Whole Exome Sequencing Clinically Practical in the Management of Pediatric Crohn's Disease?

Seak Hee Oh*, Jiwon Baek[†], Kyung Mo Kim*, Eun-Ju Lee[†], Yusun Jung[†], Yeoun Joo Lee*, Hyun-Seung Jin[‡], Byong Duk Ye[§], Suk-Kyun Yang[§], Jong-Keuk Lee[‖], Eul-Ju Seo[¶], Hyun Taek Lim[#], Inchul Lee**, and Kyuyoung Song[†]

**Department of Pediatrics, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, [†]Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, [‡]Department of Pediatrics, GangNeung Asan Hospital, Gangneung, [§]Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, [‖]Asan Institute for Life Sciences, University of Ulsan College of Medicine, Departments of [¶]Laboratory Medicine, [#]Ophthalmology, and **Pathology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea*

Background/Aims: The aim of this study was to identify the profile of rare variants associated with Crohn's disease (CD) using whole exome sequencing (WES) analysis of Korean children with CD and to evaluate whether genetic profiles could provide information during medical decision making. **Methods:** DNA samples from 18 control individuals and 22 patients with infantile, very-early and early onset CD of severe phenotype were used for WES. Genes were filtered using panels of inflammatory bowel disease (IBD)-associated genes and genes of primary immunodeficiency (PID) and monogenic IBD. **Results:** Eighty-one IBD-associated variants and 35 variants in PID genes were revealed by WES. The most frequently occurring variants were carried by nine (41%) and four (18.2%) CD probands and were *ATG16L2* (rs11235604) and *IL17REL* (rs142430606), respectively. Twenty-four IBD-associated variants and 10 PID variants were predicted to be deleterious and were identified in the heterozygous state. However, their functions were unknown with the exception of a novel p.Q111X variant in XIAP (X chromosome) of a male proband. **Conclusions:** The presence of many rare variants of unknown significance limits the clinical applicability of WES for individual CD patients. However, WES in children may be beneficial for distinguishing CD secondary to PID. **(Gut Liver 2015;9:767-775)**

Key Words: Exome sequencing; Crohn disease; Rare variant; Primary immunodeficiency; Child

INTRODUCTION

Recent advances in genome-wide association studies (GWAS) and meta-analyses have identified 140 susceptibility loci for Crohn's disease (CD), an intestinal chronic inflammatory disease, in Caucasians;[1-4] however, the currently identified loci explain less than 30% of the heritable risk and account for relatively small increments in the risk of inflammatory bowel disease (IBD). Existing GWAS have focused on common variants (minor allele frequency [MAF] >0.05), so strategies to enhance the identification of rare (MAF <0.01) and low-frequency (MAF, 0.01 to 0.05) variants with increasing effect sizes are critical for the discovery of the remaining inherited factors.[5] Direct genotyping by targeted array, metabochip, immunochip using low-frequency variants, and genome sequencing are the methods currently available to investigate disease-causing rare variants linked to complex traits.[6] Genome seuquencing technologies have developed rapidly in recent years and this strategy can be used for a wide range of investigations, from monogenic Mendelian disorders to diseases with high degrees of genetic heterogeneity.

The human exome constitutes less than 5% of the genome, and whole exome sequencing (WES) studies can therefore be more cost effective than whole genome sequencing for focused research. In addition, protein-coding regions are more evolutionarily conserved and are more sensitive to genetic changes[7,8] than nongenetic regions, making WES potentially more valuable for uncovering deleterious mutations. WES has been recently employed to circumvent the "diagnostic odyssey" by providing genetic diagnoses for hearing loss, muscular dystrophy, neuromuscular disease, retinitis pigmentosa, and mitochondrial

Correspondence to: Kyuyoung Song
Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea
Tel: +82-2-3010-4277, Fax: +82-2-3010-8029, E-mail: kysong@amc.seoul.kr
Received on April 13, 2015. Revised on July 8, 2015. Accepted on July 20, 2015.
pISSN 1976-2283 eISSN 2005-1212 http://dx.doi.org/10.5009/gnl15176
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

disease. Mitochondrial disease was particularly notable because it was associated not only with mitochondrial genes, but also with hundreds of nuclear DNA genes.[9] Recently, a variety of primary immunodeficiencies (PIDs) and monogenic diseases were revealed to cause refractory infantile colitis.[10,11] Therefore, WES is rapidly becoming a common clinical test for individuals with rare genetic disorders.[12,13]

Despite these advances, the ability of WES models to uncover disease-causing variants associated with complex conditions, such as CD and type 2 diabetes, has not been established for all populations.[14,15] Methods such as GWAS have been used to validate whether identified high-effect variants are common enough to be carried by large populations with CD. Rare and low-frequency variants may occur too infrequently to be identified as contributory for complex traits. In addition, genotypical and phenotypical differences exist between Caucasian and Asian populations with CD. For example, mutations within the nucleotide-binding-oligomerisation-domain (*NOD2/CARD15*) and autophagy-related 16-like 1 (*ATG16L1*) sequences were not associated with CD in Asian populations.[16-18] In addition, the prevalence of small bowel involvement and perianal fistula was higher in Asian patients than in Caucasian patients.[19,20]

Herein, we used WES analysis of Korean children with CD with the aim of identifying rare variants associated with CD. Genetic susceptibility plays a more important role in the etiology of pediatric CD than adult CD, probably as a consequence of a higher burden of disease-causing mutations in affected children.[21] We therefore focused on patients with early-onset CD and severe symptoms such as more extensive disease at onset and rapid progression. In addition, we also asked whether genetic profiling of variants could assist in the medical decision-making process to determine optimal treatment of pediatric CD.

MATERIALS AND METHODS

1. Study population

Twenty-two early-onset CD cases were diagnosed at the IBD Clinic of the Seoul Asan Medical Center. The basic characteristics and clinical phenotypes of the study subjects are summarized in Table 1. Among 230 CD children <14 years of age, youngest children with severe phenotype were included. The severe phenotype was defined as Pediatric Crohn's Disease Activity Index scores were >30 and simple endoscopic scores for CD were >20 at the time of diagnosis. Independent DNA samples from unrelated individuals were collected and sequenced for use as reference exomes to allow evaluation of the burden of mutation in patient samples. Reference exomes were from controls with no history of gastrointestinal or autoimmune disease. Informed consent was obtained from the parents of all the patients and the study was approved by the local ethics committees.

2. Whole exome sequencing

The WES analysis pipeline involved quality checks, alignments, and annotation to identify nucleotides that differed between the patient and reference sequences.[22] Exome capture was performed using the Sure Select Human All Exon 38Mb kit (Agilent Technologies, Santa Clara, CA, USA). The captured, purified and amplified exome-targeting library from each patient was sequenced using an Illumina HiSeq2000 platform. Capture and sequencing were performed by Macrogen, Seoul, Korea. Paired-end sequences produced by HiSeq2000 were mapped to the University of California Santa Cruz human genome assembly hg19 (NCBI build 37.1) using the mapping program BWA (version 0.5.9rc1, http://bio-bwa.sourceforge.net). Picard tools (version 1.59, http://picard.sourceforge.net/) was used for removing PCR duplicates, SAMtools (version 0.1.18, http://samtools.sourceforge.net) was used for the creation of reads uniquely mapped to the reference genome, and BED tools (version 2.15.0, http://bedtools.readthedocs.org) was used to filter out reads that did not map to the targeted exonic regions. Variants were subsequently annotated using ANNOVAR (ver. November 2011, http://www.openbioinformatics.org/annovar/),[23] from file conversion to its input format, filtering with database of single nucleotide polymorphisms (dbSNPs) for the version of 135, and SNPs from the 1000 genomes (1000G) project (http://genome.ucsc.edu/cgi-bin/hgLiftOver). Candidate mutations were selected as those coding nonsynonymous, stop, and insertion/deletion (indels) variants that were present at an allele frequency of <0.05 in the 1000G project database.

3. Panels of genes associated with IBD and PID

To prioritize rare IBD-associated variants of potentially high impact, a comprehensive panel of known IBD-associated genes was selected from previous GWAS IBD data[1-4] and from our GWAS database for Korean-specific susceptibility genes (*ATG16L2*, *DUSP5*, and *TBC1D1*).[24,25] The comprehensive list contained 267 IBD genes, and was used for cross-referencing with exome data from patients and controls (Supplementary Table 1). One-fourth of very young IBD or IBD-mimic colitis cases are related to loss-of-function mutations in critical immune genes, and variants in known PID genes were therefore analyzed to exclude PID in patients with CD.[26] A comprehensive panel of 236 PID genes was assembled according to the 2014 report from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency and Monogenic IBD genes (Supplementary Table 2).[10,11] The specific variants in PID and monogenic IBD gene panels were classified according to the OMIM database (http://omim.org).

4. Prediction of potential functionality

Three *in silico* prediction algorithms were used to predict the effect of each amino acid change on protein function.[27-29] SIFT

Table 1. Characteristics of Children with Crohn's Disease

Proband no.	Age at diagnosis, yr	Symptom onset, yr	Sex	Family history	Paris classification				Evolution of disease phenotype	Biologics
					L	B	P	G		
Proband 1	9	6	F	No	L3→L2	B1→B2B3	P0	G1	Refractory colitis, colectomy	IFX→HMR
Proband 2	9.3	8.7	M	No	L3	B1	P0	G0	-	-
Proband 3	15.8	12	M	No	L1→L4b	B1→B2B3	P1	G1	Stricture, perforation	-
Proband 4	2.8	2.3	F	No	L3	B1→B3	P0	G0	Rectovaginal fistula	-
Proband 5	0.5	0.3	F	No	L2	B1→B3	P1	G1	Rectovaginal fistula, colostomy, severe colitis, death	IFX
Proband 6	10.8	9.4	M	No	L3→L2	B1→B2B3	P0	G1	Severe colitis, ileostomy	IFX→IFX+MTX
Proband 7	13	9.4	M	No	L3→L4b	B1→B3	P0	G0	Recurrent surgeries	IFX
Proband 8	9.9	9.6	F	No	L3	B1	P1	G0	-	IFX→HMR+MTX
Proband 9	1.3	0.8	F	No	L2→L3	B1→B2B3	P0	G1	Refractory colitis, ileostomy, death	IFX→HMR
Proband 10	14.9	14.2	M	Yes	L3	B1	P1	G0	Repeated perianal disease	IFX
Proband 11	11	8.5	M	No	L3	B1	P1	G0	Refractory colitis	IFX+AZA→HMR+MTX
Proband 12	6.1	5.4	M	No	L1	B1	P1	G0	-	-
Proband 13	11	8	M	No	L2→L3	B1→B3	P0	G0	Refractory colitis, colectomy	IFX→HMR
Proband 14	13.1	12.8	F	No	L3	B1→B2B3	P1	G0	Severe colitis	IFX→HMR
Proband 15	12.6	11.9	F	Yes	L3	B3→B1	P0	G0	Severe colitis, ileostomy	IFX→HMR
Proband 16	11.1	10.8	F	Yes	L1	B1	P1	G0	Perianal abscess	IFX→IFX+MTX
Proband 17	9.3	8.8	F	No	L1	B1	P0	G0	-	-
Proband 18	10	9.5	M	No	L3	B1	P1	G0	Repeated perianal disease	IFX
Proband 19	10.3	9.7	M	No	L3	B1	P1	G0	-	IFX→IFX+AZA
Proband 20	10.8	9.2	M	No	L3	B1	P0	G0	-	IFX
Proband 21	11.8	11.1	M	No	L4b	B1	P0	G0	-	IFX
Proband 22	9	8	M	No	L3→L4b	B1→B2	P0	G1	Repeated stricture and surgeries	-

L, location; B, behavior; P, perianal disease; G, growth; F, female; IFX, infliximab; HMR, adalimumab; M, male; MTX, methotrexate; AZA, azathiopurine.

(sorts intolerant from tolerant substitutions) and/or PolyPhen2 (polymorphism phenotyping 2) ratings of "deleterious" indicated a predicted disease-causing effect. A Mutation-Taster prediction of "disease-causing" was subsequently used as a more detailed pathogenicity score. A higher PhyloP score was indicative of higher levels of evolutionary conservation.[30] In this study, highly conserved loci (PhyloP score ≥1.5) were considered to have potentially deleterious mutations if a rating of "deleterious" or "disease-causing" resulted from at least one of the three *in silico* prediction models.

RESULTS

1. Exome sequencing

Exome data were analyzed from 22 pediatric patients with CD and 18 reference individuals. Total read average was 78,473,095 bp. Seventy-eight percent of mappable reads were on-target reads and 86% of targeted bases were covered at 10× read depth. Each exome had, on average, 66,289 SNPs, with 20,196 found in exonic regions. Following a series of quality-control steps (SNP quality >50, total read depth >10, alternative read depth >3), 171,898 variants were identified across the 40 exomes. Of those, we focused on 32,794 missense/nonsense/indel variants within exons. After 24,317 of these variants were removed due to their presence in the 18 control exomes, 8,477 unique variants from 5,625 genes were identified across 22 CD exomes.

2. Characteristics of coding variants in IBD-associated genes

Of the 8,477 unique variants from 5,625 genes, the 22 probands carried 81 rare and low-frequency variants, of which MAF were less than 0.05 among 56 IBD-associated genes (Supplementary Table 3). Two probands each carried nonsense mutations in *ATG16L1* and *NOD2*; however, these were not deleterious and were not highly conserved according to *in silico* prediction algorithms. With the exception of *ATG16L2* (rs11235604) and *TBC1D1* (rs117452860),[24] the remaining variants were of unknown significance (VUS), and their functional roles in mucosal immunity remain to be elucidated.

Among the 81 variants of the 56 IBD-associated genes, the most frequently occurring variants were carried by nine and four CD probands, and were found in *ATG16L2* (rs11235604) and *IL17REL* (rs142430606), respectively (Table 2). *ATG16L2*, a homolog of *ATG16L1*, was identified as a novel candidate gene for CD in a recent Korean GWAS.[24] *ATG16L1* functions in autophagy alongside *ATG5*.[31] In addition, *ATG16L1* is closely related to *NOD2*, which functions in an autophagy-mediated antibacterial pathway in CD.[32] However, little is known regarding the function of mutated *ATG16L2*. An additional SNP in *IL17REL*, rs142430606 (c.C785T; p.P262L), has not previously been associated with IBD and was not predicted to be deleteri-

Table 2. Rare Variants in Inflammatory Bowel Disease-Associated Genes according to Frequency

No. of case	SNP									*In silico* prediction			
	Gene	Chromosome	Exon	Variant type	Base pair position in hg18	dbSNP135	Nucleotide change	Amino acid change	UCSC frequency >1%	1000G 2011Oct allele frequency	SIFT	Polyphen-2	Mutation Taster	PhyloP
9	*ATG16L2*	11	6	Ns	72,533,536	rs11235604	C658T	R220W	0	0.03	D	B	N	1.38
4	*IL17REL*	22	11	Ns	50,436,488	rs142430606	C785T	P262L	.	0.01	T	B	N	-0.21
3	*HLA-DRB5*	6	2	Ns	32,489,856	rs77853982	G196A	D66N	0	.	T	.	P	0.61
2	*IL10RA*	11	3	Ns	117,860,269	.	C301T	R101W	.	.	D	D	D	2.46
2	*IL10RA*	11	6	Ns	117,866,312	rs41354146	G697A	V233M	.	0.01	T	D	N	0.67
2	*TBC1D1*	4	11	Ns	38,053,599	rs117452860	C1990T	P664S	.	0.01	T	P	D	2.66
2	*IFIH1*	2	10	Ns	163,134,021	.	G1948A	D650N	.	.	T	B	N	0.72
2	*IL31RA*	5	13	Ns	55,204,174	rs140524514	G1379A	S460N	.	0	T	B	N	-0.23
2	*MANBA*	4	16	Ns	103,556,114	rs142248415	T2246A	L749H	.	0	D	D	D	0.849
2	*MLH3*	14	5	Ns	75,506,696	rs28757011	G3488A	G1163D	.	0.01	T	B	D	1.44
2	*MST1*	3	14	Ns	49,722,469	.	C1598G	T533S	.	.	T	B	.	2.63
2	*NOS2*	17	19	Ns	26,093,543	rs28944173	A2239G	T747A	.	0.01	T	B	N	1.94
2	*SLC11A1*	2	14	Ns	219,259,458	rs142636978	G1492A	G498S	.	0.01	T	B	N	-0.54

SNP, single nucleotide polymorphism; UCSC, university of California Santa Cruz; SIFT, sorts intolerant from tolerant substitutions (D, damaging with low confidence or damaging; T, tolerated); Ns, nonsynonymous; Polyphen-2, polymorphism phenotyping 2 (B, benign; D, probably damaging; P, possibly damaging); Mutation Taster (N, not disease-causing; P, possibly disease-causing; D, disease-causing).

Table 3. Deleterious Rare Variants in Inflammatory Bowel Disease-Associated Genes

Gene	SNP									Genotype in each proband																						*In silico* prediction			
	Chromosome	Exon	Variant type	Base pair position in hg18	dbSNP135	Nucleotide change	Amino acid change	UCSC frequency >1%	1000G_2011 allele_freq	Proband 1	Proband 2	Proband 3	Proband 4	Proband 5	Proband 6	Proband 7	Proband 8	Proband 9	Proband 10	Proband 11	Proband 12	Proband 13	Proband 14	Proband 15	Proband 16	Proband 17	Proband 18	Proband 19	Proband 20	Proband 21	Proband 22	SIFT	Polyphen-2	Mutation Taster	PhyloP
ADAD1	4	8	Ns	123,332,475	.	T893A	L298H	.	.	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	D	D	D	2.193
CREB5	7	4	Ns	28,610,098	.	C308T	P103L	.	.	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	D	D	D	2.583
CXCR2	2	3	Ns	219,000,272	.	A748C	M250L	.	.	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	.	D	N	2.045
DUSP5	10		Ns	112,262,500	.	C401T	S134L	.		–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	D	D	D	2.85
FAM55A	11	5	Ns	114,393,621	.	A662G	Y221C	.	.	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	T	D	N	2.031
FAM55A	11	6	Ns	114,392,694	rs79916924	G1214T	C405F	.	0	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	D	D	N	2.511
HNF4A	20	4	Ns	43,042,364	rs1800961	C350T	T117I	O	0.02	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	T	B	D	2.388
IFIH1	2	6	Ns	163,138,942	.	G1240C	A414P	.	0	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	T	D	D	2.822
IL10RA	11	3	Ns	117,860,269	.	C301T	R101W	.	.	–	–	–	–	1	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	D	D	D	2.461
LRRK2	12	32	Ns	40,707,861	rs33958906	C4624T	P1542S	O	0.02	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	T	D	D	2.753
MST1	3	4	Ns	49,724,902	.	G365A	R122Q	.	.	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	T	D	D	2.547
NOS2	17	19	Ns	26,093,543	rs28944173	A2239G	T747A	.	0.01	–	–	1	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	D	B	N	1.941
PLCL1	2	2	Ns	198,948,882	rs150675435	G641T	W214L	.	0	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	.	D	.	2.836
RFTN2	2	1	Ns	198,540,106	.	C77T	P26L	.	.	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	D	D	D	2.747
SH2B1	16	1	Ns	28,878,223	.	C808T	R270W	.	.	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	D	D	N	2.06
SLC11A1	2	3	Ns	219,248,982	.	G167A	R56Q	.	.	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	D	P	D	2.62
SULT1A1	16	4	Ns	28,619,655	.	C329T	P110L	.	.	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	D	D	D	1.566
SULT1A2	16	2	Ns	28,607,104	.	G148A	G50S	.	.	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	.	D	D	2.148
SULT1A2	16	7	Stopgain	28,603,710	rs138147609	G649T	E217X	.	0.01	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	D	.	D	2.3
TAB1	22	5	Ns	39,813,741	rs145235801	C437T	P146L	.	0	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	T	B	D	2.659
TBC1D1	4	11	Ns	38,053,599	rs117452860	C1990T	P664S	.	0.01	–	–	–	1	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	T	P	D	2.664
THADA	2	20	Ns	43,776,463	rs143275203	C2992G	R998G	.	0	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	T	D	D	2.753
TNFSF18	1	1	Ns	173,020,010	.	G93A	M31I	.	.	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	D	B	N	2.76
ZMIZ1	10	24	Ns	81,070,858	rs149174704	G3013A	D1005N	.	0	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	D	D	D	2.327

Novel variants are shown in gray.

1, heterozygote; 2, homozygote; SNP, single nucleotide polymorphism; UCSC, university of California Santa Cruz; SIFT, sorts intolerant from tolerant substitutions (D, damaging with low confidence or damaging; T, tolerated); Ns, nonsynonymous; Polyphen-2, polymorphism phenotyping 2 (B, benign; D, probably damaging; P, possibly damaging); Mutation Taster (N, not disease-causing; D, disease-causing).

ous through *in silico* prediction. Variant rs5771069 was previously associated with ulcerative colitis.[33] However, there is no linkage disequilibrium between rs5771069 and rs142430606.

The association of *ATG16L2* and *IL17REL* with CD was confirmed using an internal CD GWAS database (n=533). The rs11235604 variant of *ATG16L2* was strongly associated with CD in a previous Korean GWAS (odds ratio [OR], 1.63; 95% confidence interval [CI], 1.27 to 2.10; imputed p-value=1.17×10^{-4}) (Supplementary Table 4).[24] The newly identified variant (rs142430606 in *IL17REL*) showed a marginal association with CD (OR, 2.04; 95% CI, 1.001 to 4.14; imputed p-value=4.53×10^{-2}).

Twenty-four unique deleterious variants (10 low-frequency SNPs and 14 novel variants) in 21 genes were identified in the 22 probands (Table 3). The 10 low-frequency SNPs have not previously been reported as associated with IBD. All the variants were in evolutionarily conserved regions of IBD-associated genes; however, it remains to be determined whether heterozygous incidence is deleterious for these variants. No dose effects have previously been reported for these genes with respect to CD phenotypes.

3. Characteristics of coding variants in PID and monogenic IBD genes

Using a PID and monogenic IBD gene panel, 35 variants in 24 PID genes were identified in the 22 probands and among the 35 variants, 10 variants in eight PID genes were predicted to be deleterious (Table 4); however, all the variants were VUS in the heterozygous state with the exception of *XIAP* (p.Q111X; XIAP deficiency), which were identified on the X chromosome of a male patient (proband 13). The XIAP protein plays an important role in activating the nuclear factor κB signaling pathway that leads to proinflammatory cascades.[34] The stopgain mutation (c.C331T; p.Q111X) in proband 13 was confirmed by Sanger sequencing (Fig. 1) and was strongly indicative of XIAP deficiency. The mutation was located prior to the BIR2 and BIR3 domains, which play a role in the recruitment of RIP2 and apoptosis.[35] Proband 13 was diagnosed as having severe CD with perianal fistula at the age of 10 years. He presented reduced natural killer cell activity and recurrent episodes of bicytopenia with bacterial infections. A deleterious p.V561M variant in *CYBB*, of which can cause chronic granulomatous disease, was identified in proband 7. However, his respiratory burst tests were normal.

4. Correlation of patient profiles with deleterious variants

Comorbidity of perianal issues in the 22 CD probands was related to the presence of a heterozygous variant (rs11235604) in *ATG16L2* (p<0.002, chi-square test); however, proband 16, who was homozygous for the rs11235604 variant, did not suffer perianal problems. Probands 5 and 9 died of severe infantile IBD and perianal fistula at the age of 14 months and 8 years, respectively. These two probands carried a heterozygous variant

Table 4. Deleterious Rare Variants in Genes of Primary Immune Deficiency and Monogenic Inflammatory Bowel Disease

Gene	Inheritance	SNP					Genotype in each proband																*In silico* prediction			
		Chromosome	Variant type	Base pair position in hg18	dbSNP135	Nucleotide change	Amino acid change	Proband 1	Proband 2	Proband 3	Proband 4	Proband 5	Proband 6	Proband 7	Proband 8	Proband 9	Proband 10	Proband 11	Proband 12	Proband 13	Proband 14	Proband 15	Proband 16	Proband 17	Proband 18	Proband 19	Proband 20	Proband 21	Proband 22	SIFT	Polyphen-2	Mutation Taster	PhyloP
BLM	AR	15	Ns	91,328,232	.	C2744T	A915V	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	P	D	2.775
CYBB	XL	X	Ns	37,670,138	.	G1681A	V561M	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	D	D	2.31
DOCK8	AR	9	Ns	328,113	rs75352090	C782T	A261V	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	T	D	D	2.798
DOCK8	AR	9	Stop gain	463,655	rs79568455	C5907A	Y1969X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	T	.	D	2.508
FOXN1	AR	17	Ns	26,851,668	.	T271G	F91V	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	D	D	N	2.174
IL10RA	AR	11	Ns	117,860,269	.	C301T	R101W	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	D	D	D	2.461
NCF2	AR	1	Ns	183,556,112	.	G175A	A59T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	D	P	D	2.562
NCF2	AR	1	Ns	183,536,358	rs13306581	C836T	T279M	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	D	D	2.535
RAG1	AR	11	Ns	36,595,188	rs146457887	C334T	R112C	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	D	D	D	2.937
XIAP	XL	X	Stop gain	123,019,843	.	C331T	Q111X	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	D	.	D	1.7

1, heterozygote; 2, homozygote; SNP, single nucleotide polymorphism; SIFT, sorts intolerant from tolerant substitutions (D, damaging with low confidence or damaging; T, tolerated); Polyphen-2, polymorphism phenotyping 2 (D, probably damaging; P, possibly damaging); Mutation Taster (N, not disease-causing; D, disease-causing); Ns, nonsynonymous; AR, autosomal recessive; XL, X-linked.

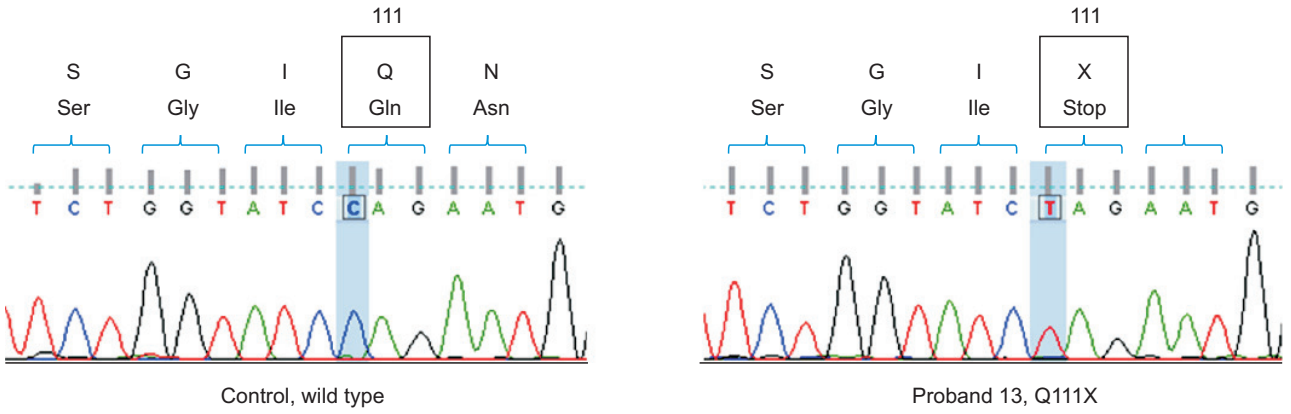

Fig. 1. Sanger sequencing of *XIAP*, showing hemizygous Q111X variants in proband 13.

of *IL10RA* (c.C301T; p.R101W), which was previously reported to be a causative gene for refractory infantile IBD when present in the homozygous state.[36] We therefore performed Sanger sequencing on *IL10RA* in the two probands and their healthy parents; however, no additional homozygote or compound heterozygote mutations in *IL10RA* were identified. In summary, no genotype-phenotype associations were noted in the probands with the exception of XIAP deficiency in proband 13.

DISCUSSION

In this study, we performed WES analysis on samples from 22 children with CD, and identified 81 IBD-associated gene variants and 35 PID genes. One variant, rs11235604 in *ATG16L2*, was already identified as a CD susceptibility locus in our GWAS database.[24] A further variant, rs142430606 in *IL17REL*, was newly identified as a probable disease-causing rare variant. The GWAS dataset confirmed this variant to be marginally associated with CD (OR, 2.04; imputed p-value=4.53×10^{-2}). PID genes were also examined, and a novel p.Q111X variant in *XIAP* was noted in a patient with CD. The majority of the rare variants, particularly 24 unique deleterious variants in conserved loci, were VUS. Further study is needed to determine the functional effects of these mutations. The identification of numerous VUS in the small study population suggests that WES might not yet be applicable to clinical decision making in the treatment of pediatric CD.

One possible explanation for the difficulties in interpreting interesting rare variants is that IBD-associated variants are too rare and genetically heterogeneous to allow statistically significant observation in a small population. Recent GWAS successfully found common disease-causing variants in populations with CD,[1-4] but those common variants accounted for less than 30% of the heritability of CD.[37] The majority of polymorphisms in the human genome are rare variants, but, due to the limited statistical power, the effects of rare variants on polygenic CD are not clear. Lack of information regarding gene function also hampers the interpretation of WES data. Approximately 5,000 genes are prioritized in databases such as OMIM, and functional interpretation of VUS not listed in the databases is difficult. This lack of functional information hampers the prioritization of candidate mutations for further analysis. Unlike diseases exhibiting Mendelian inheritance patterns that have clear genotype-phenotype correlations, complex diseases are affected by the regulatory variation of non-coding regions, cumulative effects of polygenic determinants, gene-gene interactions, gene-environment interactions, and epigenetic gene modification mechanisms, all of which present huge challenges for the study of complex traits. Due to this complexity, WES may not be sufficient to uncover critical determinants. For example, recent WES for complex traits such as type 2 diabetes and idiopathic epilepsy failed to identify any significant rare variants despite the use of large study populations.[14,15]

The scope of our study was additionally limited by the challenges presented by WES analysis.[8,38] First, WES involves applied computational genomics. Different sequencing methods produce sequences of varying length and depth and the results of "loss-of-function" predictions can vary with data formats and annotation software. In addition, detection of short sequence indels is limited to one third of the read length in WES. The very large amount of data required for WES analysis also poses a challenge in determining disease-causing mutations. Capturing specific genomic regions and the exome may reduce the complexity of the data and simplify the computational analysis. In the present study, the analysis was simplified by prioritizing IBD-associated genes from recent GWAS studies and genes of PID and monogenic IBD. Second, *in silico* prediction models show substantial disagreements.[39] In the present study, we used three programs to assess the deleterious extent of the identified mutations. SIFT predictions correlated with Polyphen-2 and Mutation-Taster predictions at levels of 40% to 67%. Care must be taken to avoid false hypotheses that primarily rely on current filtering parameters and variable interpretations of WES data.[38] Therefore, in addition to validation by Sanger sequencing, func-

tional studies are important for the full assessment of deleterious variants; however, it is difficult to perform functional studies on the numerous variants presented in the current study. The identification of numerous VUS does not alleviate the "diagnostic odyssey" needed for some patients.

Nonetheless, based on the fact that IBD-mimicking colitis is frequently observed in immunodeficient infants, WES-based diagnosis for patients with monogenic IBD may be clinically practical.[26] The identification of mutations in *IL10RA* and *XIAP* by WES highlighted the need for hematopoietic stem cell transplantation in affected children.[40,41] One-third of chronic granulomatous disease and one-fifth of XIAP-deficient patients develop a noninfectious chronic IBD similar to CD.[42,43] Common variable immune deficiency, dyskeratosis congenita, immunodysregulation polyendocrinopathy enteropathy X-linked syndrome, and Wiskott-Aldrich syndrome are also frequently accompanied by infantile enterocolitis.[26] Using a panel of PID and monogenic IBD genes, our WES identified a novel *XIAP* variant carried by proband.[13] Detailed guidelines for the diagnosis of IBD using WES remain to be established.

In conclusion, although pediatric patients with severe phenotypes carried a wide spectrum of genetic susceptibility factors for CD, the numerous heterozygous VUS in IBD-associated genes remain to be functionally characterized. Subsequently, those VUS limit the practical clinical application of WES for CD patients and hamper any personalized application of our findings to individual CD patients; however, using WES, a Korean-specific variant in *ATG16L2* was found in CD patients with early-onset and severe phenotype, and a probable candidate variant in *IL17REL* was newly identified. In addition, WES in children may be beneficial for distinguishing CD secondary to PID, for example as a result of the loss of XIAP protein.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGEMENTS

This work was supported by the Mid-Career Researcher Program through National Research Foundation of Korea grant funded by the Ministry of Education, Science and Technology, Korea (2010-0015648). We would like to thank all participating patients and healthy donors who provided the DNA and clinical information necessary for this study.

Specific author contributions: K.S. obtained financial support, conceived, and designed the study and supervised all the sample collection, data analysis, and interpretation. K.M.K. and S.K.Y. participated in study design and supervised all the sample selection. K.M.K., Y.J.L., H.S.J., S.D.Y., S.K.Y., I.L. recruited subjects and participated in diagnostic evaluation. J.K.L., E.J.S., and H.T.L. provided their exome data and critical comments. S.H.O., J.B., E.L., and H.C. performed data analyses. J.B. prepared DNA samples. S.H.O. and J.B. wrote the manuscript as co-first authors. K.S. revised the manuscript.

REFERENCES

1. 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.
2. Imielinski M, Baldassano RN, Griffiths A, et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. Nat Genet 2009;41:1335-1340.
3. 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.
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.
5. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747-753.
6. Panoutsopoulou K, Tachmazidou I, Zeggini E. In search of low-frequency and rare variants affecting complex traits. Hum Mol Genet 2013;22:R16-R21.
7. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447:799-816.
8. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? J Med Genet 2011;48:580-589.
9. Koopman WJ, Willems PH, Smeitink JA. Monogenic mitochondrial disorders. N Engl J Med 2012;366:1132-1141.
10. Al-Herz W, Bousfiha A, Casanova JL, et al. Primary immunodeficiency diseases: an update on the classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency. Front Immunol 2014;5:162.
11. Uhlig HH, Schwerd T, Koletzko S, et al. The diagnostic approach to monogenic very early onset inflammatory bowel disease. Gastroenterology 2014;147:990-1007.e3.
12. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA 2014;312:1880-1887.
13. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 2014;312:1870-1879.
14. Lohmueller KE, Sparsø T, Li Q, et al. Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. Am J Hum Genet 2013;93:1072-1086.
15. Heinzen EL, Depondt C, Cavalleri GL, et al. Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. Am J Hum

Genet 2012;91:293-302.
16. Yamazaki K, Takazoe M, Tanaka T, Kazumori T, Nakamura Y. Absence of mutation in the NOD2/CARD15 gene among 483 Japanese patients with Crohn's disease. J Hum Genet 2002;47:469-472.
17. Yamazaki K, Takazoe M, Tanaka T, et al. Association analysis of SLC22A4, SLC22A5 and DLG5 in Japanese patients with Crohn disease. J Hum Genet 2004;49:664-668.
18. Prescott NJ, Dominy KM, Kubo M, et al. Independent and population-specific association of risk variants at the IRGM locus with Crohn's disease. Hum Mol Genet 2010;19:1828-1839.
19. Ye BD, Yang SK, Cho YK, et al. Clinical features and long-term prognosis of Crohn's disease in Korea. Scand J Gastroenterol 2010; 45:1178-1185.
20. Kim BJ, Song SM, Kim KM, et al. Characteristics and trends in the incidence of inflammatory bowel disease in Korean children: a single-center experience. Dig Dis Sci 2010;55:1989-1995.
21. Kaser A, Zeissig S, Blumberg RS. Genes and environment: how will our concepts on the pathophysiology of IBD develop in the future? Dig Dis 2010;28:395-405.
22. D'Antonio M, D'Onorio De Meo P, Paoletti D, et al. WEP: a high-performance analysis pipeline for whole-exome data. BMC Bioinformatics 2013;14 Suppl 7:S11.
23. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
24. Yang SK, Hong M, Zhao W, et al. Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. Gut 2014;63:80-87.
25. Yang SK, Hong M, Zhao W, et al. Genome-wide association study of ulcerative colitis in Koreans suggests extensive overlapping of genetic susceptibility with Caucasians. Inflamm Bowel Dis 2013;19:954-966.
26. Denson LA. How does knowledge from translational research impact our clinical care of pediatric inflammatory bowel disease patients? Curr Gastroenterol Rep 2012;14:275-281.
27. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812-3814.
28. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7: 248-249.
29. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010;7:575-576.
30. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005;15:1034-1050.
31. Ishibashi K, Fujita N, Kanno E, et al. Atg16L2, a novel isoform of mammalian Atg16L that is not essential for canonical autophagy despite forming an Atg12–5-16L2 complex. Autophagy 2011;7:1500-1513.
32. Homer CR, Richmond AL, Rebert NA, Achkar JP, McDonald C. ATG16L1 and NOD2 interact in an autophagy-dependent antibacterial pathway implicated in Crohn's disease pathogenesis. Gastroenterology 2010;139:1630-1641.e2.
33. Franke A, Balschun T, Sina C, et al. Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL). Nat Genet 2010;42:292-294.
34. Rigaud S, Fondanèche MC, Lambert N, et al. XIAP deficiency in humans causes an X-linked lymphoproliferative syndrome. Nature 2006;444:110-114.
35. Filipovich AH, Zhang K, Snow AL, Marsh RA. X-linked lymphoproliferative syndromes: brothers or distant cousins? Blood 2010;116:3398-3408.
36. Kotlarz D, Beier R, Murugan D, et al. Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy. Gastroenterology 2012;143:347-355.
37. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 2012;337:64-69.
38. Goh G, Choi M. Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. Genomics Inform 2012;10:214-219.
39. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum Mutat 2011;32:894-899.
40. Worthey EA, Mayer AN, Syverson GD, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med 2011;13:255-262.
41. Dinwiddie DL, Bracken JM, Bass JA, et al. Molecular diagnosis of infantile onset inflammatory bowel disease by exome sequencing. Genomics 2013;102:442-447.
42. Marciano BE, Rosenzweig SD, Kleiner DE, et al. Gastrointestinal involvement in chronic granulomatous disease. Pediatrics 2004; 114:462-468.
43. Pachlopnik Schmid J, Canioni D, Moshous D, et al. Clinical similarities and differences of patients with X-linked lymphoproliferative syndrome type 1 (XLP-1/SAP deficiency) versus type 2 (XLP-2/XIAP deficiency). Blood 2011;117:1522-1529.