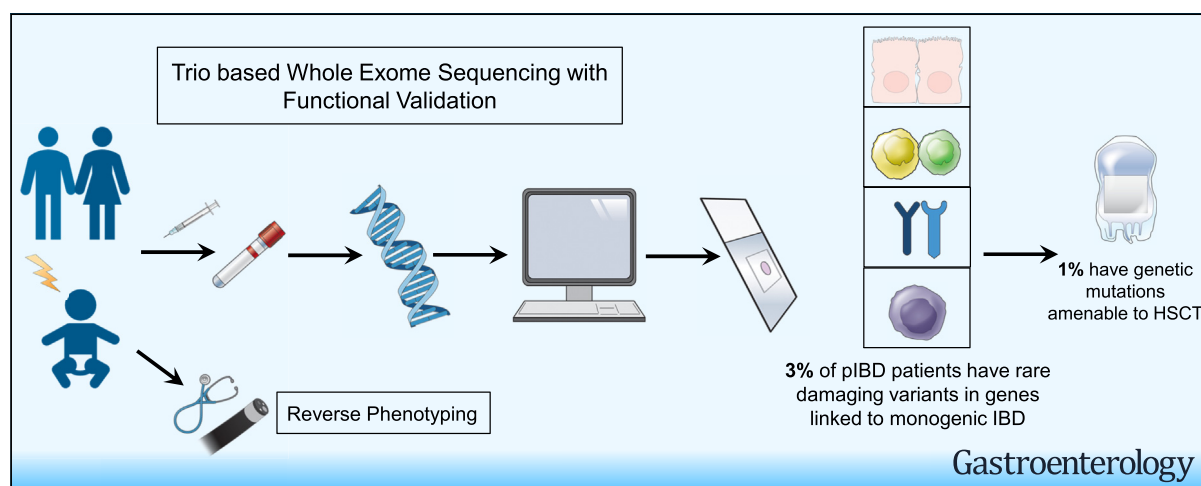# Prevalence and Clinical Features of Inflammatory Bowel Diseases Associated With Monogenic Variants, Identified by Whole-Exome Sequencing in 1000 Children at a Single Center

Eileen Crowley,[1,2,3] Neil Warner,[1] Jie Pan,[1] Sam Khalouei,[4] Abdul Elkadri,[1,5] Karoline Fiedler,[1] Justin Foong,[4] Andrei L. Turinsky,[4] Dana Bronte-Tinkew,[1] Shiqi Zhang,[1] Jamie Hu,[1] David Tian,[1] Dalin Li,[6] Julie Horowitz,[7] Iram Siddiqui,[8] Julia Upton,[9] Chaim M. Roifman,[9] Peter C. Church,[1] Donna A. Wall,[10] Arun K. Ramani,[4] Daniel Kotlarz,[11] Christoph Klein,[11] Holm Uhlig,[12,13] Scott B. Snapper,[14,15] Claudia Gonzaga-Jauregui,[7] Andrew D. Paterson,[16] Dermot P. B. McGovern,[6] Michael Brudno,[4,17] Thomas D. Walters,[1] Anne M. Griffiths,[1] and Aleixo M. Muise[1,18,19]

[1]SickKids Inflammatory Bowel Disease Center, The Hospital for Sick Children, Toronto, Ontario, Canada; [2]School of Medicine, Conway Institute, University College Dublin, Dublin, Ireland; [3]Division of Pediatric Gastroenterology, Western University, Children's Hospital, London Health Sciences Centre, London, Ontario, Canada; [4]Centre for Computational Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; [5]Division of Pediatric Gastroenterology, Medical College of Wisconsin, Milwaukee, Wisconsin; [6]F. Widjaja Foundation Inflammatory Bowel Disease Center and Immunobiology Research Institute at Cedars-Sinai Medical Center, Los Angeles, California; [7]Regeneron Genetics Center, Regeneron Pharmaceuticals Inc, Tarrytown, New York; [8]Division of Pathology, The Hospital for Sick Children, Toronto, Ontario, Canada; [9]Division of Immunology, Department of Pediatrics, University of Toronto, The Hospital for Sick Children, Toronto, Ontario, Canada; [10]Blood and Marrow Transplant/Cellular Therapy, Haematology/Oncology, Department of Pediatrics, University of Toronto, The Hospital for Sick Children, Toronto, Ontario, Canada; [11]Dr von Hauner Children's Hospital, Department of Pediatrics, University Hospital, Ludwig Maximilian University of Munich, Munich, Germany; [12]Translational Gastroenterology Unit, University of Oxford, UK; [13]Department of Pediatrics, University of Oxford, UK; [14]Division of Gastroenterology, Hepatology and Nutrition, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts; [15]Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts; [16]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; [17]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada; [18]Cell Biology Program, Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada; and [19]Department of Pediatrics, Institute of Medical Science and Biochemistry, University of Toronto, The Hospital for Sick Children, Toronto, Ontario, Canada

Trio based Whole Exome Sequencing with Functional Validation

Reverse Phenotyping

1% have genetic mutations amenable to HSCT

3% of pIBD patients have rare damaging variants in genes linked to monogenic IBD

Gastroenterology

**BACKGROUND & AIMS:** A proportion of infants and young children with inflammatory bowel diseases (IBDs) have subtypes associated with a single gene variant (monogenic IBD). We aimed to determine the prevalence of monogenic disease in a cohort of pediatric patients with IBD. **METHODS:** We performed whole-exome sequencing analyses of blood samples from an unselected cohort of 1005 children with IBD, aged 0–18 years (median age at diagnosis, 11.96 years) at a single center in Canada and their family members (2305 samples total). Variants believed to cause IBD were validated using Sanger sequencing. Biopsies from patients were analyzed by immunofluorescence and histochemical analyses.

**RESULTS:** We identified 40 rare variants associated with 21 monogenic genes among 31 of the 1005 children with IBD (including 5 variants in *XIAP*, 3 in *DOCK8*, and 2 each in *FOXP3, GUCY2C,* and *LRBA*). These variants occurred in 7.8% of children younger than 6 years and 2.3% of children aged 6–18 years. Of the 17 patients with monogenic Crohn's disease, 35% had abdominal pain, 24% had nonbloody loose stool, 18% had vomiting, 18% had weight loss, and 5% had intermittent bloody loose stool. The 14 patients with monogenic ulcerative colitis or IBD-unclassified received their diagnosis at a younger age, and their most predominant feature was bloody loose stool (78%). Features associated with monogenic IBD, compared to cases of IBD not associated with a single variant, were age of onset younger than 2 years (odds ratio [OR], 6.30; $P$ = .020), family history of autoimmune disease (OR, 5.12; $P$ = .002), extra-intestinal manifestations (OR, 15.36; $P$ < .0001), and surgery (OR, 3.42; $P$ = .042). Seventeen patients had variants in genes that could be corrected with allogeneic hematopoietic stem cell transplantation. **CONCLUSIONS:** In whole-exome sequencing analyses of more than 1000 children with IBD at a single center, we found that 3% had rare variants in genes previously associated with pediatric IBD. These were associated with different IBD phenotypes, and 1% of the patients had variants that could be potentially corrected with allogeneic hematopoietic stem cell transplantation. Monogenic IBD is rare, but should be considered in analysis of all patients with pediatric onset of IBD.

---

Inflammatory bowel diseases (IBD), composed of Crohn's disease (CD), ulcerative colitis (UC), and IBD-unclassified (IBD-U), are common chronic relapsing inflammatory conditions of the gastrointestinal tract that affect both children and adults.[1,2] Recent studies have demonstrated that IBD is a global disease with a growing prevalence in developed countries and accelerating incidence in newly industrialized countries.[3] It has been estimated that IBD develops during childhood or adolescence in >25% of patients.[4] Canadian health administrative data showed that IBD incidence continues to increase rapidly throughout the pediatric age range, with the greatest increase in children diagnosed before 5 years of age.[5]

Genome-wide association studies have identified >230 risk loci[6–8] explaining only 20%–25% of genetic heritability in complex adult-onset IBD.[9] As with other complex disorders, host genetics are recognized to play a more predominant role in younger children termed *very early onset IBD* (VEOIBD; defined as diagnosis before 6 years of age).[10,11] Next-generation sequencing,[12] including whole-exome sequencing (WES), is now used clinically to investigate monogenic causes in complex diseases,[13,14] including VEOIBD.[15] Utilizing WES, a number of genetic disorders associated with VEOIBD[16,17] (termed *monogenic IBD* for this study; see Supplementary Table 1 for detailed list of monogenic IBD genes) have been identified. Monogenic IBD genes are categorized into those resulting in epithelial[18] and/or immune[16,17] defects and knowledge of tissue-specific gene expression and protein function is critical in determining precision treatment approaches.[15] Curative allogeneic hematopoietic stem cell transplantation (HSCT) is now considered standard of care for select patients with functionally validated immune-related monogenic IBD defects.[17]

The prevalence of monogenic IBD across the full age range (0–18 years) of pediatric IBD is unknown. Therefore, the primary aim of this study was to use WES to interrogate the currently known monogenic IBD genes to determine the prevalence in a single-center cohort of 1005 pediatric IBD patients diagnosed before 18 years of age and identify phenotypic characteristics predictive of monogenic disease.

## Methods

Detailed methods and clinical descriptions are provided in Supplementary Material (Supplementary Tables 1–11 and Supplementary Figures 1–6).

### Patient Population

The study was conducted with Research Ethics Board (REB 1000024905) approval at the Hospital for Sick Children

---

> ### WHAT YOU NEED TO KNOW
>
> #### BACKGROUND AND CONTEXT
>
> A proportion of infants and young children with inflammatory bowel diseases (IBD) have subtypes associated with a single gene variant driving disease (monogenic IBD).
>
> #### NEW FINDINGS
>
> In whole-exome sequencing analyses of more than 1000 children with IBD, we found that 3% had rare variants in genes previously associated with pediatric IBD. These caused different phenotypes, and 1% of the patients had variants that could be corrected with gene therapy and hematopoietic stem cell transplantation
>
> #### LIMITATIONS
>
> Further studies are needed to determine what genetic factors might cause the other 97% of cases of pediatric IBD.
>
> #### IMPACT
>
> Monogenic IBD is rare but should be considered in analysis of all patients with pediatric onset of IBD. Knowledge of genetic factors can be used in prognosis and selection of therapy.

(SickKids). The SickKids IBD Center is the only IBD primary care and referral center for the greater Toronto area and most children in the greater Toronto area with IBD are both diagnosed and followed at SickKids. All IBD patients (and families) diagnosed and/or followed at SickKids younger than 18 years were eligible for inclusion regardless of disease severity or age of diagnosis and enrolled during a 13-year period (2003–2015; Table 1 and Supplementary Tables 2–4 for cohort characteristics). Inclusion criteria were being diagnosed with IBD between 0 and 18 years of age and followed at the SickKids. Exclusion criteria were IBD with known chromosomal abnormalities, diagnosed syndromic disease, previous diagnosed primary immunodeficiency, referral for a second opinion from other regions of Canada, diagnosed with other forms of monogenic intestinal disease, and refusal to consent for genetics for any reason.

### DNA Extraction

Genomic DNA was extracted from peripheral venous blood samples collected in EDTA. DNA concentration was estimated using the Qubit 2.0 Fluorometer and a 260:280 ratio calculated using a NanoDrop spectrophotometer. The mean DNA yield obtained was 150 $\mu$g/mL, and approximately 2 $\mu$g of each patient DNA was extracted for next-generation sequencing.

### Whole-Exome Sequencing

WES was carried out using high-quality genomic DNA that was mechanically fragmented by adaptive focused acoustic technology to a mean size of approximately 150 base pairs. The sheared libraries were prepared for exome capture using a custom-designed, highly automated approach to produce a library appropriately indexed for pooled exome capture and sequencing. The exomes were captured with the NimbleGen VCRome 2.1 reagent. The probe set targets 42 Mb of DNA covering the Vega, CCDS, and RefSeq gene models, microRNAs, and some regulatory regions. Captured samples were polymerase chain reaction–amplified to ensure that exome enrichment and genome de-enrichment were successful. Samples that pass quality control were sequenced on the Illumina HiSeq 2500 platform using paired-end 75 bp reads and two indexing reads.[19]

### Whole-Exome Sequencing and Bioinformatics Pipeline

WES was performed in collaboration with the Regeneron Genetics Center. Regeneron Pharmaceuticals did not contribute to the data analysis, interpretation of findings, or writing of this article. Genes were selected based on known association with monogenic IBD (outlined and referenced in Supplementary Table 1). Genes identified as risk or without validated monogenic association were not included in this analysis. The data files were processed using the Care4Rare bioinformatics pipeline at SickKids, which is composed of the following components; alignment, variant calling, and annotation. Sequencing reads were aligned to human reference genome (GRCh38/hg38) using BWA-mem (Burrows-Wheeler Aligner, version 0.7.12) followed by indel realignment using Genome Analysis Toolkit (GATK, version 3.5), marking polymerase chain reaction duplicates using Picard and base recalibration by GATK. A BED file corresponding to the library preparation capture kit was used in the pipeline to limit the analysis to exonic intervals. The following variant callers were run on the BAM files of each family to produce family-based VCF files: GATK HaplotypeCaller[20] (version 3.5), Vardict[21] (version 1.4.6), Varscan[22]

**Table 1.** Phenotypic Characteristics of Probands in the Sequenced Cohort and the Monogenic Cohort

| Characteristic | Total pediatric IBD Cohort | Non-monogenic pediatric IBD Cohort | Monogenic pediatric IBD Cohort | OR (95% CI) | Bonferroni P Value |
|---|---|---|---|---|---|
| Total n | 1005 | 974 | 31 | — | — |
| Age at diagnosis, y, median (IQR) | 11.96 (8.96–14.21) | 12.04 (9.05–14.25) | 10.83 (3.45–12.53) | — | .12 |
| Age at diagnosis, 0–1.9 y (infantile), % | 2.9 | 2.5 | 13 | 6.30 (1.98–20.08) | **.020** |
| Age at diagnosis 2–5.9 y (VEOIBD), % | 11.2 | 10.7 | 22 | 2.62 (1.06–6.46) | .405 |
| Age at diagnosis 6–9.9 y (EOIBD), % | 17.8 | 18.3 | 10 | 0.67 (0.20–2.32) | 1 |
| Age at diagnosis 10-17.9 years, % | 68.1 | 68.5 | 55 | — | — |
| Sex (% male), % | 60 | 59.6 | 71 | 1.65 (0.78–3.82) | 1 |
| Family history of IBD, % | 32 | 31.7 | 42 | 1.55 (0.74–3.19) | 1 |
| First-degree family history of IBD, % | 14 | 14 | 7 | 0.41 (0.07–1.40) | 1 |
| Family history of autoimmune disease, % | 7 | 6.3 | 26 | 5.12 (2.07–11.48) | **.002** |
| Disease type, % | CD 59 | CD 60 | CD 55 | — | — |
| | UC/IBD-U 41 | UC/IBD-U 40 | UC/IBD-U 45 | | |
| Any EIM, % | 11 | 9.3 | 61 | 15.36 (7.31–33.52) | **<.0001** |
| >1 EIM, % | 3.3 | 2.6 | 26 | 13.20 (5.13–31.51) | **<.0001** |
| Progression to biologic therapy, % | 38 | 38.2 | 32 | 0.77 (0.34–1.62) | 1 |
| Progression to surgical therapy, % | 9.8 | 9.2 | 26 | 3.42 (1.40–7.57) | **.042** |

NOTE. Comparisons are made between monogenic and non-monogenic groups. OR estimates and their 95% confidence intervals (CIs) were computed using logistic regression models (refer to Supplementary Materials). P values in bold type are statistically significant (P < .05 after Bonferroni correction).
EIM, extra-intestinal manifestation; EOIBD, early onset inflammatory bowel disease; IQR, interquartile range.

(version 2.3.9), Samtools[23] (version 1.3), and Freebayes[24] (version 1.0.0). An ensemble approach using bcbio.variation.recall (https://github.com/bcbio/bcbio-nextgen) was then used where only variants called by 2 or more variant callers were retained. Annovar (annovar/2016.02.01) was used to annotate the VCF files (Supplementary Table 6A).[25] These files were subsequently normalized with the vt tool[26] and further annotated with Ensembl VEP[27] (version 82). VCF[28] anno added the annotations in Supplementary Table 5A and B. The vcf2db program (https://github.com/quinlan-lab/vcf2db) was then used to generate GEMINI[29] (-compatible databases from family-based VCF files. GEMINI was used to query the different inheritance models in Supplementary Table 5C using the filters in Supplementary Table 5D. Further quality assessment measures are outlined in the Supplementary Materials (Supplementary Figures 1A–C and 2).

### Coverage Analysis

In order to assess the read depth of the exonic intervals for the 67 VEOIBD genes, the average coverage values from the sample_interval_summary file, generated by the GATK DepthOfCoverage tool, were extracted for all 2309 samples. Using custom Python and R scripts, coverage box-plots were generated for each VEOIBD exonic interval (Supplementary Figure 1C); median read depth was 78.0.

### Sequencing Quality Assessment

Sequencing Quality Assessment is outlined in Supplementary Figures 1 and 2. Transition to transversion ratios per sample were obtained using the bcftools[30] (version 1.6) stats tool. The number of variants called according to functional consequence estimated using Ensembl VEP, version 75. Functional variants included: stop gained, splice donor and acceptors, frameshift, missense, inframe insertion and deletions, initiator codon and splice region variants. The distribution of total number of variants per sample along with the number of heterozygous and homozygous variants were calculated. Sex analysis was performed, by investigating the number of homozygous and heterozygous variants in X chromosome for both male and females. When assessing the number of heterozygous variants in X chromosome, 5 outliers were identified; 1 was incorrectly coded in the database and 4 were subsequently excluded from further analysis. The coefficient of inbreeding was calculated using plink (version 2.0) software. We applied the "Plink –make-king" tool to the project-level merged VCF file containing all the variants of the entire dataset. Only autosomes were considered by the program. The spouse pairs showing a high coefficient of inbreeding value (>0.7) were reviewed for consanguinity. The genotype quality of variants were extracted from merged VCF file using a custom script and the distribution was plotted as a box plot.

### Sanger Sequencing

Selected variants predicted to be pathogenic and assessed as deleterious by annotation tools outlined above were verified by Sanger sequencing in probands and relatives if sufficient DNA was available. Primers were designed using primerBLAST and polymerase chain reaction was performed in our research laboratory. The Sanger sequencing service was provided by ACGT (Toronto, ON, Canada).

### Genotype Phenotype Analysis

After high-quality filtering, each patient deemed to have a protein coding or splice variant that was high quality and rare (minor allele frequency < 0.01) were reverse phenotyped. This meant that clinical data were extracted from the database with deep phenotyping performed on any outstanding clinical details via access to an electronic medical record system. Immune and pathology workup was clinically driven, and results were accessed on the electronic medical system.

### Paris Classification

Accurate phenotype classification is essential in determining the utility of genotype-phenotype correlation. The Paris Classification (Supplementary Table 4) was developed by a group of experts in pediatric IBD in 2009.[31] This was an update on the previously published Montreal Classification of IBD.[32] The Paris Classification considers age at diagnosis (A1a <10 years, A1b 10–17 years), location (L1 distal 1/3 ileum ± limited cecal disease; L2 colonic, L3 ileocolonic, L4 isolated upper disease) and behaviour of disease (B1 nonstricturing, nonpenetrating; B2 stricturing; B3 penetrating; p perianal disease modifier), along with consideration of linear growth impairment (G0 no evidence of growth delay, G1 growth delay). This aims to capture the more dynamic features of the pediatric IBD phenotype resulting in uniform standards for defining IBD phenotypes.

### Statistical Methods

Descriptive statistics were provided with medians and interquartile range for continuous variables. Mann-Whitney U test was used for non-normally distributed continuous variables. Chi-square or Fisher's exact test was applied for categorical variables. Categorical variables were compared by calculating an odds ratio (OR) using logistic regression models. For the analysis of age in pediatric IBD cohorts, the age group 10–17.99 years was used as the baseline level, to which other age groups were compared using logistic regression. Results were considered statistically significant when $P < .05$ after Bonferroni correction for multiple testing. Statistical analyses were performed by using the SPSS software, version 22.0 (SPSS, Chicago, IL), as well as the R function glm and R package nnet for logistic regression modeling.

## Results

### Cohort Characteristics

In total, 2305 (99.8%) participants (1005 pediatric IBD patients, and 1300 parents and siblings) were analyzed (4 of 2309 individuals failed quality control; Figure 1A, Supplementary Table 2). Forty-nine percent of pediatric IBD patients were part of complete trios (patient and both of their parents) including 26 quads (trio plus sibling; Figure 1B) and 77% of patients had at least 1 first-degree family member sequenced (including 105 affected first-degree family members; Supplementary Tables 2 and 3). Pediatric IBD patients had a 1.5:1 male to female ratio and were diagnosed with CD (n = 601 [60%]) and UC/IBD-U (n = 404 [40%]) (Supplementary Table 4). The median age at diagnosis was 11.96 years and the median age at symptom onset was 10.65 years (Table 1; see age
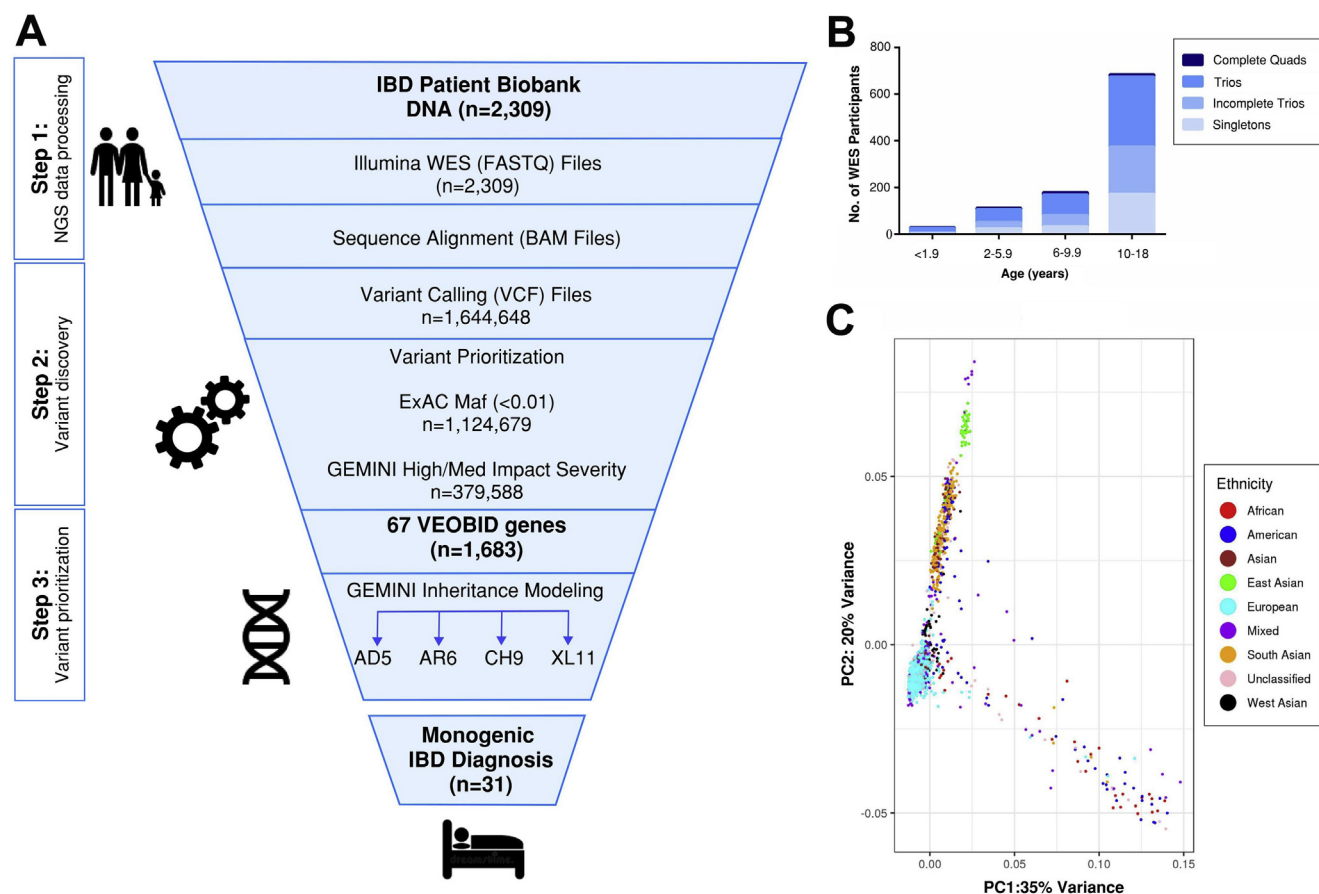
**Figure 1.** Flowchart of WES pediatric IBD study. (*A*) Flowchart of WES in pediatric IBD study; 1,644,648 variants were called. Following variant prioritization 1,124,679 were ExAC MAF < 0.01 and 379,588 high/med impact severity. Variants then underwent GEMINI Pedigree Analysis and inheritance modeling. In the known monogenic IBD genes 5 were autosomal dominant; 6 autosomal recessive; 9 compound heterozygotes (note compound heterozygotes in Table 2 is denoted by autosomal recessive a/b); and 11 X-linked recessive (XL) inheritance. (*B*) Familial inheritance-based analysis by age group of the sequenced cohort. *Bar graph* displays the sequencing of the cohort based on singletons (probands alone); incomplete trios; complete trios; complete quads by age group. A "trio" is defined as a sequenced proband and both parents. An "incomplete trio" is a proband and any relative. A "complete quad" is a proband, both parents and a relative. (*C*) Principal component analysis. Ethnic diversity demonstrated among the cohort.

distribution of the patients in Figure 1*B*, Supplementary Table 4). Principal component analysis of the cohort demonstrated broad ethnic diversity including European, East Asian, or South Asian ethnicity (Figure 1*C*, Supplementary Figure 2, and Supplementary Table 5*E*).

## Identification of Rare Damaging Variants in Monogenic Inflammatory Bowel Disease Genes

GEMINI analysis and initial filtering of the WES data was based on rare, protein coding variants, deleterious predictions, and damaging scores (including combined annotation-dependent depletion score >15[33]) and a secondary manual filtering based on confirmatory inheritance pattern, segregation, concurrence with clinical features associated with phenotypes of known genetic disease, as described in Supplementary Table 1, and pathogenicity was based on the American College of Medical Genetics and Genomics classification.[34] In 31 (3%) of the 1005 pediatric IBD patients, we identified 40 distinct rare damaging

variants in 21 of the known 67 monogenic IBD genes (Table 2). Of the 31 patients with monogenic IBD variants, 23 patients (74%) were sequenced as trios, 4 patients (13%) as incomplete trios, and 4 patients (13%) as singletons. All 40 predicted pathogenic monogenic IBD variants were orthogonally validated using Sanger sequencing (data not shown) and transmission was confirmed to be either autosomal dominant, autosomal recessive (all bi-allelic), or X-linked recessive. Functional de novo variants in the IBD monogenic genes were not identified in any patient in this cohort. Functionally, 67% were missense variants and 32% were predicted loss-of-function alterations, stop-gained, frameshift, splice-site, or in-frame indels (Table 2, Figure 2*A*). As shown in Figure 2*B*, among the 31 children harboring variants in known monogenic IBD genes, those most represented were *XIAP* (5 of 31 [16%]); *DOCK8* (3 of 31 [10%]); *ARPC1B*, *FOXP3*, *GUCY2C*, and *LRBA* (2 of 31 [6%]). Overall, 3% of the 1005 pediatric IBD patients were suspected to have disease-causing variants in monogenic IBD genes.

**Table 2.** Variants Identified Among Monogenic Inflammatory Bowel Disease Genes

| Biological category | Patient | Sex | Age at diagnosis | Chrom | Gene | Inheritance model | aa mutation | Impact | ExAC maf | CADD 1-3 Phred score | Family history | Reported Inheritance | Causal evidence |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Epithelial barrier & response defects | 1 | M | 15 | chr2 | ALPI | AR a | A360V | missense | 0.0006363 | 26.8 | + | AR | G, F, P |
| | | | | | | AR b | Q439X | stop gained | 0.0001295 | 29.7 | | | |
| | 2 | F | 11.91 | chr3 | COL7A1 | AR a | R1696C | missense | 0.0003 | 28.7 | - | AR/AD (milder) | P |
| | | | | | | AR b | R523W | missense | None | 25.5 | | | |
| | 3 | M | 12.3 | chr12 | GUCY2C | AD | G549S | missense | 0.000045 | 33 | + | AD | N/A |
| | 4 | F | 15.1 | chr12 | GUCY2C | AD | F525L | missense | 0.000067 | 25.5 | - | AD | N/A |
| | 5 | M | 11.38 | chr3 | SLCO2A1 | AR a | R314W | missense | 0.0000164 | 26.2 | - | AR | P |
| | | | | | | AR b | M539V | missense | 0.0003 | 27.1 | | | |
| | 6 | F | 0 | chr2 | TTC7A | AR a | E71K | missense | 0.0000082 | 34 | - | AR | G, F, P |
| | | | | | | AR b | Q526X | stop gained | 0 | 36 | | AR | |
| T and B cell differentiation defects | 7 | M | 2.75 | chr7 | ARPC1B | AR | A105V | missense | 0 | 25.1 | - | AR | G, F, P |
| | 8** | M | 3.5 | chr7 | ARPC1B | AR | V91WfsX121 | frameshift | 0 | none | - | AR | G, F, P |
| | 9 | M | 15.52 | chrX | BTK | XL | N484K | missense | 0 | 23 | - | XL | F, P |
| | 10 | M | 5.5 | chrX | DKC1 | XL | T49M | missense | 0 | 26 | - | XL | G, F, P |
| | 11 | F | 12.38 | chr9 | DOCK8 | AR a | | splice | 0.0000411 | 35 | - | AR/AD (milder) | G, P |
| | | | | | | AR b | S97L | missense | 0.0005 | 26.4 | | | |
| | 12 | F | 14.73 | chr9 | DOCK8 | AR a | R947C | missense | 0.0007 | 22.6 | + | AR/AD (milder) | P |
| | | | | | | AR b | R532Q | missense | 0.0000414 | 26.4 | | | |
| | 13 | F | 14.65 | chr9 | DOCK8 | AR a | R59H | missense | 0.0021 | 26.7 | + | AR/AD (milder) | P |
| | | | | | | AR b | Y19C | missense | 0 | 22.1 | | | |
| | 14 | M | 1.63 | chr4 | LRBA | AR | V737I | missense | 0.0002 | 23.2 | - | AR | P |
| | 15** | M | 9.33 | chr4 | LRBA | AR* | E1916X | missense | 0 | None | - | AR | G, F, P |
| | 16 | F | 2.5 | chr2 | STAT1 | AD | V389A | missense | 0 | 19.69 | - | AD | G, F, P |
| Hyper- & Auto-inflammatory disorders | 17 | M | 14 | chr10 | HPS1 | AR a | E9D | missense | 0.0016 | 20.4 | + | AR | p |
| | | | | | | AR b | Y81F | missense | 0.0000695 | 23.4 | | | |
| | 18 | M | 12.75 | chr1 | PIK3CD | AD | V616A | missense | 0.0000083 | 27.1 | +^ | AD | N/A |
| | 19** | M | 8.4 | chrX | SH2D1A | XL | M1T | start Met loss | 0 | 23.4 | - | XL | G, F, P |
| | 20 | M | 11.86 | chrX | XIAP | XL | R29K | missense | 0 | 25.7 | + | XL | P |
| | 21 | M | 9.1 | chrX | XIAP | XL | S278X | stop gained | 0 | 37 | + | XL | G, P |
| | 22 | M | 10.83 | chrX | XIAP | XL | A216T | missense | 0 | 25.6 | + | XL | P |
| | 23 | M | 10.67 | chrX | XIAP | XL | ΔE349 | inframe deletion | 0.0004 | None | - | XL | G, P |
| | 24** | M | 0 | chrX | XIAP | XL | V298fsX306 | frameshift | 0 | None | - | XL | G, F, P |
| | 25** | M | 3.4 | chrX | CYBB | XL | A84A | splice | 0.000012 | 19.34 | - | XL | G, F, P |
| Regulatory T cells and immune regulation | 26 | M | 12 | chr2 | CTLA4 | AD | | splice | 0 | 28 | - | AD | G, P |
| | 27 | M | 3.05 | chrX | FOXP3 | XL | L198P | missense | 0 | 24.4 | + | XL | F, P |
| | 28 | M | 12.68 | chrX | FOXP3 | XL | H121Y | missense | 0.0000782 | 21 | + | XL | F, P |
| | 29** | F | 0.17 | chr21 | IL10RB | AR | W3X | stop gained | 0 | 27.6 | +^ | AR | G, F, P |
| Others | 30 | F | 4.42 | chr2 | HSPA1L | AR | K495EfsX13 | frameshift | 0.0003 | None | - | | G, P |
| | 31 | M | 12.26 | chr1 | MASP2 | AR a | D244N | missense | 0.000068 | 28.7 | + | AR | N/A |
| | | | | | | AR b | W513C | missense | 8.24E-06 | 24.7 | | | |

+, yes; −, no; ^, first degree; aa, amino acid; AD, autosomal dominant; AR, autosomal recessive; ARa/ARb, compound heterozygous; CADD, combined annotation dependent depletion; Chrom/chr, chromosome; F, functional; G, genetic; P, pathology; XL, X-linked recessive.
**Patient had allogeneic HSCT.
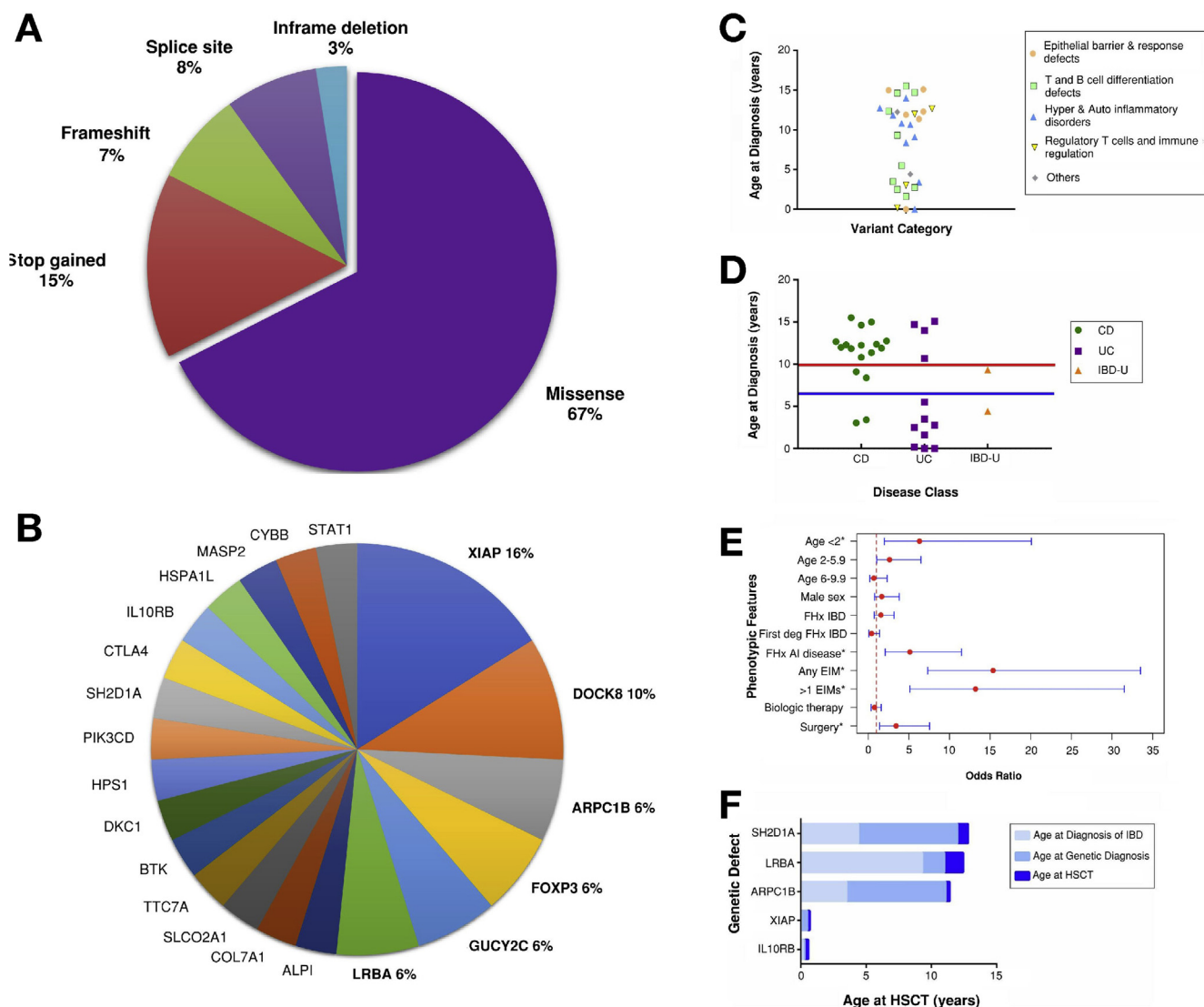
**Figure 2.** Characteristics of identified monogenic IBD population. (*A*) Variant type. Variant types identified using WES in the pediatric IBD cohort. (*B*) Gene variants identified on WES analysis. *Graph* demonstrates gene variants with the most common highlighted in *bold*. (*C*) Age at diagnosis as per variant class. Epithelial barrier response defects; T- and B-cell differentiation defect; hyper- and auto-inflammatory disorders; regulatory T cells and immune regulation; and other. *Colors* represent the color used in Table 2. (*D*) Age at diagnosis of IBD and disease phenotype. *Graph* displays monogenic pediatric IBD variant diagnosed in VEOIBD age group (35%) indicated by *blue line* and older than 10 years of age (55%) (Paris 1b) indicated by *red line*. (*E*) Odds ratio analysis of phenotypic features of identified monogenic IBD. ORs and their confidence intervals (CIs) were computed using logistic regression models (see Supplementary Materials). AI, autoimmune; EIM, extra-intestinal manifestation; FHx, family history. Each horizontal line represents the 95% CI for OR (*red dot*). *Dashed line* indicates expected value of 1.0. *Statistical significance (*P* < .05 after Bonferroni correction). (*F*) Age at HSCT. Age of patient at time of HSCT (in years) with identified monogenic IBD variant in *ARPC1B* (patient 8), *IL10RB* (patient 29), *LRBA* (patient 15), *SH2D1A* (patient 19), *XIAP* (patient 24). *Green line* refers to age of diagnosis of IBD. *Red line* refers to age of monogenic diagnosis. *Blue line* refers to age of HSCT.

## Functional Validation

While in silico pathogenicity of variants identified using WES was based on combined annotation-dependent depletion score,[33] inheritance pattern, impact, frequency, and the American College of Medical Genetics and Genomics classification,[34] we attempted to further validate each variant through retrospective clinical assessment of the probands, functional testing, and/or pathologic examination of biopsy samples (Tables 2 and Supplementary Tables 6–8, and

Supplementary Figures 3–6 for detailed review of each patient). All patients had clinical evidence supporting the known monogenic disease phenotype (clinical features[5,16] are outlined in Supplementary Table 1 and detailed for each patient in Supplementary Table 8), although some patients had milder or incomplete forms of the disease as previously demonstrated in patients with monogenic-forms of IBD.[15] The majority of patients had multiple levels of support, including 16 patients with genetic support of

BASIC AND
TRANSLATIONAL AT

**Table 3.** Phenotypic Features of Identified Monogenic Inflammatory Bowel Disease Cohort

| Monogenic cohort (n = 31) | CD | UC/IBD-U |
|---|---|---|
| Disease class, n | 17 | 14 |
| Sex, % *male* | 82 | 57 |
| Age at diagnosis, *y*, median (IQR) | 12 (9.5-12.7) | 3.9 (0.9-11.5) |
| Age at diagnosis, 0–5.9 y (VEOIBD), n (%) | 2 (6) | 9 (29) |
| Age at diagnosis, 6–9.9 y (EOIBD), n (%) | 2 (6) | 1 (4) |
| Age at diagnosis, 10–17.9 y, n (%) | 13 (42) | 4 (13) |
| Most common presenting feature, % | Abdominal pain 35 | Bloody loose stool 78 |
| Extra-intestinal manifestation, % | 76 | 42 |
| Family history, IBD, % | 58 | 21 |
| Family history, autoimmune disease, % | 29 | 21 |
| Personal history, autoimmune disease, % | 23 | 35 |
| Consanguinity, % | 0 | 14 |
| Ethnicities, % | Caucasian 70 | Caucasian 37 |
|  | South Asian 18 | African 21 |
|  | East Asian 6 | Mixed 21 |
|  | Mixed 6 | East Asian 14 |
|  |  | South Asian 7 |
| Abnormality on immune workup, % | 18 | 28 |
| Serology, % | Positive ASCA 17 | ANCA positive 36 |
|  | Positive ANA 6 | ANA positive 14 |
| Endoscopy, % | L3 35 | E4 42 |
|  | L3L4a 29 | E3 28 |
|  | L3L4b 12 | E2 7 |
|  | L2 12 | Limited due to disease severity 23 |
|  | L1 12 |  |
| Histopathology, % | Granulomatous inflammation, 35 | Apoptosis 28 |
| Therapies, % | Surgery 29 | Surgery 21 |
|  | Biologic agent 53 | Biologic agent 14 |
|  | HSCT 12 | HSCT 28 |
|  | Other 6 | Other 37 |
| Outcome | — | 1 death |

ANA, anti-nuclear antibody; ANCA, anti-neutrophil cytoplasm antibodies; ASCA, anti-saccharomyces cerevisiae antibodies

causality with either known ClinVar pathogenicity and/or loss-of-function variants. Functional testing including protein expression, immunological testing, and biochemical assays were carried out in 14 patients, including 3 patients without further genetic evidence. To provide additional support for the patients without available clinical testing, samples were examined for both known histologic features of disease[35] and protein expression/localization based on known RNA/protein expression defined by Protein Atlas (www.proteinatlas.org). We identified either pathological features associated with monogenic disease and/or altered protein staining in all 27 patient samples examined, including 8 patients without other supporting genetic or functional evidence (Supplementary Figure 3–6). Therefore, only 4 patients with variants in 3 genes (*GUCY2C, PIK3CD,* and *MASP2*) did not have supporting evidence, as these 3 genes had neither clinical functional tests nor validated antibodies available to examine pathology. Together this further supports the in silico prediction in 27 of 31 patients and the role of these genes in the development of IBD.

## Implications of Genetic Diagnoses

Demographic and phenotypic clinical characteristics of the probands with putative monogenic IBD are summarized in Table 3 and Supplementary Table 8 for detailed phenotyping. Figure 2*C* and *D* and Table 1 showed that in the monogenic group the median age at diagnosis was 10.83 years and median age of symptom onset of 9.69 years and 20 of 31 patients (64%) were diagnosed at older than 6 years of age. In the monogenic IBD variant group, 71% were male, 17 (55%) were diagnosed with Crohn's disease, and 14 (45%) were UC/IBD-U (Figure 2*D*). The presenting clinical features for monogenic CD patients (n = 17) were abdominal pain (35%), nonbloody loose stool (24%), vomiting (18%), weight loss, (18%), and intermittent bloody loose stool (5%) (Table 3). The monogenic UC/IBD-U patients (n = 14) were diagnosed at a younger age and the most predominant presenting clinical feature was bloody loose stool (78%) (Supplementary Table 3). Features associated with monogenic disease in comparison to the remaining pediatric IBD cohort were age of onset of disease younger than 2 years (OR, 6.30; $P = .022$), family history of autoimmune disease (OR, 5.12; $P = .002$), extra-intestinal manifestations of IBD (OR, 15.36; $P < .0001$) and surgery (OR, 3.42; $P = 0.046$) (Figure 2*E* and Table 1).

In total, 17 of 31 monogenic IBD patients (>1% of total cohort) had variants in genes known to be amenable to allogeneic stem cell transplant (*ARPC1B, IL10RB, LRBA,*

SH2D1A, XIAP, CYBB, CTLA4, STAT1, BTK, and FOXP3; Supplementary Table 9). All 17 patient/families via their most responsible physicians (including those transitioned to adult care) were informed of the genetic variants and patients are undergoing clinical genetic validation and counseling where appropriate, as some patients had milder disease and transplantation may not be a recommended or preferable treatment option. Six patients (19% of monogenic IBD cohort) already had allogeneic HSCT (Table 2 and Supplementary Table 8: patient 8, ARPC1B; patient 29, IL10RB; patient 15, LRBA; patient 19, SH2D1A; patient 24, XIAP; patient 25, CYBB; all unpublished; patient 25 was transplanted elsewhere and not included in the subsequent analysis). The median age at transplantation in these 5 patients was 11.5 years (interquartile range, 0.45–12.62 years), and there was often a delay between onset of disease and ultimate genetic diagnosis and HSCT (Figure 2F). Of note, monogenic epithelial IBD or combined epithelial-immune defects (ie, TTC7A deficiency and the 1 death in this study population; patient 6[36]) may not respond to HSCT or biologic therapies and may present specific therapeutic challenges.[37]

## Discussion

In this single-center cohort study of 1005 pediatric IBD patients we utilized WES to determine a 3% prevalence of damaging variants in genes linked to monogenic IBD. A number of studies have shown an estimated prevalence of monogenic IBD between 0 to 70%[38–44] (reviewed in Supplementary Table 10). However, these studies are difficult to compare as they examine only subsets of monogenic IBD genes and use a number of sequencing methodologies, including genetic panels, WES, and mixed methodologies. Moreover, these studies may have significant selection bias, as the highest rates of monogenic disease are within cohort studies of patients referred with severe disease and/or very young age of onset increasing the likelihood to identify monogenic disease. Before our study, few studies have examined older pediatric patients in a systematic way. In Toronto, Canada, there are very few community pediatric gastroenterologists; therefore, the vast majority of pediatric IBD patients in the greater Toronto area (catchment area population of approximately 6 million) are diagnosed and followed until 18 years of age at SickKids and make up the cohort described here. This patient cohort is a major strength of this study, as it is a large heterogeneous, multi-ethnic, well-characterized, unselected cohort of children diagnosed from a single pediatric IBD center and patients/families were enrolled regardless of age of diagnosis and disease severity. Another strength of our pediatric IBD cohort was that the majority of patients had at least 1 family member sequenced allowing for family-based genetic analysis.

The frequency of monogenic variants in VEOIBD (7.8%) reinforces that exome sequencing should become part of standard of care for this group of patients diagnosed with IBD, especially children diagnosed younger than 2 years of age. Pediatric gastroenterologists may screen for monogenic

forms of IBD in very young children; however, previous studies have not ascertained the prevalence of these genes across the entire pediatric age range. As described, most monogenic IBD studies have focused only on very young children or young children with the most severe forms of disease,[38,40,45,46] while this study examined an unbiased cohort of patients and extends the age of onset of monogenic IBD throughout the pediatric age range. Another key finding of our study was that 64% (20 of 31) of the monogenic pediatric IBD patients presented after 6 years of age. We found an unexpected prevalence of 2.3% of monogenic variants in all children aged 6 years and older. For these older children, the phenotypic features, including extra-intestinal manifestations of IBD and family history of autoimmune disease, may be used to select patients for consideration of WES analysis.

For each monogenic disorder associated with chronic IBD, the bowel inflammation often has variable penetrance and is only 1 component of a disease that may manifest with a wide spectrum of phenotypes[15–18,47–49] (Supplementary Table 1). Detailed phenotyping of each patient with a monogenic IBD variant (Supplementary Tables 8A–U, Supplementary Figures 4–6) suggests that within each group of genes, the phenotypic variation is likely due to the genetic heterogeneity of disease-causing variants, genetic disease modifiers, and undetermined environmental factors. Interestingly, Huang et al[50] demonstrated that patients with chronic granulomatous disease who developed IBD had a higher polygenic risk score for IBD genome-wide association studies variants when compared to chronic granulomatous disease patients without IBD. We similarly developed a polygenic risk score and compared non-IBD controls (n = 7492), Toronto cohort patients with monogenic IBD variant carriage (n = 31), and Toronto cohort IBD patients without monogenic IBD variant carriage (n = 974). However, we did not identify any significant differences in polygenic risk score between Toronto cohorts maybe due to the small number of patients with each type of monogenic disease (data not shown). It is interesting to speculate that IBD genome-wide association studies risk variants coupled to environmental factors are driving the IBD presentation in some patients with monogenic forms of disease. Overall, our findings suggest a wide heterogeneity in monogenic IBD clinical presentations with earlier age of onset, a family history of autoimmune disease, extra-intestinal manifestations of IBD, and surgery as indicators of monogenic disease (Table 1, Figure 2E). However, these features are common in pediatric IBD and detailed phenotyping of the monogenic cohort may have resulted in an over-representation in these patients.

Special consideration should be given to specific gene expression of epithelial vs immune monogenic forms of IBD (Table 2, Figure 2C). The intestinal epithelial barrier is composed of a layer of columnar cells that function as a gateway between the gut lumen and the lamina propria. Of those epithelial defects purported to be associated with monogenic IBD,[18] we identified ALPI, COL7A1, TTC7A, GUCY2C, and SLCO2A1 in our cohort. Variants in these genes may cause perturbations in the epithelial barrier leading to

immune dysregulation resulting in IBD.[18] This group of patients present a clinical challenge as there are no specific epithelial treatments available that improve the epithelial barrier dysfunction and HSCT is not a viable treatment option. However, a recently published preclinical study has identified leflunomide as a potential therapy for TTC7A-deficiency.[51]

We also identified monogenic immune IBD genes involved in a number of cellular processes outlined in Table 2 and Figure 2C and each patient discussed in detail in Supplementary Figures 4–6. For example, we identified variants in T- and B-cell differentiation including BTK (Supplementary Figure 4) and DOCK8. DOCK8 has diverse roles in the immune system, including regulation of the actin cytoskeleton and can present from infancy to adulthood with variable symptoms including severe infections, atopy, autoimmunity, cancer, and IBD.[52] In this study, 3 patients had bi-allelic DOCK8 variants with clinical features consistent with DOCK8-deficiency (Supplementary Table 8H, Supplementary Figure 3G), including eczema and food allergy although they did not have truncating variants classically associated with this disease. FOXP3 is a transcription factor that is specifically expressed in regulatory T cells, which play a critical role in T-cell tolerance.[53,54] Variants in FOXP3 can lead to X-linked, immune dysregulation, polyendocrinopathy, and enteropathy, IPEX syndrome. In this study, 2 male patients with chronic colitis and other autoimmune and extra-intestinal features were identified (Supplementary Table 8I, Supplementary Figure 5). Variants in XIAP cause an X-linked recessive disorder with a widely reported age of onset and diverse phenotypes including infantile onset and predisposition to hemophagocytic lymphohistiocytosis and lymphoproliferative syndrome. Zeissig et al[55] reported XIAP variants in 4% of all male pediatric CD patients. Here we demonstrated that 1% (5 of 391) of male pediatric CD patients had XIAP variants of which 1 patient with a V298fsX306 XIAP variant was successfully transplanted and currently has no active disease (patient 24, Table 2, Supplementary Table 8U, and Supplementary Figure 6A).

A major difficulty in utilizing genetics in clinical care of children with IBD is the lack of standardized functional testing. This is a critical step in precision medicine, especially when recommending major alternative treatment strategies, such as allogeneic bone marrow transplant in patients with primary immunodeficiencies genes associated with monogenic IBD[16] or palliation in patients with severe forms of TTC7A-deficiency[56] or PLVAP-deficiency.[57] There are a few genes where the protein product can be easily assayed. For example, in IL10R deficiency STAT3 phosphorylation can be measured after IL10 stimulation[58] and reactive oxygen species can be measured in chronic granulomatous disease, although disease-causing variants in NCF4 may have normal reactive oxygen species production.[59] While for others genes, only experimental biochemical assays are available in selected research laboratories, for example, XIAP,[60] TTC7A,[36] and ARPC1B.[61] Protein expression of a number of monogenic IBD genes associated with primary immunodeficiencies can be measured using flow cytometry-based assays, such as LRBA, FOXP3, and XIAP[17]; however, missense variants, which may result in normal gene expression and deleterious protein function will not be identified using this methodology. In an attempt to further validate causative variants, we utilized a combination of genetic, functional and/or pathological approaches; however, further standardized testing is necessary for all patients with monogenic IBD variants and critical for those with variants potentially amenable to allogeneic HSCT. Furthermore, when functional testing is not available, as with most of the variants described here, patients must be fully informed of the inherent risks of genetic interpretation on therapeutic decisions.

There are limitations with the WES methodology used in this study with approximately 5% of exons are poorly covered[12] (see Supplementary Figure 1B for exon coverage of monogenic IBD genes). In this study, this limitation was illustrated by the poor WES coverage of exon 1 in SH2D1A. Using manual review of raw WES data and Sanger sequencing validation, we identified a variant (M1T) resulting in the loss of the start methionine in exon 1 of SH2D1A in a patient with severe ileitis, colitis, and growth failure (patient 19, Table 2 and Supplementary Table 8Q, Supplementary Figure 6A). The identification of this SH2D1A variant resulted in curative HSCT for this patient. Also WES does not cover noncoding yet potentially functional regions of the genome and has limited capacity to identify copy number changes and structural variants.[12] Furthermore, there has been a rapid increase in the discovery of monogenic IBD genes[47] and we anticipate that many more genes will be discovered. Therefore, this study likely underestimates the contribution of monogenic gene disorders in pediatric IBD.

Overall, this single pediatric IBD center study supports a 3% prevalence of damaging variants in genes linked to monogenic IBD. Most importantly, this study demonstrates that 1% of monogenic pediatric IBD patients have variants in genes associated with primary immunodeficiency that are potentially curable through allogeneic HSCT (Supplementary Table 9). We believe this data supports the diagnosis of monogenic disease beyond the very early onset IBD population, especially in children with a family history of autoimmune diseases and those with evidence of extra-intestinal manifestations of IBD. Molecular identification of disease-causing variants in monogenic disease genes can inform patient management and improve outcomes by targeting definitive and personalized treatment strategies.

## Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2020.02.023.

## References

1. Baumgart DC, Sandborn WJ. Crohn's disease. Lancet 2012;380:1590–1605.

2. Ordas I, Eckmann L, Talamini M, et al. Ulcerative colitis. Lancet 2012;380:1606–1619.

3. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 2017;390:2769–2778.

4. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. Gastroenterology 2004;126:1504–1517.

5. Benchimol EI, Bernstein CN, Bitton A, et al. Trends in epidemiology of pediatric inflammatory bowel disease in Canada: distributed network analysis of multiple population-based provincial health administrative databases. Am J Gastroenterol 2017;112:1120–1134.

6. 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.

7. **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.

8. Huang H, Fang M, Jostins L, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature 2017;547:173–178.

9. Park JH, Wacholder S, Gail MH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 2010;42:570–575.

10. Uhlig HH. Monogenic diseases associated with intestinal inflammation: implications for the understanding of inflammatory bowel disease. Gut 2013;62:1795–1805.

11. Thiagarajah JR, Kamin DS, Acra S, et al. Advances in evaluation of chronic diarrhea in infants. Gastroenterology 2018;154:2045–2059.e6.

12. Adams DR, Eng CM. Next-generation sequencing to diagnose suspected genetic disorders. N Engl J Med 2018;379:1353–1362.

13. Groopman EE, Marasa M, Cameron-Christie S, et al. Diagnostic utility of exome sequencing for kidney disease. N Engl J Med 2019;380:142–151.

14. Splinter K, Adams DR, Bacino CA, et al. Effect of genetic diagnosis on patients with previously undiagnosed disease. N Engl J Med 2018;379:2131–2139.

15. Uhlig HH, Muise AM. Clinical genomics in inflammatory bowel disease. Trends Genet 2017;33:629–641.

16. 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.

17. Crowley E, Muise A. Inflammatory bowel disease: what very early onset disease teaches us. Gastroenterol Clin North Am 2018;47:755–772.

18. Leung G, Muise AM. Monogenic intestinal epithelium defects and the development of inflammatory bowel disease. Physiology (Bethesda) 2018;33:360–369.

19. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 2016;354.

20. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–1303.

21. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res 2016;44:e108.

22. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568–576.

23. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–2209.

24. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv 2012:1207; 3907.

25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.

26. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. Bioinformatics 2015;31:2202–2204.

27. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17:122.

28. Pedersen BS, Layer RM, Quinlan AR. Vcfanno: fast, flexible annotation of genetic variants. Genome Biol 2016;17:118.

29. Paila U, Chapman BA, Kirchner R, et al. GEMINI: integrative exploration of genetic variation and genome annotations. PLoS Comput Biol 2013;9:e1003153.

30. Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics 2017; 33:2037–2039.

31. 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.

32. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005;19(Suppl A):5A–36A.

33. Rentzsch P, Witten D, Cooper GM, et al. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019;47:D886–D894.

34. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–424.

35. Ensari A, Kelsen J, Russo P. Newcomers in paediatric GI pathology: childhood enteropathies including very early onset monogenic IBD. Virchows Arch 2018;472:111–123.

36. **Avitzur Y, Guo C,** Mastropaolo LA, et al. Mutations in tetratricopeptide repeat domain 7A result in a severe form of very early onset inflammatory bowel disease. Gastroenterology 2014;146:1028–1039.

37. Miot C, Imai K, Imai C, et al. Hematopoietic stem cell transplantation in 29 patients hemizygous for

hypomorphic IKBKG/NEMO mutations. Blood 2017; 130:1456–1467.

38. Kammermeier J, Drury S, James CT, et al. Targeted gene panel sequencing in children with very early onset inflammatory bowel disease–evaluation and prospective analysis. J Med Genet 2014;51:748–755.

39. Kelsen JR, Dawany N, Moran CJ, et al. Exome sequencing analysis reveals variants in primary immunodeficiency genes in patients with very early onset inflammatory bowel disease. Gastroenterology 2015; 149:1415–1424.

40. Ostrowski J, Paziewska A, Lazowska I, et al. Genetic architecture differences between pediatric and adult-onset inflammatory bowel diseases in the Polish population. Sci Rep 2016;6:39831.

41. Xiao Y, Wang XQ, Yu Y, et al. Comprehensive mutation screening for 10 genes in Chinese patients suffering very early onset inflammatory bowel disease. World J Gastroenterol 2016;22:5578–5588.

42. Petersen BS, August D, Abt R, et al. Targeted gene panel sequencing for early-onset inflammatory bowel disease and chronic diarrhea. Inflamm Bowel Dis 2017;23:2109–2120.

43. Charbit-Henrion F, Parlato M, Hanein S, et al. Diagnostic yield of next-generation sequencing in very early-onset inflammatory bowel diseases: a multicenter study. J Crohns Colitis 2018;12:1104–1112.

44. Lega S, Pin A, Arrigo S, et al. Diagnostic approach to monogenic inflammatory bowel disease in clinical practice: a ten-year multicentric experience. Inflamm Bowel Dis 2019 Aug 3 [Epub ahead of print].

45. Kammermeier J, Dziubak R, Pescarin M, et al. Phenotypic and genotypic characterisation of inflammatory bowel disease presenting before the age of 2 years. J Crohns Colitis 2017;11:60–69.

46. Ashton JJ, Andreoletti G, Coelho T, et al. Identification of variants in genes associated with single-gene inflammatory bowel disease by whole-exome sequencing. Inflamm Bowel Dis 2016;22:2317–2327.

47. Batura V, Muise AM. Very early onset IBD: novel genetic aetiologies. Curr Opin Allergy Clin Immunol 2018; 18:470–480.

48. Turner D, Muise AM. Very early onset IBD: how very different 'on average'? J Crohns Colitis 2017;11:517–518.

49. **Ouahed J, Spencer E,** Kotlarz D, et al. Very early onset inflammatory bowel disease: a clinical approach with a focus on the role of genetics and underlying immune deficiencies. Inflamm Bowel Dis 2019 Dec 3 [Epub ahead of print].

50. Huang C, De Ravin SS, Paul AR, et al. Genetic risk for inflammatory bowel disease is a determinant of Crohn's disease development in chronic granulomatous disease. Inflamm Bowel Dis 2016;22:2794–2801.

51. Jardine S, Anderson S, Babcock S, et al. Drug screen identifies leflunomide for treatment of inflammatory bowel diseases caused by TTC7A deficiency. Gastroenterology 2020;158:1000–1015.

52. Biggs CM, Keles S, Chatila TA. DOCK8 deficiency: Insights into pathophysiology, clinical features and management. Clin Immunol 2017;181:75–82.

53. Murugan D, Albert MH, Langemeier J, et al. Very early onset inflammatory bowel disease associated with aberrant trafficking of IL-10R1 and cure by T cell replete haploidentical bone marrow transplantation. J Clin Immunol 2014;34:331–339.

54. Bennett CL, Christie J, Ramsdell F, et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat Genet 2001;27:20–21.

55. Zeissig Y, Petersen BS, Milutinovic S, et al. XIAP variants in male Crohn's disease. Gut 2015;64:66–76.

56. Jardine S, Dhingani N, Muise AM. TTC7A: steward of intestinal health. Cell Mol Gastroenterol Hepatol 2019; 7:555–570.

57. **Elkadri A, Thoeni C,** Deharvengt SJ, et al. Mutations in plasmalemma vesicle associated protein result in sieving protein-losing enteropathy characterized by hypoproteinemia, hypoalbuminemia, and hypertriglyceridemia. Cell Mol Gastroenterol Hepatol 2015; 1:381–394 e7.

58. **Glocker EO, Kotlarz D,** Boztug K, et al. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. N Engl J Med 2009;361:2033–2045.

59. van de Geer A, Nieto-Patlan A, Kuhns DB, et al. Inherited p40phox deficiency differs from classic chronic granulomatous disease. J Clin Invest 2018;128:3957–3975.

60. Schwerd T, Pandey S, Yang HT, et al. Impaired antibacterial autophagy links granulomatous intestinal inflammation in Niemann-Pick disease type C1 and XIAP deficiency with NOD2 variants in Crohn's disease. Gut 2017;66:1060–1073.

61. Kahr WH, Pluthero FG, Elkadri A, et al. Loss of the Arp2/3 complex component ARPC1B causes platelet abnormalities and predisposes to inflammatory disease. Nat Commun 2017;8:14816.

**Correspondence**
Address correspondence to: Aleixo M. Muise, MD, PhD, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada, M5G 1X8. e-mail: aleixo.muise@utoronto.ca; fax: 416-813-6531.