# Genetic Sequencing of Pediatric Patients Identifies Mutations in Monogenic Inflammatory Bowel Disease Genes that Translate to Distinct Clinical Phenotypes

James J. Ashton, MRCPCH[1,2], Enrico Mossotto, PhD[1,3], Imogen S. Stafford, MSci[1], Rachel Haggarty, RN[3], Tracy A.F. Coelho, PhD[2], Akshay Batra, MD[2], Nadeem A. Afzal, MD[2], Matthew Mort, PhD[4], David Bunyan, PhD[5], Robert Mark Beattie, FRCPCH[2] and Sarah Ennis, PhD[1]

| | |
|---|---|
| OBJECTIVES: | **Monogenic inflammatory bowel disease (IBD) comprises rare Mendelian causes of gut inflammation, often presenting in infants with severe and atypical disease. This study aimed to identify clinically relevant variants within 68 monogenic IBD genes in an unselected pediatric IBD cohort.** |
| METHODS: | **Whole exome sequencing was performed on patients with pediatric-onset disease. Variants fulfilling the American College of Medical Genetics criteria as "pathogenic" or "likely pathogenic" were assessed against phenotype at diagnosis and follow-up. Individual patient variants were assessed and processed to generate a per-gene, per-individual, deleteriousness score.** |
| RESULTS: | **Four hundred one patients were included, and the median age of disease-onset was 11.92 years. In total, 11.5% of patients harbored a monogenic variant. *TRIM22-related* disease was implicated in 5 patients. A pathogenic mutation in the Wiskott-Aldrich syndrome (*WAS*) gene was confirmed in 2 male children with severe pancolonic inflammation and primary sclerosing cholangitis. In total, 7.3% of patients with Crohn's disease had apparent autosomal recessive, monogenic *NOD2*-related disease. Compared with non-*NOD2* Crohn's disease, these patients had a marked stricturing phenotype (odds ratio 11.52, significant after correction for disease location) and had undergone significantly more intestinal resections (odds ratio 10.75). Variants in *ADA*, *FERMT1*, and *LRBA* did not meet the criteria for monogenic disease in any patients; however, case-control analysis of mutation burden significantly implicated these genes in disease etiology.** |
| DISCUSSION: | **Routine whole exome sequencing in pediatric patients with IBD results in a precise molecular diagnosis for a subset of patients with IBD, providing the opportunity to personalize therapy. *NOD2* status informs risk of stricturing disease requiring surgery, allowing clinicians to direct prognosis and intervention.** |

## INTRODUCTION

Inflammatory bowel disease (IBD) is a chronic, relapsing, and remitting disease characterized by intestinal inflammation. Most patients with IBD harbor an underlying genetic risk affected by environmental factors, including the microbiome (1). To date, in excess of 230 genes have been associated with IBD, mostly through genome-wide association studies (GWAS) (2,3). The first locus implicated in the risk of developing disease was on chromosome 16 and was mapped to *NOD2* in the early 2000s (4–6). There are limited data implicating homozygote and compound heterozygote *NOD2* variants as disease-causing in an autosomal recessive (AR) inheritance pattern (7,8). The success of prospective projects based on microbiome and RNA sequencing data, such as PROTECT, has brought into focus the need for improving predictive algorithms by also utilizing precise genetic diagnoses (9,10).

[1]Department of Human Genetics and Genomic Medicine, University of Southampton, Southampton, UK; [2]Department of Paediatric Gastroenterology, Southampton Children's Hospital, Southampton, UK; [3]NIHR Southampton Biomedical Research Centre, University Hospital Southampton, Southampton, UK; [4]Human Genetic Mutation Database, Cardiff University, Cardiff, UK; [5]Wessex Regional Genetics Laboratory, Salisbury NHS Foundation Trust, Salisbury, UK. **Correspondence:** Sarah Ennis. E-mail: s.ennis@southampton.ac.uk.

High throughput next generation sequencing (NGS) technologies are powerful for detection of genetic conditions. NGS is already being routinely exploited in mainstream diagnostics of rare disease to substantial patient benefit (11). As yet, NGS technology has seen little clinical implementation in complex diseases such as IBD (12). However, it has aided discovery of IBD risk genes and identified precise causative variants, alongside informing genotype-phenotype correlations (13–15). Molecular diagnosis using NGS relies on accurate clinical phenotyping and functional assessment of mutations. *De novo* and homozygous recessive inheritance is most easily detected, but detection of compound heterozygosity (supplementary figure 1, http://links.lww.com/CTG/A174) is more difficult (16).

The vanguard of NGS application in IBD is in the identification of a rare subset of conditions that are Mendelian disorders, masquerading as IBD (17,18). These are a group of diseases (currently underpinned by variation in 68 genes) typically detected in very early onset IBD (VEOIBD) with severe and atypical features (17,19). Monogenic forms of IBD are often the manifestation of an underlying immune deficiency or epithelial barrier dysfunction and have specific management considerations (19).

Personalizing medication from the increasing array of targets that now include the JAK-STAT pathway (tofacitinib), IL12/IL23 signaling (ustekinumab), and anti-integrins (α4β7, vedolizumab), alongside anti-TNF and immunomodulators, must move to target the specific patient's underlying cause for disease (20).

This study aimed to apply exome sequencing to a cohort of typical pediatric patients with IBD to identify clinically relevant variants within monogenic IBD genes, using standard guidelines, and correlate with patient phenotype. Furthermore, we apply a novel *per-gene* deleteriousness score to assess the contribution of monogenic variation to disease phenotype.

## METHODS

Patients were recruited from the Wessex regional pediatric IBD service at Southampton Children's Hospital to the genetics of pediatric IBD study (2010 to present). The eligibility criteria for recruitment was a confirmed histological diagnosis of either Crohn's disease (CD), ulcerative colitis (UC), or IBD unclassified (IBDU), in line with the Porto criteria, and age less than 18 years (21,22).

### DNA extraction

Patient DNA was extracted from peripheral venous blood samples collected in EDTA using the salting-out method, or from saliva, as previously described (23).

### Whole exome sequencing data processing

Raw FASTQ sequencing data from patients with pediatric-onset IBD were processed using our in-house pipeline (24). verifyBamID was used to check the presence of DNA contamination across the cohort (25). Alignment was performed against the human reference genome (hg19 assembly) using BWA-MEM (26) (version 0.7.12). Aligned BAM files were sorted and duplicate reads were marked using Picard tools (version 1.97). Following GATK v3.7 (27) best practice recommendations (28), variants were called using GATK HaplotypeCaller to produce a gVCF file for each sample and later jointly genotyped.

Annotation of this composite file applied ANNOVAR v2016Feb01 using default databases RefSeq gene transcripts (RefGene), deleteriousness scores databases (dbnsfp33a, CADD 1.3, and DANN), dbSNP147, and the human genetic mutation database (HGMD Pro 2018) flat file (29). Variant allele frequencies were sourced through the genome aggregation database (gnomAD) (30). HaplotypeCaller default settings were used corresponding to variants with a minimum Phred base quality score of 20 being called.

### Monogenic IBD gene list

A list of 68 genes previously implicated in monogenic IBD was established, supplementary table 1, http://links.lww.com/CTG/A175. This list combined genes reported by Uhlig et al. (2014) (n = 50), Uhlig and Muise (2017) (n = 15), Girardelli et al. (2018) (n = 1), and through direct correspondence with the International Early-Onset Paediatric IBD Cohort Study consortium (NEOPICS) (n = 2) (18,19,31). The reported inheritance pattern for each monogenic disease gene was determined as either autosomal dominant (AD), AR, or X-linked (XL). The *NOD2* inheritance pattern was treated as AR (7,31,32).

### Variant filtering

A total of 1,405 high-quality variants (Phred >20) were called across the 68 monogenic IBD genes in 401 pediatric patients with IBD. We applied a crude preliminary filter to exclude variants with no previous evidence for causality in publicly available databases (HGMD Pro 2018 and ClinVar 2018) or those that are common and have *in silico* evidence of being benign (29,33). Variants with the following annotation in HGMD and/or ClinVar were retained for further investigation:

1. HGMD—Disease-associated polymorphism with supporting functional evidence—*DFP* or disease-causing mutation-*DM* or probable/possible pathological mutation—*DM?*
2. ClinVar—Pathogenic, Likely Pathogenic

Any variants fulfilling these criteria were scrutinized to confirm their pathogenic status was in the context of IBD or monogenic disorders with bowel inflammation, whereas variants achieving pathogenic status because of an unrelated clinical phenotype were excluded.

Because HGMD and ClinVar fail to annotate a subset of variants, a second filtering strategy was applied. Variants without any HGMD or ClinVar annotation were retained based on the following criteria: (i) Coding context—(ExonicFunc.knownGene) "Exonic" or "Splicing" AND; (ii) CADD Phred score >20 AND; (iii) gnomAD "all genomes" frequency <0.01 or Novel.

Variants withstanding the filtering strategies above were only retained if they were inherited in the correct zygosity to be disease causing. For genes reported as AD, heterozygous variants were retained; for genes reported as AR, homozygous variants were retained; and for genes reported as XL, hemizygous men (one allele on the X chromosome in males) or homozygous females were retained. Patients harboring 2 or more different variants within the same gene throughout were tested for compound heterozygosity using Sanger sequencing of the proband and parental DNA. Confirmed compound heterozygous variants were retained (see supplementary dataset 2, http://links.lww.com/CTG/A173).

This substantially reduced the number of patients and variants that warranted close scrutiny for consistency with the American College of Medical Genetics (ACMG) guidelines (34). An independent literature review was conducted to collate validated functional evidence for all 35 variants (supplementary table 2, http://links.lww.com/CTG/A176). Validated functional evidence was defined as one or more report(s) describing reduced/absent

protein function including impact on downstream signaling/protein expression, nonsense-mediated decay, or deletions. These data were used to annotate each variant according to the ACMG criteria for pathogenicity. Each patient underwent final classification to determine if their variant profile fulfilled the criteria for "pathogenic" or "likely pathogenic" according to the ACMG rules for combing criteria to classify sequence variants (34).

### Phenotypic characterization
In depth, longitudinal, clinical phenotyping was extracted for all patients in the cohort including diagnostic and follow-up information (Supplementary data 1, http://links.lww.com/CTG/A172). Phenotypic characteristics were transformed to binary or continuous data for use in regression analyses. The follow-up duration was calculated for each patient based on the date of diagnosis and last recorded clinical contact.

### Application of GenePy *in silico* score
GenePy is a novel software that combines the effect of multiple variants occurring within any given set of genes into *a single score for each gene for each individual* (24). By scoring whole gene pathogenicity within an individual, GenePy allows for interpretation on a patient-by-patient basis. GenePy incorporates biological information on variant deleteriousness (using an *in silico* predictor such as DANN (35)), population frequency, and observed zygosity for each variant. All variants meeting minimum genotyping quality (>20) were retained for GenePy using VCFtools. Because GenePy scores can be applied in a case-control comparison within ethnic subgroups, *peddy* software was used to infer relatedness and ethnicity (with a probability >90%) for all patients with IBD (36). A cohort of 173 non-IBD Caucasian ethnicity individuals (EUCLIDS consortium) for whom whole exome sequencing (WES) data were available were used as controls.

GenePy scores are quantitative values that follow a *Poisson* distribution, whereby for any given gene, most patients have a score close to zero and high scores are rare. It is expected that most patients will have scores in the same range because controls with a small subset of patients incur high scores. It is possible to assess evidence for gene causality by selecting the most extreme scores in right tail of the GenePy distributions in cases and compare with the same proportion in controls using a one-tailed Mann-Whitney $U$ test.

Statistical significance was corrected for multiple testing using the false discovery rate. Enrichment for a diagnosis of either CD or UC was assessed using the Fisher exact test. Forward stepwise linear regression was performed using R (v3.6.0) and SPSS (v24, IBM) software.

### Ethical approval and patient involvement
The study has ethics approval from Southampton and South West Hampshire Research Ethics Committee (09/H0504/125). Patients and families are involved in guiding research strategy through local research events and in dissemination of results through our research website.

## RESULTS
Four hundred one patients were included in the analysis. Mean age at diagnosis was 11.92 years (range 1.3–17.39), 40.9% were women, and 64.8% had a diagnosis of CD. Children diagnosed before the age of 6 years (VEOIBD) accounted for 7.5% of patients (n = 30), and a further 17.2% (n = 69) were diagnosed with early onset IBD

(EOIBD), aged 6 or older and less than 10 years. The remaining 75.3% patients (n = 302) were diagnosed between the age of 10 and 18 years and designated pediatric-onset IBD (POIBD) (Table 1). The median follow-up time for the entire cohort was 4.6 years (range 0.15–17.7).

### ACMG "pathogenic" or "likely pathogenic" monogenic IBD gene variants
Initial filtering excluded 1,345 variants across 312 patients. Subsequent variant confirmation by zygosity and Sanger sequencing excluded 27 patients, and application of the ACMG guidelines excluded a further 16 patients (Figure 1). Twenty-nine variants fulfilled the ACMG standards to be classified as "pathogenic" or "likely pathogenic" across 46 patients (11.5% of the cohort) and are discussed in detail below (Table 2).

"Pathogenic" or "likely pathogenic" variants were observed in 16.7% of patients with VEOIBD, 11.6% of patients with EOIBD, and in 10.9% of those with POIBD (Supplementary tables 3–5, http://links.lww.com/CTG/A177, http://links.lww.com/CTG/A178, http://links.lww.com/CTG/A179 provide variant annotation). Recurrent variants were observed in *NOD2* (20 patients), *TRIM22* (5 patients), *CD40LG* (5 patients), *WAS* (4 patients), *NCF2* (3 patients), *STAT1* (3 patients), *DKC1* (2 patients), and *DCLRE1C* (2 patients). One patient was identified with variant(s) in each of *XIAP*, *NCF1*, and *MASP2*. A single patient harbored a hemizygous variant in each of *WAS* and *STAT1* genes.

Twenty-three patients had "pathogenic" or "likely pathogenic" compound heterozygous variants confirmed through Sanger sequencing in *DCLRE1C*, *NCF2*, *TRIM22*, and *NOD2*. One additional patient (harboring *TRIM22* R317K/R442K) was assumed to be compound heterozygote, but segregation analysis was not possible because of a lack of parental DNA. Without exception, all *potential* compound heterozygote variant pairs within *NOD2*, *TRIM22*, and *DCLRE1C* (representing 16, 3, and 2 patients, respectively) were confirmed after segregation analysis. Conversely, potential compound heterozygote variants in *NCF2* correctly segregated in only 2 of 8 patients, where failure to segregate was consistently because of the *NCF2* P454S variant co-occurring on the same parental haplotype as the H389G variant.

### Phenotypic characteristics of monogenic variants
Phenotypic characteristics of the 46 patients with a "pathogenic" or "likely pathogenic" monogenic IBD variant(s) are detailed in Table 2.

### NOD2 variants—a monogenic stricturing disease phenotype
Twenty patients (5% of all IBD) harbored one or more of 11 variants consistent with an AR pattern of inheritance; 19 of 20 (95%) patients had a diagnosis of CD, representing 7.3% of patients with CD. A novel variant (E963G) predicted to be highly deleterious (CADD 27.3) was observed in a single patient.

In the 19 patients with CD with "pathogenic" or "likely pathogenic" *NOD2* variants, 13 had stricturing disease (68.4%). Stricturing disease behavior was seen in 38 of 240 (15.8%) of the remaining patients with CD, translating to an odds ratio (OR) of 11.52 (relative risk [RR] 4.32) in patients with monogenic *NOD2* CD ($\chi^2 = 30.3$, $P = 2.0 \times 10^{-6}$). To assess whether this stricturing phenotype was solely a function of disease location, we tested the rate of stricturing disease in monogenic *NOD2*-related disease patients with ileal location compared with those non-*NOD2* patients with ileal location. Where approximately 20% of non-

INFLAMMATORY BOWEL DISEASE

**Table 1.** Demographic characterization of patient cohort

|  | All patients (%) | VEOIBD (%) | EOIBD (%) | POIBD (%) |
|---|---|---|---|---|
| No. of patients | 401 | 30 | 69 | 302 |
| Mean age at diagnosis (range) | 11.92 (1.3–17.39 years) | NA | NA | NA |
| Number female (%) | 164 (40.9%) | 12 (40%) | 35 (50.7%) | 117 (38.7%) |
| Crohn's disease (%) | 259 (64.6%) | 13 (43.3%) | 48 (69.6%) | 198 (65.6%) |
| Ulcerative colitis (%) | 125 (31.2%) | 15 (50%) | 17 (24.6%) | 93 (30.8%) |
| IBDU (%) | 17 (4.2%) | 2 (6.7%) | 4 (5.8%) | 11 (3.6%) |

EOIBD, early onset inflammatory bowel disease, ≥6, <10 years; IBD, inflammatory bowel disease; POIBD, pediatric-onset inflammatory bowel disease, ≥10, <18 years; VEOIBD, very early onset inflammatory bowel disease, <6 years.

*NOD2* patients with CD with ileal disease developed strictures, 70% of patients with "pathogenic" or "likely pathogenic" *NOD2* variation, and ileal location were subsequently diagnosed with stricturing disease ($\chi^2 = 20.4$, $P = 6.0 \times 10^{-6}$, Supplementary Table 6, http://links.lww.com/CTG/A180).

Patients with monogenic *NOD2*-related disease were at significantly increased risk of undergoing intestinal resection (right hemicolectomy). Surgical resection had occurred in 12 of 19 (63.2%) monogenic *NOD2* patients with CD compared with 33 of 259 (13.8%) non-*NOD2* patients with CD (OR 10.75, RR 4.59, $\chi^2 = 29.8$, $P = 4.9 \times 10^{-8}$, Supplementary Table 7, http://links.lww.com/CTG/A181).

### TRIM22 variants—severe variable disease phenotype
All 5 patients with "pathogenic" or "likely pathogenic" *TRIM22* variants had at least one copy of the R321K variant, with a single patient harboring this variant in homozygote form. A variable but severe disease phenotype was seen in all 5 patients. Patient #19 had a moderate-severe disease course requiring treatment with anti-TNF monoclonal therapy, but no fistulating or stricturing disease emerged during the follow-up period. Patient #20 was diagnosed with CD at 12 years of age, was treated with monoclonals but developed a stricturing phenotype within 2.5 years of follow-up. Patient #21 was diagnosed at 14 years of age with UC and had a mild disease phenotype requiring 5-ASA and thiopurine treatment. Patient #22 had very early onset of CD, a severe fistulating perianal phenotype requiring multiple surgical procedures, and anti-TNF therapy consistent with the phenotypic spectrum previously reported in 3 *TRIM22* cases (37). Patient #23 was diagnosed with CD at 9 years of age and a severe disease course leading to subtotal colectomy for refractory disease at 11 years of age.

### WAS variant (P460S)—severe ulcerative colitis with liver disease
Of the 4 patients with "pathogenic" or "likely pathogenic" *WAS* alleles, 2 patients (hemizygous for P460S) presented with a markedly distinct phenotype. Patient #6 was diagnosed with severe and extensive UC at 11 years of age, requiring an early colectomy within 2 years and a subsequent diagnosis of primary sclerosing cholangitis (PSC). Intermittent thrombocytopenia and recurrent infections was recorded throughout their disease course consistent with the previously described phenotype for this variant in Wiskott-Aldrich syndrome (38,39). Patient #7 was diagnosed at 14 years of age and also had severe UC refractory to

treatment that required a colectomy. This patient was also diagnosed with PSC.

### Additional monogenic variants
Patients with *XIAP*, *MASP2*, and *NCF2* "pathogenic" or "likely pathogenic" variants had a phenotype largely consistent with previous reports for those genes (40–42), whereas those with *CD40LG*, *DKC1*, *DCLRE1C*, *NCF1*, *and STAT1* variants were more heterogenous in their clinical profile.

### Monogenic genes harbor significantly higher mutation burden in IBD patients
Forty-four patients were of nonwhite ethnicity and excluded from association analyses. GenePy scores were successfully generated for 67 of the 68 monogenic IBD genes, where at least one high-quality missense or insertion/deletion variant was annotated in exonic regions (Supplementary table 1, http://links.lww.com/CTG/A175).

When comparing the top 10% of GenePy scores between IBD and controls, 8 genes accrued significantly higher scores in IBD cases. Following false discovery rate correction, *ADA*, *FERMT1*, *LRBA*, and *NOD2* remained significant (Table 3). Patients identified as having extreme GenePy scores within *NOD2* were significantly enriched for patients with CD (0.0046). No other genes were enriched for either IBD subtype. Of the 20 patients with "pathogenic" or "likely pathogenic" *NOD2* variant(s), 15 were present in the top 10% *NOD2* GenePy scores. We excluded all patients with monogenic *NOD2* variants and recalculated the Mann-Whitney *U* statistics. This confirmed a persistent significant difference ($P = 0.0035$) in cases compared with controls and confirms that those patients failing the threshold to have "pathogenic" or "likely pathogenic" *NOD2* variation harbor a significant excess of pathogenic *NOD2* mutations.

### Phenotypic assessment of patients with extreme GenePy scores
After correction for multiple testing, 4 genes maintained evidence for a significant burden of gene pathogenicity scores (*ADA*, *FERMT1*, *LRBA*, and *NOD2*). Evidence for distinctive phenotypic characteristics conferred by each of these genes was tested using linear regression. GenePy scores for all patients (regardless of IBD subtype) were regressed against clinical features (Supplementary data 1, http://links.lww.com/CTG/A172), and significant associations are detailed in Table 4. Patients with inflated GenePy scores in the *ADA* gene were enriched for men ($P = 0.021$) and presented with isolated colonic disease ($P = 0.033$). No clinical
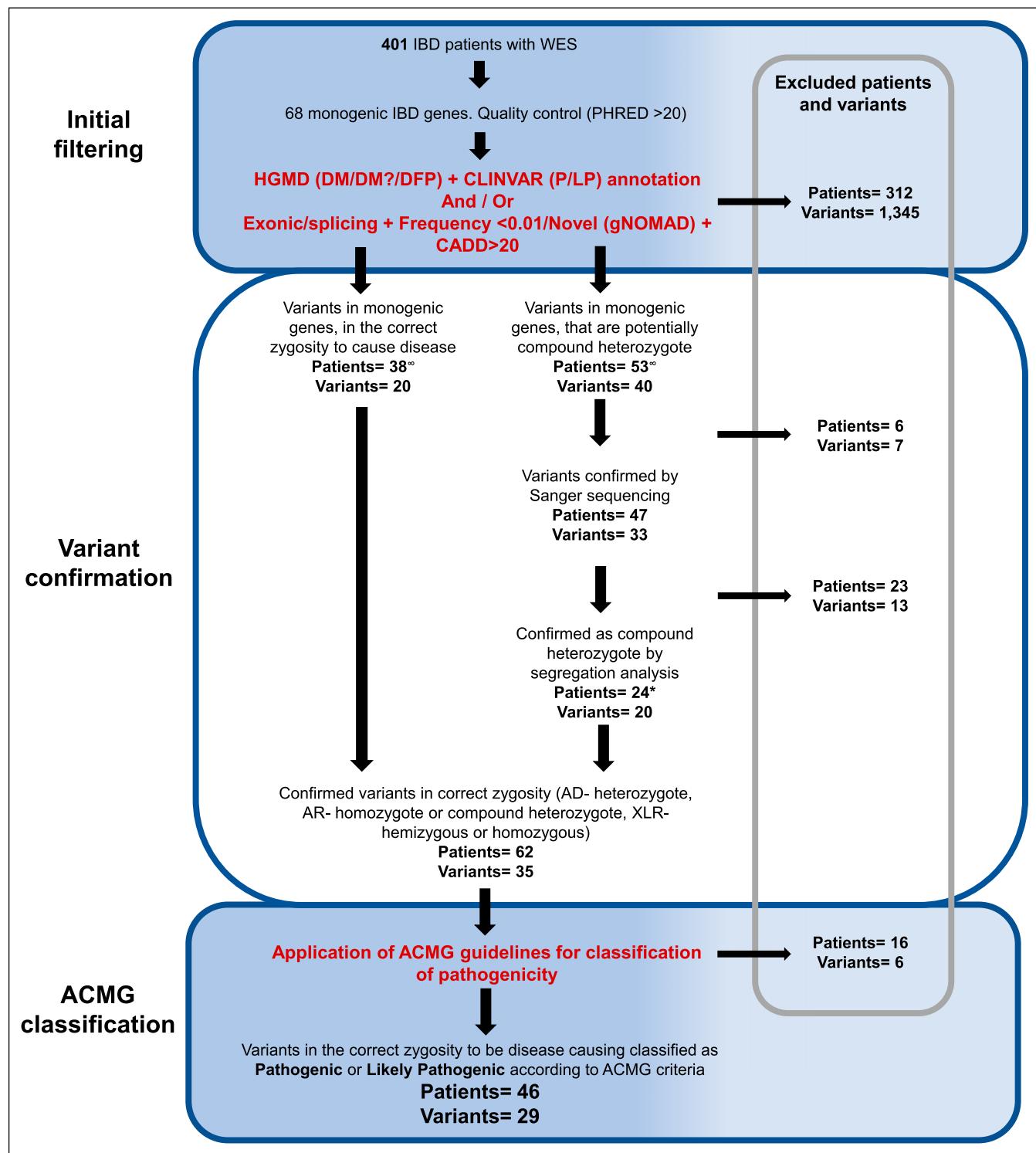
**Figure 1.** Flowchart of variant-filtering detailing variant/patient exclusions at each filtering stage. ∞Two patients appear in both variant confirmation pathways (correct zygosity and potential compound heterozygote) of the flowchart. *Includes one patient (harboring *TRIM22* R317K/R442K) who was assumed to be compound heterozygote but segregation analysis was not possible because of lack of parental DNA.

characteristics were significantly associated with *FERMT1* scores. Patients with higher *LRBA* gene pathogenicity scores more often underwent any IBD-related surgery ($P = 0.006$).

Across all European ancestry patients, higher mutational burden within *NOD2* was associated with lower use of 5-ASA medication ($P = 0.002$), consistent with this drug being of primary use in UC. These patients also had a significantly higher rate of stricturing disease ($P = 6.6 \times 10^{-7}$) confirming the association observed at the monogenic variant level. We hypothesized that GenePy may identify an association between stricturing disease

**Table 2.** Genetic and phenotypic characterization of 29 variants across 46 patients with "pathogenic" or "likely pathogenic" monogenic IBD gene variants

| Patient | Type of defect | Gene | Chromosome | Variant(s) | ACMG individual variant classification | ACMG pathogenicity classification | Sex | Age at diagnosis | Follow-up time (years) | CD | UC | Disease location at diagnosis (Paris Classification) | Stricturing disease | Perianal fistula/ abscess | Thiopurine use | Monoclonal therapy | Surgery | Skin lesions | Liver disease | Other |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | T + B cell defects | CD40LG | X | G219R | S | LPa | M | 15.3 | 1.25 | X | | L2 + L4a | — | — | X | — | — | — | — | |
| 2 | | CD40LG | X | G219R | S | LPa | M | 8.0 | 6.5 | X | | L2 | — | — | — | — | — | — | — | Mild disease—no medications |
| 3 | | CD40LG | X | G219R | S | LPa | M | 12.2 | 6.15 | X | | L3 | — | — | X | I, a | — | — | — | — |
| 4 | | CD40LG | X | G219R | S | LPa | M | 14.0 | 2.05 | | X | E2 | — | — | — | — | — | Eczema | — | — |
| 5 | | CD40LG | X | G219R | S | LPa | M | 13.3 | 5.3 | | X | E4 | — | — | X | I | — | Eczema | — | Asthma |
| 6 | | WAS | X | P460S | S | LPa | M | 11.5 | 7.95 | | X | E4 | — | — | X | I, a | SC | — | PSC | Recurrent infections, intermittent thrombocytopenia |
| 7[b] | | WAS | X | P460S | S | LPa | M | 14.9 | 5.6 | | X | E4 | — | — | X | — | SC | — | PSC | — |
| 8[b] | | WAS | X | E131K | S | LPa | M | 14.3 | 10.4 | X | | L2 + L4a | — | X | X | — | P | — | — | Also has STAT1 variant |
| 9 | | WAS | X | E131K | S | LPa | M | 13.5 | 5.4 | X | | L3 | X | — | X | I | — | — | — | — |
| 10 | | DKC1 | X | UTR5 142C>G | S | LPa | M | 13.6 | 10.1 | | X | E4 | — | — | X | — | — | — | — | — |
| 11 | | DKC1 | X | UTR5 142C>G | S | LPa | M | 9.2 | 6.4 | X | | L1 | — | — | X | — | — | — | — | Asthma, oral disease |
| 12 | | DCLRE1C | 10 | G153R / P171R | M / S | LPa | M | 10.7 | 5.95 | X | | L3 | — | — | X | I | — | — | — | — |
| 13 | | DCLRE1C | 10 | G153R / P171R | M / S | LPa | F | 13.2 | 1.9 | | X | E4 | — | — | X | — | — | — | — | — |
| 14 | Autoinflammatory | XIAP | X | T470S | S | LPa | M | 4.9 | 4.35 | X | | L1 + L4a | — | X | X | — | P | — | — | — |
| 15 | Phagocytic defects | NCF1 | 7 | R90H / R90H | S / S | Pa | M | 2.9 | 0.9 | | X | E4 | — | — | X | — | — | — | — | — |
| 16 | | NCF2 | 1 | H389Q | S | Pa | M | 11.0 | 4.15 | X | | L3 + L4a | — | — | X | I | — | — | — | Aplastic anemia, initially abnormal liver function—now resolved |
| 17 | | NCF2 | 1 | H389Q / H389Q / R395Q | S / S / M | LPa | F | 13.2 | 9.9 | | X | E2 | — | — | X | — | — | — | — | Coeliac disease |
| 18 | | NCF2 | 1 | H389Q / N419I | S / S | Pa | M | 14.8 | 2.0 | X | | L3 + L4a | — | — | X | I | — | — | — | — |
| 19 | Immunoregulation | TRIM22 | 11 | R321K / R321K | S / S | Pa | M | 15.3 | 7.45 | X | | L3 + L4a | — | — | X | A | — | — | — | — |
| 20 | | TRIM22 | 11 | R321K / R442C | S / S | Pa | M | 12.2 | 6.15 | X | | L3 + L4a | X | — | X | I | — | — | — | — |
| 21[a] | | TRIM22 | 11 | R321K / R442C | S / S | Pa | F | 14.1 | 1.75 | | X | E4 | — | — | X | — | — | — | — | — |
| 22 | | TRIM22 | 11 | R321K / S244L | S / S | Pa | M | 2.4 | 5.15 | X | | L2 | — | X | X | I | P | — | — | — |
| 23 | | TRIM22 | 11 | R321K / P484S | S / M | Pa | F | 9.6 | 9.3 | X | | L2 | — | — | X | — | SC | — | — | Epilepsy |
| 24 | | STAT1 | 2 | V266I | S | LPa | M | 5.5 | 10.1 | | X | E2 | — | — | X | — | — | — | — | — |
| 8[b] | | STAT1 | 2 | V266I | S | LPa | M | 14.3 | 10.4 | X | | L2 + L4a | — | X | X | — | P | — | — | Also has WAS variant |
| 25 | | STAT1 | 2 | V266I | S | LPa | F | 12.6 | 4.8 | X | | L3 | — | — | — | — | — | — | — | Hypothyroidism, mild disease—no medication |

## Table 2. (*continued*)

| Patient | Type of defect | Gene | Chromosome | Variant(s) | ACMG individual variant classification | ACMG pathogenicity classification | Sex | Age at diagnosis | Follow-up time (years) | CD | UC | Disease location at diagnosis (Paris Classification) | Stricturing disease | Perianal fistula/abscess | Thiopurine use | Monoclonal therapy | Surgery | Skin lesions | Liver disease | Other |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 26 | | *MASP2* | 1 | D120G | S | Pa | F | 10.7 | 1.7 | X | | E1 | — | — | X | I | — | — | — | Mild left-sided disease consistent with MASP2 deficiency |
| | | | | D120G | S | | | | | | | | | | | | | | | |
| 27 | | *NOD2* | 16 | R702W | S | Pa | F | 16.0 | 3.42 | X | | L1 | X | — | X | — | RH | — | — | — |
| | | | | R702W | S | | | | | | | | | | | | | | | |
| 28 | | *NOD2* | 16 | R702W | S | Pa | F | 14.5 | 6.35 | X | | L3 | X | — | X | — | RH | — | — | — |
| | | | | R702W | S | | | | | | | | | | | | | | | |
| 29 | | *NOD2* | 16 | R702W | S | Pa | F | 7.4 | 4.3 | X | | L1 | X | — | X | I | RH | — | — | — |
| | | | | R702W | S | | | | | | | | | | | | | | | |
| 30 | | *NOD2* | 16 | R702W | S | Pa | F | 9.8 | 1.6 | X | | L3 + L4a | X | — | X | I | RH | — | — | — |
| | | | | R702W | S | | | | | | | | | | | | | | | |
| 31 | | *NOD2* | 16 | R702W | S | Pa | M | 5.8 | 6.1 | X | | L3 + L4a | — | — | X | I | — | — | — | — |
| | | | | A755V | S | | | | | | | | | | | | | | | |
| 32 | | *NOD2* | 16 | R702W | S | Pa | F | 15.5 | 0.3 | X | | L1 | X | — | X | I | — | Eczema | — | — |
| | | | | A755V | S | | | | | | | | | | | | | | | |
| 33 | | *NOD2* | 16 | R702W | S | Pa | F | 9.7 | 10.7 | X | | L3 | X | — | X | — | RH | — | — | Asthma |
| | | | | V955I | S | | | | | | | | | | | | | | | |
| 34 | | *NOD2* | 16 | R702W | S | Pa | M | 5.9 | 5.8 | X | | L2 | — | — | — | — | — | — | — | Mild disease |
| | | | | V955I | S | | | | | | | | | | | | | | | |
| 35 | | *NOD2* | 16 | R702W | S | Pa | F | 9.4 | 6.05 | X | | L3 | — | — | X | I | — | Psoriasis | — | — |
| | | | | V955I | S | | | | | | | | | | | | | | | |
| 36 | | *NOD2* | 16 | R702W | S | LPa | M | 15.7 | 1.4 | X | | L3 + L4a | X | — | X | I | RH | — | — | Asthma |
| | | | | N852S | Su | | | | | | | | | | | | | | | |
| 37 | | *NOD2* | 16 | R702W | S | Pa | F | 14.5 | 2.7 | X | | L3 + L4a | X | — | X | I | RH | — | — | — |
| | | | | A1007 fs | VS | | | | | | | | | | | | | | | |
| 38 | | *NOD2* | 16 | R702W | S | Pa | M | 11.5 | 2.55 | X | | L1 | X | — | X | I | RH | Eczema | — | — |
| | | | | A1007 fs | VS | | | | | | | | | | | | | | | |
| 39 | | *NOD2* | 16 | V955I | S | Pa | F | 15.1 | 8.0 | X | | L3 + L4a | X | — | X | — | RH | — | — | — |
| | | | | A1007 fs | VS | | | | | | | | | | | | | | | |
| 40 | | *NOD2* | 16 | R702W | S | LPa | M | 13.3 | 1.85 | X | | L3 | X | — | X | I | RH | Eczema | — | Asthma |
| | | | | H352R | Su | | | | | | | | | | | | | | | |
| 41 | | *NOD2* | 16 | A755V | S | Pa | M | 15.0 | 5.9 | X | | L3 | X | — | X | — | RH | — | — | — |
| | | | | G908R | S | | | | | | | | | | | | | | | |
| 42 | | *NOD2* | 16 | V955I | S | Pa | M | 13.1 | 2.6 | | X | E3 | — | — | X | I | — | — | — | — |
| | | | | G908R | S | | | | | | | | | | | | | | | |
| 43 | | *NOD2* | 16 | V955I | S | LPa | F | 11.7 | 3.45 | X | | L1 | — | — | X | — | — | Eczema | — | Asthma |
| | | | | R744W | Su | | | | | | | | | | | | | | | |
| 44 | | *NOD2* | 16 | V955I | S | LPa | M | 13.6 | 5.7 | X | | L3 | X | X | — | — | RH, P | — | — | — |
| | | | | E963G | M | | | | | | | | | | | | | | | |
| 45 | | *NOD2* | 16 | D824N | Su | LPa | M | 9.6 | 7.0 | X | | L1 | — | — | X | — | — | — | — | — |
| | | | | G908R | S | | | | | | | | | | | | | | | |
| 46 | | *NOD2* | 16 | G908R | S | LPa | M | 15.0 | 1.7 | X | | L2 + L4a | — | — | — | — | — | — | — | — |
| | | | | R708H | Su | | | | | | | | | | | | | | | |

I, infliximab; A, adalimumab; SC, subtotal colectomy; P, perianal procedure (drainage or seton); RH, right hemicolectomy; dash (—) signifies absence of that feature; data were available for all patients.

ACMG individual variant classification: evidence of pathogenicity; VS, very strong; S, strong; M, moderate; Su, supporting.

ACMG classification of pathogenicity: Pa, pathogenic; LPa, likely pathogenic.

IBD, inflammatory bowel disease; PSC, primary sclerosis cholangitis.

[a]Segregation not performed due to lack of parental DNA.

[b]Patient #8 is hemizygous for both a *WAS* and *STAT1* variant.

**Table 3.** GenePy score comparison between top 10% of patients with IBD (n = 36) with top 10% of controls (n = 18)

| Gene | Cases vs Controls significance | FDR correction | CD | UC | IBDU | CD/UC enrichment |
|---|---|---|---|---|---|---|
| ADA | 0.0023 | **0.0364** | 20 | 14 | 2 | 0.3504 |
| COL7A1 | 0.0339 | 0.2671 | 23 | 12 | 1 | 1 |
| FERMT1 | 0.0023 | **0.0364** | 23 | 11 | 2 | 1 |
| LRBA | 0.0014 | **0.0364** | 28 | 6 | 2 | 0.0812 |
| NOD2 | 0.0015 | **0.0364** | 32 | 4 | 0 | **0.0046** |
| STXBP2 | 0.0279 | 0.2515 | 22 | 13 | 1 | 0.7081 |
| TRIM22 | 0.0141 | 0.1538 | 28 | 8 | 0 | 0.1946 |
| TTC37 | 0.0146 | 0.1538 | 26 | 10 | 0 | 0.5785 |

False discovery rate (FDR) correction for multiple testing was applied. Enrichment for CD/UC subtypes was investigated using the Fisher exact test.
Bold results indicate statistical significance at $P < 0.05$.
IBD, inflammatory bowel disease.

driven by patients who carry a high *NOD2* mutational burden but did not fulfill the criteria for "pathogenic" or "likely pathogenic" monogenic *NOD2*-related disease. Therefore, we excluded patients with "pathogenic" or "likely pathogenic" monogenic *NOD2* variants, and for the remaining 338 patients, we repeated the regression analysis of *NOD2* GenePy scores. Despite limiting this analysis to patients who do not fulfill the criteria for "pathogenic" or "likely pathogenic" NOD2 variation, there endures a striking negative correlation between a high *NOD2* gene pathogenicity score and the use of 5-ASA ($P = 0.006$) and a strong positive correlation with stricturing disease ($P = 7.3 \times 10^{-4}$).

## DISCUSSION

In our unselected cohort of pediatric patients, we identified a "pathogenic" or "likely pathogenic" variant, in a known monogenic IBD gene, in 11.5% of patients. When considering VEOIBD only, this rate was 16.7%. Over the recent years, NGS has identified Mendelian causes of patients presenting with particularly severe IBD-like phenotypes (43,44). The perceived wisdom is that these are exceptional cases masquerading as IBD, and contemporary molecular diagnostics are unlikely to yield clinically relevant diagnostic rates within the general IBD population (18). In this study, we applied extremely stringent filtering criteria insisting on validated functional evidence for variants and our observations are likely to underestimate the true prevalence of monogenic gene variants. Whether this nontrivial rate is maintained in adult cohorts remain to be seen.

*NOD2* was the first genetic locus identified in IBD and was designated the IBD1 locus through linkage studies in the 1990s (4). These analyses were focused on pedigrees with early onset and severe disease and suggested an AR inheritance pattern (45,46). More recently, *NOD2* has been the most consistent hit in GWAS of IBD and fueled the argument for common variation predisposing to disease, with many studies focusing on R702W, G908R, and 1007fs only (47,48). Our findings are consistent with monogenic *NOD2*-related disease representing the molecular basis of 5% of all pediatric IBD cases, increasing to 7.3% for CD. Most cases harbored compound heterozygous variants, having one low frequency variant and a second different very rare/novel mutation. This supports previous data from the 2000s, where studies independently identified a RR of 9.8–44 in individuals with compound heterozygous or homozygous *NOD2* variants. However, these studies were limited by identifying only commonly reported *NOD2* variants and did not account for additional rare or novel variants (5,6,8). Our results reinforce the recent evidence by Horowitz et al. (7) who suggested that up to 7.8% of pediatric patients had monogenic *NOD2*-related disease, and the modest differences in diagnostic rates between both studies are likely because of our application of conservative and stringent filtering criteria. Our data additionally report the distinct clinical characteristics that segregate with monogenic *NOD2*-related disease. We describe the E963G variant as "likely pathogenic" based on *in silico* evidence, segregation with a known pathogenic variant, and the presence

**Table 4.** Clinical phenotype characteristics associated with genes overburdened with pathogenic mutations in patients with inflammatory bowel disease (IBD)

| Gene | Effect | B | SE | P |
|---|---|---|---|---|
| ADA | Gender (male) | 0.277 | 0.119 | 0.021 |
|  | Isolated colonic disease | 0.253 | 0.118 | 0.033 |
| LRBA | Surgery | 0.011 | 0.004 | 0.006 |
| NOD2 | 5-ASA | −0.056 | 0.018 | 0.002 |
|  | Stricturing disease | 0.134 | 0.026 | $6.6 \times 10^{-7}$ |
| NOD2 excluding reportable monogenic patients | 5-ASA | −0.047 | 0.017 | 0.006 |
|  | Stricturing disease | 0.094 | 0.028 | $7.3 \times 10^{-4}$ |

with a distinct stricturing phenotype; however, classification of novel variants without functional validation remains challenging and confirmation of the impact of this variant is important.

A relationship between *NOD2* variation and stricturing disease phenotype was first discussed in 2002. Abreu et al. (49) reported ORs of 2.4 and 7.4 for heterozygous and pooled compound heterozygous/homozygote variants, respectively, in an analysis limited to R702W, G908R, and 1007 fs mutations. Subsequent studies, summarized elsewhere, have questioned whether this association is because of *NOD2* predisposing to ileal disease, rather than fibrostenotic disease *per se*, with conflicting results (50). Many of these studies examined a limited number of variants (R702W, G908R, or 1007 fs), did not correct for ileal disease location, or did not differentiate between heterozygous and compound heterozygous/homozygote variants (50). To our knowledge, this study is the first to analyze stricturing phenotype while considering *NOD2* as a monogenic cause of disease, including all rare and novel variants passing the stringent filtering criteria. After correction for disease location, we observe a striking increase in stricturing disease and surgical resection risk in monogenic *NOD2* patients, with the highest reported RR to date (4.32 and 4.59, respectively).

Additional monogenic causes for disease were identified in this cohort, with clear phenotype-genotype correlation observed for some variants. We describe the second report of monogenic IBD associated with variants in *TRIM22* (37). One of our patients has a phenotype consistent with the recent description of severe early onset perianal CD; however, we describe 4 patients with a severe but variable phenotype. This suggests a spectrum of *TRIM22*-related disease presenting throughout childhood. We identify a novel relationship between severe extensive UC, PSC, and the *WAS* P460S variant not previously reported in IBD, which may have treatment implications for patients presenting with this genotype (38). This variant allele has a frequency of 0.0023 and may impart variable penetrance similar to other X-linked IBD genes (18).

Application of a whole gene pathogenicity scoring tool enabled us to assess the burden for any given gene in individuals, rather than assessing single variants only. Despite no individual patient fulfilling strict criteria for monogenic disease, *ADA*, *FERMT1*, and *LRBA* accumulated significantly higher mutation burden in cases compared with controls. Either the observed variants in these genes play a role in polygenic IBD risk or there are additional noncoding mutations undetected by exome sequencing that lead to AR disease in patients with higher mutation burden. Nevertheless, we discern clinically informative significant associations between *LRBA* pathogenic burden in children with increased rates of IBD-related surgery.

Using GenePy, we confirm a significant role for *NOD2* in stricturing disease, excluding monogenic *NOD2* diagnoses. Our results indicate that within the set of patients not achieving a *NOD2* monogenic disease diagnosis, there remain patients whose disease is underpinned by deficient *NOD2* signaling. These data provide further evidence that *NOD2* heterozygosity has either some penetrance, that cumulative burden of mild *cis* or *trans* variants impact on disease, or more likely, an undetected mutation in noncoding regulatory region(s) constitute the "second hit" under a recessive model. Although GenePy provides a contemporary method for assessing pathogenicity across a gene, it is limited by the imperfection of deleteriousness metrics, as evidenced by the modest CADD score assigned to the *NOD2* V955I variant, despite this variant's known role in disease. It is possible that variants such as V955I are in linkage disequilibrium with additional intronic or promotor variants, not detected through whole exome sequencing, which is the true "second hit" in a recessive model in these patients. There is a clear necessity for more functional workup of variant impact, both on a *per variant* basis and in combination (both *cis* and *trans*). Multiple variants within a single haplotype have the potential to behave diversely—acting in synergy to reduce or increase functionality or may mutually compensate in ways that cannot be predicted by single variant functional analyses.

Missing heritability of IBD remains. Twin studies estimate heritability at 0.75 in CD and 0.67 in UC compared with 0.37 and 0.27 from the GWAS data (51). Because GWAS are only powered to detect features attributable to common variation, some of this missing heritability is likely because of very rare or private mutation, as observed through identification of rare variants and increased mutation burden in patients in this study.

Accumulating data provide compelling evidence to advocate for *NOD2* screening in newly diagnosed IBD to inform predictive algorithms for treatment including the need for surgical resection (7). Enabling personalized therapy becomes more important with the development of new drugs, such as *RIPK2* inhibitors, that modulate the *NOD2* signaling pathway (52,53). However, focusing only on *NOD2* would limit the potential benefits of precision diagnostics by excluding analysis of other genes, as evidenced by clinically relevant variation in *TRIM22* and *WAS* in this modest cohort. Any monogenic IBD gene panel would need to adapt flexibly alongside gene discovery. National programs are committing to providing NGS for any child admitted to intensive care with an unknown diagnosis because this approach has achieved a diagnostic rate of 25%. The value of a molecular diagnosis, especially one that provides certain prognosis and bespoke management to a child with a serious chronic disease, is invaluable. Our data deliver persuasive evidence for NGS diagnostics as a standard of care in pediatric-onset IBD, providing a precise diagnosis and personalized therapy for a substantial number of patients.

## CONFLICTS OF INTEREST

INFLAMMATORY BOWEL DISEASE

INFLAMMATORY BOWEL DISEASE

## Study Highlights

### WHAT IS KNOWN

✓ Mendelian disorders are the cause of IBD in a subset of very young patients.

✓ *NOD2* is the risk gene mostly associated with Crohn's disease.

✓ Making a precise molecular diagnoses leads to personalized therapy in some patients.

### WHAT IS NEW HERE

✓ Variants classified as "pathogenic" or "likely pathogenic" in monogenic IBD genes using the American College of Medical Genetics guidelines were identified in 11.5% of pediatric patients, median age 11.92 years.

✓ Pediatric-onset *NOD2*-related disease led to an increased risk of stricturing (OR 11.52) and intestinal resection (OR 10.75) compared with non-*NOD2* Crohn's disease.

✓ *TRIM22* variants are replicated for the first time and were seen in ~1% of patients.

✓ A *WAS* variant was newly observed in 2 patients with severe ulcerative colitis and primary sclerosis cholangitis.

## REFERENCES

1. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119–24.
2. Khor B, Gardet A, Xavier RJ, et al. Genetics and pathogenesis of inflammatory bowel disease. Nature 2011;474:307–17.
3. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119–24.
4. Hugot JP, Laurent-Puig P, Gower-Rousseau C, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 1996;379:821–3.
5. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599–603.
6. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001;411:603–6.
7. Horowitz JE, Warner N, Staples J, et al. Mutation spectrum of NOD2 reveals recessive inheritance as a main driver of early onset Crohn's Disease. bioRxiv 2017:098574.
8. Ahmad T, Armuzzi A, Bunce M, et al. The molecular classification of the clinical manifestations of Crohn's disease. Gastroenterology 2002;122:854–66.
9. Kugathasan S, Denson LA, Walters TD, et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: A multicentre inception cohort study. Lancet 2017;389:1710–8.
10. Hyams JS, Davis Thomas S, Gotman N, et al. Clinical and biological predictors of response to standardised paediatric colitis therapy: A multicentre inception cohort study. Lancet 2019;393:P1708–1720.
11. Peplow M. The 100,000 genomes project. BMJ 2016;353:i1757.
12. Ashton JJ, Ennis S, Beattie RM. Early-onset paediatric inflammatory bowel disease. Lancet Child Adolesc Heal 2017;1:147–158.
13. Takahashi S, Andreoletti G, Chen R, et al. De novo and rare mutations in the HSPA1L heat shock gene associated with inflammatory bowel disease. Genome Med 2017;9:8.
14. Worthey EA, Mayer AN, Syverson GD, et al. Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med 2011;13:255–62.
15. Kotlarz D, Beier R, Murugan D, et al. Loss of interleukin-10 signaling and infantile inflammatory bowel disease: Implications for diagnosis and therapy. Gastroenterology 2012;143:347–55.
16. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nat Publ Gr 2013;14:415.
17. Uhlig HH. Monogenic diseases associated with intestinal inflammation: Implications for the understanding of inflammatory bowel disease. Gut 2013;62:1795–805.
18. Uhlig HH, Muise AM. Clinical genomics in inflammatory bowel disease. Trends Genet 2017;33:629–41.
19. 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.
20. Ashton JJ, Mossotto E, Ennis S, et al. Personalising medicine in inflammatory bowel disease—current and future perspectives. Transl Pediatr 2019;8:56.
21. IBD Working Group of the European Society for Paediatric Gastroenterology, Hepatology and Nutrition. Inflammatory bowel disease in children and adolescents: Recommendations for diagnosis—the porto criteria. J Pediatr Gastroenterol Nutr 2005;41:1–7.
22. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. J Pediatr Gastroenterol Nutr 2014;58:795–806.
23. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.
24. Mossotto E, Ashton JJ, O'Gorman L, et al. GenePy—a score for estimating gene pathogenicity in individuals using next-generation sequencing data. BMC Bioinformatics 2019;20:254.
25. Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet 2012;91:839–48.
26. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv (Genomics) 2013:1303.3997.
27. 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;20:1297–303.
28. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–8.
29. Stenson PD, Mort M, Ball EV, et al. The human gene mutation database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet 2017;136:665–77.
30. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–91.
31. Girardelli M, Loganes C, Pin A, et al. Novel NOD2 mutation in early-onset inflammatory bowel phenotype. Inflamm Bowel Dis 2018;24:1204–12.
32. Frade-Proud'Hon-Clerc S, Smol T, Frenois F, et al. A novel rare missense variation of the NOD2 gene: Evidences of implication in Crohn's disease. Int J Mol Sci 2019;20:835.
33. Landrum MJ, Lee JM, Benson M, et al. ClinVar: Public archive of interpretations of clinically relevant variants. Nucleic Acids Res 2016;44:D862–8.
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;17:405–24.
35. Quang D, Chen Y, Xie X. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics 2015;31:761–3.
36. Pedersen BS, Quinlan AR. Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with Peddy. Am J Hum Genet 2017;100:406–13.
37. Li Q, Lee CH, Peters LA, et al. Variants in TRIM22 that affect NOD2 signaling are associated with very early onset inflammatory bowel disease HHS public access. Gastroenterology 2016;150:1196–207.
38. Zheng Y, Wu S, Yu X, et al. The WASP P460S mutation causes a new phenotype of WASP mutations related disorder: X-linked pancytopenia. Blood 2017;130:1044.

39. Ohya T, Yanagimachi M, Iwasawa K, et al. Childhood-onset inflammatory bowel diseases associated with mutation of Wiskott-Aldrich syndrome protein gene. World J Gastroenterol 2017;23:8544–52.

40. 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;22:2317–27.

41. Thiel S, Steffensen R, Christensen IJ, et al. Deficiency of mannan-binding lectin associated serine protease-2 due to missense polymorphisms. Genes Immun 2007;8:154–63.

42. Dhillon SS, Fattouh R, Elkadri A, et al. Variants in nicotinamide adenine dinucleotide phosphate oxidase complex components determine susceptibility to very early onset inflammatory bowel disease. Gastroenterology 2014;147:680–9.e2.

43. Oh SH, Baek J, Liany H, et al. A synonymous variant in IL10RA affects RNA splicing in paediatric patients with refractory inflammatory bowel disease. J Crohns Colitis 2016;10:1366–71.

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

45. Lesage S, Zouali H, Cézard JP, et al. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. Am J Hum Genet 2002;70:845–57.

46. Brant SR, Panhuysen CIM, Bailey–Wilson JE, et al. Linkage heterogeneity for the IBD1 locus in Crohn's disease pedigrees by disease onset and severity. Gastroenterology 2000;119:1483–90.

47. Rivas MA, Beaudoin M, Gardet JE, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 2011;43:1066–73.

48. Adler J, Rangwalla SC, Dwamena BA, et al. The prognostic power of the NOD2 genotype for complicated Crohn's disease: A meta-analysis. Am J Gastroenterol 2011;106:699–712.

49. Abreu MT, Taylor KD, Lin YC, et al. Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. Gastroenterology 2002;123:679–88.

50. Verstockt B, Cleynen I. Genetic influences on the development of fibrosis in Crohn's disease. Front Med 2016;3:24.

51. Gordon H, Trier Moller F, Andersen V, et al. Heritability in inflammatory bowel disease: From the first twin study to genome-wide association studies. Inflamm Bowel Dis 2015;21:1428–34.

52. Salla M, Aguayo-Ortiz R, Danmaliki GI, et al. Identification and characterization of novel receptor-interacting serine/threonine-protein kinase 2 inhibitors using structural similarity analysis. J Pharmacol Exp Ther 2018;365:354–67.

53. Hrdinka M, Schlicher L, Dai B, et al. Small molecule inhibitors reveal an indispensable scaffolding role of RIPK2 in NOD2 signaling. EMBO J 2018;37:e99372.