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
Western University, ecrowle6@uwo.ca

Neil Warner

Jie Pan

Sam Khalouei

Abdul Elkadri

See next page for additional authors

Follow this and additional works at: https://ir.lib.uwo.ca/paedpub

Part of the Pediatrics Commons

Citation of this paper:
Crowley, Eileen; Warner, Neil; Pan, Jie; Khalouei, Sam; Elkadri, Abdul; Fiedler, Karoline; Foong, Justin; Turinsky, Andrei L; Bronte-Tinkew, Dana; and Zhang, Shiqi, "Prevalence and Clinical Features of Inflammatory Bowel Diseases Associated With Monogenic Variants, Identified by Whole-Exome Sequencing in 1000 Children at a Single Center" (2020). Paediatrics Publications. 390.
https://ir.lib.uwo.ca/paedpub/390
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,9 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. Muise1,18,19

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

See Covering the Cover synopsis on page 2017.

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; \( P = 0.020 \)), family history of autoimmune disease (OR, 5.12; \( P = 0.002 \)), extra-intestinal manifestations (OR, 15.36; \( P < 0.001 \)), and surgery (OR, 3.42; \( P = 0.042 \)). 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.

**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.

**Patient Population**

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

### Methods

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

**Abbreviations used in this paper:** CD, Crohn’s disease; GATK, Genome Analysis Toolkit; HSCT, hematopoietic stem cell transplantation; IBD, inflammatory bowel disease; IBD-U, inflammatory bowel disease-unclassified; UC, ulcerative colitis; VEOIBD, very early onset inflammatory bowel disease; WES, whole-exome sequencing.

**Most current article**

© 2020 by the AGA Institute. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[0016-5085](https://doi.org/10.1053/j.gastro.2020.02.023)
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 µg/mL, and approximately 2 µg 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, miRNAAs, 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 HaploypeCaller20 (version 3.5), Vardict21 (version 1.4.6), Varscan22

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

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

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.
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. This was an update on the previously published Montreal Classification of IBD. 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
distribution of the patients in Figure 1B, Supplementary Table 4). Principal component analysis of the cohort demonstrated broad ethnic diversity including European, East Asian, or South Asian ethnicity (Figure 1C, Supplementary Figure 2, and Supplementary Table 5E).

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) 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. 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 2A). As shown in Figure 2B, 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

<table>
<thead>
<tr>
<th>Biological category</th>
<th>Patient</th>
<th>Sex</th>
<th>Age at diagnosis</th>
<th>Chrom</th>
<th>Gen</th>
<th>Inheritance model</th>
<th>aa mutation</th>
<th>Impact</th>
<th>ExAC maf</th>
<th>CADD 1-3 Phred score</th>
<th>Family history</th>
<th>Reported Inheritance</th>
<th>Causal evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epithelial barrier &amp; response defects</td>
<td>1</td>
<td>M</td>
<td>15</td>
<td>chr2</td>
<td>ALPI</td>
<td>AR a</td>
<td>A360V</td>
<td>missense</td>
<td>0.00006363</td>
<td>26.8</td>
<td>+</td>
<td>AR</td>
<td>G, F, P</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>F</td>
<td>11.91</td>
<td>chr3</td>
<td>COL7A1</td>
<td>AR a</td>
<td>R1696C</td>
<td>missense</td>
<td>0.00003</td>
<td>28.7</td>
<td>-</td>
<td>AR/AD</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>M</td>
<td>12.3</td>
<td>chr12</td>
<td>GUCY2C</td>
<td>AD</td>
<td>G549S</td>
<td>missense</td>
<td>0.000045</td>
<td>33</td>
<td>+</td>
<td>AD</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>F</td>
<td>15.1</td>
<td>chr12</td>
<td>GUCY2C</td>
<td>AD</td>
<td>F525L</td>
<td>missense</td>
<td>0.000067</td>
<td>25.5</td>
<td>-</td>
<td>AD</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>M</td>
<td>11.38</td>
<td>chr3</td>
<td>SLCO2A1</td>
<td>AR a</td>
<td>R314W</td>
<td>missense</td>
<td>0.00000164</td>
<td>26.2</td>
<td>-</td>
<td>AR</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>F</td>
<td>0</td>
<td>chr2</td>
<td>TTC7A</td>
<td>AR a</td>
<td>E71K</td>
<td>missense</td>
<td>0.00000082</td>
<td>34</td>
<td>-</td>
<td>AR</td>
<td>G, F, P</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>M</td>
<td>2.75</td>
<td>chr7</td>
<td>ARPC1B</td>
<td>AR b</td>
<td>G526X</td>
<td>stop gained</td>
<td>0</td>
<td>36</td>
<td>-</td>
<td>AR</td>
<td>P</td>
</tr>
<tr>
<td>T and B cell differentiation defects</td>
<td>8*</td>
<td>M</td>
<td>3.5</td>
<td>chr7</td>
<td>ARPC1B</td>
<td>AR b</td>
<td>V91WfsX121</td>
<td>frameshift</td>
<td>0</td>
<td>none</td>
<td>-</td>
<td>AR</td>
<td>G, F, P</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>M</td>
<td>15.52</td>
<td>chrX</td>
<td>GUCY2C</td>
<td>AD</td>
<td>N484K</td>
<td>missense</td>
<td>0</td>
<td>23</td>
<td>-</td>
<td>XL</td>
<td>F, P</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>M</td>
<td>5.5</td>
<td>chrX</td>
<td>GUCY2C</td>
<td>AD</td>
<td>T49M</td>
<td>missense</td>
<td>0</td>
<td>26</td>
<td>-</td>
<td>XL</td>
<td>G, F, P</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>F</td>
<td>12.38</td>
<td>chr9</td>
<td>DOCK8</td>
<td>AR b</td>
<td>S97L</td>
<td>missense</td>
<td>0</td>
<td>35</td>
<td>-</td>
<td>AR/AD</td>
<td>G, P</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>F</td>
<td>14.73</td>
<td>chr9</td>
<td>DOCK8</td>
<td>AR b</td>
<td>R947C</td>
<td>missense</td>
<td>0</td>
<td>26.4</td>
<td>(milder)</td>
<td>AR/AD</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>F</td>
<td>14.65</td>
<td>chr9</td>
<td>DOCK8</td>
<td>AR b</td>
<td>R59H</td>
<td>missense</td>
<td>0</td>
<td>26.7</td>
<td>(milder)</td>
<td>AR/AD</td>
<td>P</td>
</tr>
<tr>
<td>Hyper- &amp; Auto-inflammatory disorders</td>
<td>14</td>
<td>M</td>
<td>1.63</td>
<td>chr4</td>
<td>LRBA</td>
<td>AR b</td>
<td>Y71C</td>
<td>missense</td>
<td>0</td>
<td>22.1</td>
<td>(milder)</td>
<td>AR</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>15**</td>
<td>M</td>
<td>9.33</td>
<td>chr4</td>
<td>LRBA</td>
<td>AR*</td>
<td>E1916X</td>
<td>missense</td>
<td>0</td>
<td>None</td>
<td>-</td>
<td>AR</td>
<td>G, F, P</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>F</td>
<td>2.5</td>
<td>chr2</td>
<td>STAT1</td>
<td>AD</td>
<td>V389A</td>
<td>missense</td>
<td>0</td>
<td>19.69</td>
<td>-</td>
<td>AD</td>
<td>G, F, P</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>M</td>
<td>14</td>
<td>chr10</td>
<td>HPS1</td>
<td>AR b</td>
<td>E9D</td>
<td>missense</td>
<td>0.000016</td>
<td>20.4</td>
<td>+</td>
<td>AR</td>
<td>p</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>M</td>
<td>12.75</td>
<td>chr1</td>
<td>PIK3CD</td>
<td>AD</td>
<td>V616A</td>
<td>missense</td>
<td>0.0000003</td>
<td>23.4</td>
<td>+</td>
<td>AD</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>19**</td>
<td>M</td>
<td>8.4</td>
<td>chrX</td>
<td>SH2D1A</td>
<td>AD</td>
<td>M1T</td>
<td>start Met loss</td>
<td>0</td>
<td>23.4</td>
<td>-</td>
<td>XL</td>
<td>G, F, P</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>M</td>
<td>11.86</td>
<td>chrX</td>
<td>XIAP</td>
<td>AD</td>
<td>R29K</td>
<td>missense</td>
<td>0</td>
<td>25.7</td>
<td>+</td>
<td>XL</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>M</td>
<td>9.1</td>
<td>chrX</td>
<td>XIAP</td>
<td>AD</td>
<td>S278K</td>
<td>stop gained</td>
<td>0</td>
<td>37</td>
<td>+</td>
<td>XL</td>
<td>G, P</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>M</td>
<td>10.83</td>
<td>chrX</td>
<td>XIAP</td>
<td>AD</td>
<td>A216T</td>
<td>missense</td>
<td>0</td>
<td>25.6</td>
<td>+</td>
<td>XL</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>M</td>
<td>10.67</td>
<td>chrX</td>
<td>XIAP</td>
<td>AD</td>
<td>ΔE349</td>
<td>inframe deletion</td>
<td>0.0004</td>
<td>None</td>
<td>-</td>
<td>XL</td>
<td>G, P</td>
</tr>
<tr>
<td></td>
<td>24**</td>
<td>M</td>
<td>0</td>
<td>chrX</td>
<td>XIAP</td>
<td>AD</td>
<td>V298fsX306</td>
<td>frameshift</td>
<td>0</td>
<td>None</td>
<td>-</td>
<td>XL</td>
<td>G, F, P</td>
</tr>
<tr>
<td></td>
<td>25**</td>
<td>M</td>
<td>3.4</td>
<td>chrX</td>
<td>CYBB</td>
<td>AD</td>
<td>A84A</td>
<td>splice</td>
<td>0.000012</td>
<td>19.34</td>
<td>-</td>
<td>XL</td>
<td>G, F, P</td>
</tr>
<tr>
<td>Regulatory T cells and immune regulation</td>
<td>26</td>
<td>M</td>
<td>12</td>
<td>chr2</td>
<td>CTLA4</td>
<td>AD</td>
<td>L198P</td>
<td>missense</td>
<td>0</td>
<td>24.4</td>
<td>+</td>
<td>XL</td>
<td>F, P</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>M</td>
<td>3.05</td>
<td>chrX</td>
<td>FOXP3</td>
<td>AD</td>
<td>H121Y</td>
<td>missense</td>
<td>0.0000782</td>
<td>21</td>
<td>+</td>
<td>XL</td>
<td>F, P</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>M</td>
<td>12.68</td>
<td>chrX</td>
<td>FOXP3</td>
<td>AD</td>
<td>L198P</td>
<td>missense</td>
<td>0</td>
<td>24.4</td>
<td>+</td>
<td>XL</td>
<td>F, P</td>
</tr>
<tr>
<td></td>
<td>29**</td>
<td>F</td>
<td>0.17</td>
<td>chr2</td>
<td>IL10RB</td>
<td>AD</td>
<td>W3X</td>
<td>stop gained</td>
<td>0</td>
<td>27.6</td>
<td>+</td>
<td>AR</td>
<td>G, F, P</td>
</tr>
<tr>
<td>Others</td>
<td>30</td>
<td>F</td>
<td>4.42</td>
<td>chr2</td>
<td>HSPA4L</td>
<td>AR*</td>
<td>K495fsX13</td>
<td>frameshift</td>
<td>0.0003</td>
<td>None</td>
<td>-</td>
<td>G, P</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>M</td>
<td>12.26</td>
<td>chr1</td>
<td>MASP2</td>
<td>AR b</td>
<td>D244N</td>
<td>missense</td>
<td>0.000068</td>
<td>28.7</td>
<td>+</td>
<td>AR</td>
<td>N/A</td>
</tr>
</tbody>
</table>

+, 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.
Functional Validation

While in silico pathogenicity of variants identified using WES was based on combined annotation-dependent depletion score, inheritance pattern, impact, frequency, and the American College of Medical Genetics and Genomics classification, 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 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. The majority of patients had multiple levels of support, including 16 patients with genetic support of
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 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 2C 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 2D). 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 2E 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, and/or other genetic mechanisms. Therefore, this study expands the genetic landscape of monogenic IBD, which may have diagnostic, therapeutic, and public health implications.
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 636) may not respond to HSCT or biologic therapies and may present specific therapeutic challenges.77

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, multiethnic, 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 phenotypes15–18,47–49 (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 al50 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 SLC02A1 in our cohort. Variants in these genes may cause perturbations in the epithelial barrier leading to
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 8F, 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 IBD55 or palliation in patients with severe forms of TTCTA-deficiency56 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 stimulation58 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 TTCTA,56 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; 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 covered12 (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 genes59 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](http://www.gastrojournal.org), and at [https://doi.org/10.1053/j.gastro.2020.02.023](https://doi.org/10.1053/j.gastro.2020.02.023).

**References**


Author names in bold designate shared co-first authorship.

Received July 9, 2019. Accepted February 6, 2020.

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

Acknowledgments
The authors thank all of the SickKids patients and their families who have consented and participated in this study and the health care professionals at The Hospital for Sick Children who care for these IBD patients. Research Ethics Board 1000024905. Jie Pan and Sam Khaleouei contributed equally to this work.

CRediT Authorship Contributions
Eileen Crowley, MD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Writing – original draft: Lead). Neil Warner, PhD (Data curation: Supporting; Formal analysis: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting). Jie Pan, MD (Formal analysis: Supporting; Writing – review & editing: Supporting). Sam Khaleouei, PhD (Formal analysis: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting). Abdul Elkadri, MD (Conceptualization: Supporting; Data curation: Supporting; Formal analysis: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting). Justin Foong, PhD (Formal analysis: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting). Andrei Turinsky, PhD (Formal analysis: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting). Dana Bronte-Tinkew, MD (Data curation: Supporting; Writing – review & editing: Supporting). Shiqi Zhang, BSc (Data curation: Supporting; Writing – review & editing: Supporting).
Writing – review & editing: Supporting). Jamie Hu, BSc (Data curation: Supporting; Writing – review & editing: Supporting). David Tian, BSc (Data curation: Supporting; Writing – review & editing: Supporting). Dalin Li, PhD (Data curation; Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Julie Horowitz, PhD (Data curation: Supporting; Writing – review & editing: Supporting). Iram Siddiqui, MD (Data curation: Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Dalin Li, PhD (Data curation: Supporting; Writing – review & editing: Supporting). David Tian, BSc (Data curation: Supporting; Writing – review & editing: Supporting). Dalin Li, PhD (Data curation: Supporting; Writing – review & editing: Supporting). Chaim Roifman, MD PhD (Data curation: Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Peter Church, MD (Conceptualization: Supporting; Data curation: Supporting; Writing – review & editing: Supporting). Andrew Paterson, PhD (Data curation: Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Christoph Klein, MD PhD (Data curation: Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Holm Uhlig, MD PhD (Data curation: Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Scott Snapper, MD PhD (Data curation: Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Daniel Kotlarz, MD PhD (Data curation: Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Michael Brudno, PhD (Data curation: Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Thomas Walters, MD (Conceptualization: Supporting; Data curation: Supporting; Formal analysis: Supporting; Writing – review & editing: Supporting). Anne Griffiths, MD (Formal analysis: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting). Alexo M Muse, MD PhD (Conceptualization: Supporting; Data curation: Supporting; Formal analysis: Supporting; Funding acquisition: Supporting; Supervision: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting).

Conflicts of interest
These authors disclose the following: Julie Horowitz and Claudia Gonzaga-Jauregui are full-time employees of the Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc and receive stock options as part of compensation. The remaining authors disclose no conflicts.

Funding
EC was supported by a Canadian Institutes of Health Research (CIHR)/Canadian Association of Gastroenterology Fellowship. AMM is funded by a Canada Research Chair (Tier 1) in Pediatric IBD, CIHR Foundation Grant, and National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) (RC2DK118640) grant. AMM, SBS, CK, DK, HU, and DMG are supported by the Leona M. and Harry B. Helmsley Charitable Trust. CK and DK are supported by the Collaborative Research Consortium SFB1054 project A05. Bioinformatics analyses were supported in part by the Canadian Centre for Computational Genomics, part of the Genome Technology Platform funded by Genome Canada through Genome Quebec and Ontario Genomics. DMG is also supported by National Institutes of Health/NIDDK grants P01 DK046763, and U01 DK062413, and The Litwin Foundation.