INFLAMMATORY BOWEL DISEASES®
Official Journal of the Crohn's & Colitis Foundation of America, Inc.

Identification of variants in genes associated with single gene inflammatory bowel disease by whole exome sequencing

Journal:	*Inflammatory Bowel Diseases*
Manuscript ID	IBD-16-0372
Wiley - Manuscript type:	Original Research Articles - Basic Science
Date Submitted by the Author:	25-May-2016
Complete List of Authors:	Ashton, James; University of Southampton, Department of Paediatric Gastroenterology Andreoletti, Gaia; University of Southampton, Human Genetics and Genomic Medicine Coelho, Tracy; University of Southampton, Human Genetics and Genomic Medicine Haggarty, Rachel; NIHR Nutrition Biomedical Research Centre, Southampton Centre for Biomedical Research Batra, Akshay; University Hospital Southampton NHS Foundation Trust, Paediatric Gastrointestinal Unit Afzal, Nadeem; University Hospital Southampton NHS Foundation Trust, Paediatric Gastroenterology Beattie, R; University Hospital Southampton NHS Foundation Trust, Paediatric Gastroenterology Ennis, Sarah; University of Southampton, Human Genetics and Genomic Medicine
Keywords:	Inflammatory bowel disease, paediatric, monogenic IBD, exome sequencing

Identification of variants in genes associated with single gene inflammatory bowel disease by whole exome sequencing

James J Ashton[1,2]* BMBS, Gaia Andreoletti[1]*PhD, Tracy Coelho[1,2]MRCPCH, Rachel Haggarty[3], Akshay Batra[2]MD, Nadeem A Afzal[2]MD, R Mark Beattie[2]FRCPCH and Sarah Ennis[1]PhD.

*These authors have contributed equally to this study

1. Human Genetics & Genomic Medicine, University of Southampton, Duthie Building (Mailpoint 808), Southampton General Hospital, Southampton, SO16 6YD, UK
2. Department of Paediatric Gastroenterology, University Hospital Southampton NHS Foundation Trust, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK
3. NIHR Nutrition Biomedical Research Centre, Southampton Centre for Biomedical Research, University Hospital Southampton NHS Foundation Trust (Mailpoint 218), Southampton General Hospital, Southampton, SO16 6YD, UK

Correspondence to:

Professor Sarah Ennis

Human Genetics & Genomic Medicine

University of Southampton

Duthie Building (Mailpoint 808)

Southampton General Hospital

Southampton

SO16 6YD

Tel: +44 (0)23 8079 8614

Fax: +44 (0)23 8079 4264

s.ennis@southampton.ac.uk

Word count 3316

Abstract word count 249

Contributions

SE, GA and JJA conceived the study.

It was designed by SE, GA and JJA. GA and JJA collected and analysed the data. JJA wrote the manuscript in conjunction with all authors.

All authors approved the manuscript prior to submission.

Acknowledgements and sources of funding

The authors are very grateful to all participants and their families. We thank Liz Blake for assisting with pediatric recruitment, Nikki Graham and Steven Booth for technical assistance in DNA laboratory in Human Genetics & Genomic Medicine, University of Southampton, the NIHR & the Southampton Centre for Biomedical Research (SCBR). This work is supported by The Crohn's in Childhood Research Association (CIRCA) and GA is supported by The Gerald Kerkut Charitable Trust. JJA is funded by a University of Southampton National Institute of Health Research Academic Clinical Fellowship.

Conflicts of interest

There are no conflicts of interest to declare

Key words-

Inflammatory bowel disease; exome sequencing; monogenic IBD; paediatric

Abstract

Background-Most cases of inflammatory bowel disease (IBD) are caused by complex host-environment interaction. There are a number of conditions associated with a single gene mutation, most cases are very early onset (aged < 6 years), present with a unique form of disease and often have atypical features.

Methods-Whole-exome data for 147 paediatric IBD patients was interrogated for a panel of 51 genes associated with monogenic IBD. Observed variation was categorised according to the American College of Medical Genetics (ACMG) guidelines to identify rare, *novel* and known variants that might contribute to IBD.

Results-574 variants were identified across 51 genes. These were categorised in line with ACMG guidance to remove benign variants and to identify 'pathogenic' and 'likely pathogenic' variants. In six patients we observed six pathogenic variants of which *CYBA(*c.287+2T>C), *COL7A1(*c.6501+1G>C*), LIG4(*p.R814X), and *XIAP(*p.T470S) were known causative mutations and *FERMT1(*p.R271Q) and *SKIV2L*(c.354+5G>A) were novel. In the three patients with *XIAP, SKIV2L* and *FERMT1* variants, individuals' disease features resembled the monogenic phenotype. This was despite apparent heterozygous carriage of pathogenic variation for the latter two genes. The *XIAP* variant was observed in a hemizygous male.

Conclusion-Whole exome sequencing allows for identification of known and *de novo* potentially causative mutations in genes associated with monogenic IBD. Whilst these are rare conditions it is vital to identify causative mutations early in order to improve prognosis. We postulate that in a subset of IBD, heterozygous mutations (in genes thought to manifest IBD through autosomal recessive inheritance) may contribute to clinical presentation.

Introduction

Inflammatory bowel disease (IBD), comprising Crohn's disease (CD), ulcerative colitis (UC) and inflammatory bowel disease unclassified (IBDU) are a heterogeneous group of conditions carrying significant morbidity. The majority of disease is thought to be caused by a complex interaction between predisposing genetic factors (1), a dysfunctional immune system (2) and gut microbiome (3). In a small subgroup of patients, typically with early onset (age < 10 years) and very early onset (age < 6 years) IBD, there is a monogenic cause for the disease (4). Whether or not IBD should be the primary diagnosis or the condition be labelled as IBD-like phenotype of an underlying immune defect is uncertain.

The overall incidence of paediatric IBD (pIBD) is increasing (5, 6) however the presenting symptoms and severity has remained unchanged (7). The frequency of IBD caused by a single genetic variant (monogenic IBD) is very low (4). Individuals who have monogenic IBD are important to recognise as they have increased risk of developing significant concurrent problems, such as immunodeficiency, and this will impact on treatment options and prognosis (8). In addition monogenic IBD may also be associated with specific features not typically associated with IBD such as nail and hair abnormalities (9), epidermolysis bullosa (10, 11) and autoimmune haemolytic anaemia (12). Identification of atypical signs such as these should trigger further testing in these patients. To date, 51 genes have been identified linked to monogenic IBD, with the majority also associated with additional clinical features; particularly a functional immune disorder (such as chronic granulomatous disease or severe combined immunodeficiency syndrome) (8, 13). There is a potential for misdirected treatment of patients with monogenic disease; receiving escalated treatment regimens with extreme forms of surgery and medical therapies rather than treating the underlying immune or other defect.

The accessibility of next generation sequencing technology has allowed identification of rare and novel pathogenic variants in pIBD (14). Furthermore, variants in genes associated with primary immunodeficiency have been identified in patients with very early onset IBD (15), alongside specific

mutations in genes associated with monogenic IBD (16, 17). Previously whole-exome sequencing has helped identify an association between and children presenting very-early-onset IBD and homozygous mutations in the interleukin 10 receptor (*IL10)* gene, *IL10* associated receptor alpha and beta subunits (*IL10RA* and *IL10RB*), homozygous mutations in *ADAM17* and hemizygous mutations within *FOXP3* (18, 19).

This study utilises whole-exome sequencing data from a cohort of children (up to 16 years of age) diagnosed with IBD to extract all variants across 51 genes associated with monogenic IBD and identify potentially pathogenic mutations. This study aimed to look for mutations in genes associated with monogenic IBD in all paediatric cases, not merely those presenting with (very) early onset disease.

Materials and Methods

Recruitment

Children are recruited following diagnosis by the paediatric gastroenterology service at University Hospital Southampton (UHS). All children aged under 18 years are eligible for inclusion and all are diagnosed in line with the Porto criteria (20). Clinical data is recorded for each patient.

Ethical approval

The study has ethics approval from Southampton & South West Hampshire Research Ethics Committee (09/H0504/125).

DNA extraction

Genomic DNA was extracted from peripheral venous blood samples collected in EDTA using the salting out method (21). DNA concentration was estimated using the Qubit ® 2.0 Fluorometer and α 260:280 ratio calculated using a nanodrop spectrophometer. The average DNA yield obtained was 150µg/ml and approximately 20ug of each patient DNA was extracted for next generation sequencing.

Whole-exome sequencing and data processing

Whole-exome capture was performed using Agilent SureSelect Human All Exon 51 Mb (versions 4 and 5) capture kit. Raw data generated from paired-end sequencing protocol were aligned against hg19 using Novoalign (novoalign/2.08.02) as previously described (14, 22). Mapping steps produced parameters for each sequenced position, such as base quality, coverage, alternative allele, reference allele and the number of reads at that position. Sequence coverage for each sample was calculated with in-house customized scripts that applied the BedTools (23) package (v2.13.2). PICARD (picard/1.97) was used to remove duplicate reads and SAMtools (24) mpileup (samtools/0.1.18) was used to call SNPs and short INDELs from the alignment file. Variations with read depth < 4 were excluded. Good quality bases with a phred score > 20 were retained for downstream analysis (25, 26). ANNOVAR (annovar/2013Feb21) (27) was applied for variant annotation. A bespoke script was used to assign individual variants as: “novel” if

they were not previously reported in the dbSNP137 databases (28), 1000 Genomes Project phase one (1KG) (29), the Exome Variant Server (EVS) of European Americans of the NHLI-ESP project with 6500 exomes [http://evs.gs.washington.edu/EVS/], in 46 unrelated human subjects sequenced by Complete Genomics (46 CG) (30) or in the Southampton database of reference exomes. Resultant variant files for each subject were subjected to further in-house quality control tests to detect DNA sample contamination and ensure sex concordance by assessing autosomal and X chromosome heterozygosity. Variant sharing between all pairs of individuals was assessed to confirm that subjects were not related. Sample provenance was confirmed by application of a validated panel developed specifically for exome data (31).

Copy number variations (CNVs) were assessed using the software ExomeDepth.

Gene selection and filtering strategy

A list of 50 genes taken from Uhlig *et al* (8) and updated with a single gene from Li *et al* (13) gave a total of 51 genes for interrogation after comprehensive literature review (table 1). Any variation within these 51 genes was extracted from the variant call files generated for each of the pIBD patients.

The following list of filters was applied in order to exclude variation unlikely to have clinical impact: all synonymous variants; variants common within the general population (MAF_{1KG}> 0.05); variants within intron-exon splice boundaries considered unlikely to impact splicing (MaxEnt score < 3); poorly conserved variants (Phylop < 0.95); variants within homopolymer tracts or repeat regions; those representing alignment artifacts or flagged as likely false-positive (32). All remaining novel, non-synonymous, frameshift and non-frameshift insertion/deletions, splicing, stop gain and stop loss mutations were considered and grouped based on the American College of Medical Genetics (ACMG) guidelines (33) into the categories 'Pathogenic', 'Likely Pathogenic' and 'Benign'.

The ACMG guidelines on classification of variants specifies that the functional impact of mutations must have been assessed to classify the variant as pathogenic; all pathogenic variants have previous functional work and are listed in the human gene mutation database (HGMD) (34). Likely pathogenic variants have

functional impact inferred from similar mutations and demonstrate compelling clinical correlation.

Sanger sequencing and segregation analysis

Variants within the `Pathogenic` group occurring in the correct zygosity to be causal and assessed as deleterious by *in silico* annotation tools were verified by Sanger sequencing in the probands and all relatives for whom DNA was available (Fig. S1a-c). Primers were designed using primerBLAST and sequencing was outsourced at Source Bioscience, Nottingham, UK.

Results

Southampton pIBD cohort

One hundred and forty-seven individual patient exomes have been sequenced from the Southampton PIBD cohort. Demographic data for patients is shown in table 2.

Characterization of mutations within genes associated with monogenic form of IBD

Within the cohort, 574 variants were called across all 51 genes specifically associated with a single-gene IBD. A total of 67 mutations remained following standard filtering criteria (Figure 1). Following the ACMG guidelines, four of these were determined as 'Pathogenic' category, two as 'Likely Pathogenic' category and 61 as 'Benign' category. No CNVs were identified in genes with 'Pathogenic' or 'Likely Pathogenic' mutations.

'Pathogenic' and 'Likely Pathogenic' mutations

Four pathogenic, *CYBA* (c.287+2T>C), *COL7A1* (c.6501+1G>C), *LIG4* (p.R814X), and *XIAP* (p.T470S), and 2 likely pathogenic, *FERMT1* (p.R271Q) and *SKIV2L* (c.354+5G>A), variants were identified in six independent probands. *CYBA* (Chronic granulomatous disease), *COL7A (*Hallopeau-Siemens recessive dystrophic epidermolysis bullosa), *LIG4 (*Lig4 syndrome), *FERMT1* (Kindler's syndrome) and *SKIV2L (*Trichohepatoenteric syndrome) are known to cause disease in an autosomal recessive inheritance pattern. *XIAP* (X-linked lymphoproliferative disease type 2) causes disease in an X-linked recessive pattern (table 3).

As *CYBA, COL7A, LI4G, FERMT1* and *SKIV2L* are known to cause disease in homozygous state exome data for the patients harbouring heterozygous variants within these genes were further interrogated in order to identify a second, pathogenic, common variant which might contribute to the phenotype (table 3). We observed common variants all genes other than *LIG4*: *CYBA* (2), *COL7A* (1), *FERMT1* (5) and *SKIV2L* (2). Patient specific mutations, characteristics and associated disease types can be seen in table 4.

COL7A1 (c.6501+1G>C): Patient 1 is a female patient diagnosed with Crohn's disease at 12 years of age with concurrent autoimmune hypothyroidism but no skin manifestations at time of examination. She carries a rare splicing mutation in the *COL7A1* gene (c.6501+1G>C) in heterozygous state. This mutation was previously associated with Hallopeau-Siemens recessive dystrophic epidermolysis bullosa, a condition causing blistering of the skin and digestive tract (35). This patient also carries a common (MAF = 0.67) heterozygous synonymous (p.P939P) variant in *COL7A1* which is unlikely to contribute to disease phenotype.

***SKIV2L* (c.354+5G>A):** Patient 2 was diagnosed aged 5 years with ulcerative colitis. He presented an extremely severe course with prolonged periods of steroid dependency. He did not present with hair abnormalities or deranged liver function tests at most recent examination. He carries a novel splicing mutation c.354+5G>A within the *SKIV2L* gene, which is associated with trichohepatoenteric syndrome. This condition is an autosomal recessive disease consisting of intractable diarrhoea, hair and facial abnormalities and most often presents in infancy. Patient 2 harbours a second common, poorly conserved, likely benign nonsynonymous variant (p.M214L) in homozygous state and a common synonymous (p.Y1067Y) variant in homozygous state within *SKIV2L*. There was no significant difference in depth of coverage for any exon within this gene, in which mutations p.M214L and p.Y1067Y occur, for this patient and two other samples indicating a low chance of multiple exon deletion in 1 allele.

LIG4 (p.R814X): Patient 3 is a female patient diagnosed aged 12 with IBDU and she was previously diagnosed at age 6 years with vitiligo. She harbours a known stop gain (p.R814X) variant within the *LIG4* gene, associated with so called LIG4 syndrome. This syndrome consists of immunodeficiency, skin abnormalities (including photosensitivity) and IBD presenting with protracted diarrhoea (36). The patient carries the variant in a heterozygous state and she does not harbour any other common or rare mutation within *LIG4* identified using WES data available.

CYBA* (c.287+2T>C):* Patient 4, a female patient diagnosed at the age of 16 with extremely severe stricturing Crohn's disease with multiple granulomas requiring urgent surgery. She carries a heterozygous

splicing mutation within the *CYBA* gene (c.287+2T>C) known to be associated with chronic granulomatous disease (CGD) (37). This patient also harbours two common (MAF_{1KG} > 5%) variants within *CYBA*; 1 is non-synonymous (p.V174A) and assessed by *in silico* tools to be benign variant and the other is synonymous (p.E12E).

FERMT1 (p.R271Q): Patient 5 is an early onset UC patient diagnosed at age 9 years. He was subsequently diagnosed with severe oral pemphigus by immunofluorescence at age 12. He carries a rare (MAF = 0. 000116) non-synonymous mutation within the *FERMT1 (*c.G812A, p.R271Q) gene on chromosome 20 in heterozygous state. This identified variant causes a mutation in the same codon of the known *FERMT1* stop gain mutation (c.C811T, p.R271X) which is known to cause Kindler's syndrome, a blistering skin disease that may present with ulcerative colitis (4, 38, 39). Within the same gene the patient also carries a second common nonsynonymous variant (p.R526K) in homozygous state, four common synonymous variants (p.F565F, p.K525K, in homozygous state, and p.L385L and p.H38H in heterozygous state) and two splicing variants (c.532+8T>C and c.152-4G>A with a MaxEnt score of 1.2 and 1.72 respectively). There was no significant difference in depth of coverage for any exon in this gene compared between patient 5 and two other samples. Variants p.R271Q and p.R526K were confirmed by Sanger sequencing in the proband and relatives where available (Figure 2 and Figure 1a-b in supplementary).

***XIAP* (p.T470S)**: We identified a known non-synonymous variant within the *XIAP* (c.A1408T, p.T470S) gene on the X chromosome, known to cause X-linked lymphoproliferative disease type 2 in patient 6, a very early onset (age < 6 years) CD male patient. X-linked lymphoproliferative disease type 2 is an extremely rare condition that may present with fistulating IBD, accompanied by perianal disease and dysgammaglobulinemia (40). This patient presented at 4 years of age with severe fistulating Crohn's disease, he had an extremely severe disease course with recurrent perianal disease and fistulae and also suffered from isolated IgA deficiency. Segregation analysis in the proband and available family member confirmed hemizygous variant in the proband. The variant is absent in the unaffected father and present in the unaffected mother in heterozygous state (Figure 3 and Figure 1c in Supplementary). Of note patient 6

has a half-sister (same father) with Crohn's disease, she does not harbour this *XIAP* variant, does not have dysgammaglobinaemia and presented at a later age.

Discussion

This study applies whole-exome sequencing to 147 paediatric IBD patients and interrogates a panel of 51 genes identified as being associated with monogenic causes of IBD. We have rigorously applied the recently updated ACMG guidelines for variant classification to our results and identified four known and two *novel* variants that fit the 'pathogenic' or 'likely pathogenic' categories respectively (33). 'Likely pathogenic' and 'pathogenic' variant classification methodology can be seen tables 3 and 5 in Richards *et al* (33). Importantly according to ACMG guidelines variants are classified based on the specific variant, with the zygosity of that variant ignored for classification. We observe known pathogenic variants in heterozygous form in childhood onset disease concurrent with a striking overlap of the additional clinical features specific to the accepted recessively inherited disease. We postulate that in some cases, pathogenic mutations may be penetrant to varying degrees in heterozygous form and the 'all or nothing' paradigm for recessive penetrance may be an oversimplification of the genetic aetiology in selected cases.

Patient 6, harbouring a known causative mutation in the *XIAP* gene for X-linked lymphoproliferative disease type 2 has a phenotype which closely resembles that described for the condition; early onset severe Crohn's- like symptoms, dysgammaglobinaemia and perianal disease (41, 42). This condition is extremely rare, previous estimates of disease associated with *XIAP* mutations, including the more common X-linked lymphoproliferative disease type 1, have put the prevalence of *XIAP* mutations at 4% of paediatric male IBD patients, although this is likely to be a significant overestimate due to selection bias- overrepresentation of young, severe patients in studies sequencing exomes/genomes (8, 43). Here we report 1 patient with a causative XIAP mutation (1.2% of frequency in 85 EO males and 7.6% frequency in 13 VEO males), our data are subject to the same selection bias as previous studies.

The other five patients, all harbouring potentially causative mutations presented with a range of symptoms. Patient 5 presented with oral pemphigus and symptoms of IBD and was identified as being heterozygote for a novel nonsynonymous variant within the same codon as that of a known causative mutation for Kindler's syndrome (38, 39, 44, 45). Whilst this patient's phenotype, specifically their skin

manifestation, was not as severe as typical Kindler's syndrome there is reasonable suspicion that the heterozygous mutation will have contributed to the phenotype seen, perhaps in conjunction with the second variant observed in the *FERMT1* gene of this patient as concurrent pemphigus and IBD is extremely rare.

Patient 2, who presented with colitis at the age of 5 was found to have a novel mutation in *SKIV2L.* Whilst a very similar homozygote mutation is known to cause Trichohepatoenteric syndrome, presenting at 1-12 weeks of age (46). Whilst this child does not present with classical symptoms of this condition, this their disease course has been extremely severe with frequent relapses. Interestingly there have been previous reports of milder phenotype with colitis presenting at the age of 4.5 years (47). Also of note *SKIV2L* is within 1 megabase of the HLA complex genes although we do not detect any variants in this region.

If all identified 'likely pathogenic' and 'pathogenic' variants have contributed to the development of IBD in these patients we can estimate the prevalence of paediatric IBD contributed to by a monogenic variant at 4%. However, in our study, VEO patients and patients with severe phenotype were preferentially selected for exome sequencing, possibly leading to an overrepresentation of potential monogenic causes in this cohort (8). IBD has is typically considered a polygenic disorder, however the 163 IBD-associated loci identified by GWAS only account for a small part of the heritability seen in IBD (48, 49). Rare mutations in genes associated with monogenic IBD would not be detected by GWAS and therefore may account for some of this missing heritability (8).

Excluding the XIAP variant (disease causing in hemizygous state), all of the variants identified are previously reported to cause disease in a homozygous state (and have undergone functional validation), however these mutations are all in heterozygous state within our remaining five patients. By relaxing our filtering criteria we identified a second, more common variant in 4 of the 5 genes (see table 3). There is precedent in genetic causes of nystagmus for common 'benign' variants contributing to disease when they coexist with a highly deleterious heterozygote variant (50).

We postulate that heterozygote 'pathogenic' and 'likely pathogenic' variants seen in genes associated with monogenic IBD may account for an attenuated phenotype (of the full condition) with variable penetrance of the mutated allele. Our hypothesis that heterozygote variants known to cause monogenic disease in homozygous state may still effect clinical manifestation to some lesser degree has been observed in other conditions such as adenomatous polyposis coli (51), Parkinsonism (52) and disorders of eye development (53).

Exome sequencing by definition overlooks non-coding but potentially functional regions of genes; mutations in these regions as well as their intergenic regulatory sequences may be contributing to disease susceptibility, severity and co-morbidities. In addition, it is likely that additional mutations in modifier genes are also determining aspects of clinical presentation.

Extensive functional studies are required for definitive interpretation of mutations and for common diseases, it may not be informative to assess single variants in isolation. Whilst functional validation of variants has been conducted in homozygous cases (see table 3), detailed functional interrogation of all heterozygote variation (either in isolation or in the context of the entire mutational profile inherited by one individual) observed in this cohort is beyond the scope of the current paper. The bottleneck to clinical translation of personalised genomics imposed by a relative deficiency in functional assessment of variants of unknown significance, is already being observed for rare diseases and the scope of this problem increases in complexity for common disease.

Recent reviews by Uhlig have highlighted the importance of considering monogenic causes of IBD, especially in very early onset IBD (4, 8). It has long been accepted that intractable diarrhoea of infancy (early onset IBD) was likely to have a specific genetic basis in many patients with a significant proportion having underlying IBD (54). Previous work has highlighted the genetic heterogeneity of CD and hypothesised that some of the missing heritability is associated with monogenic disease (55). Our study highlights the utility of whole-exome sequencing in identification of novel and known variants in this panel of genes associated with monogenic causes of IBD.

Although all patients presented with IBD it is important to recognise that many of the monogenic conditions associated with IBD have broader phenotypes that may lead to subsequent development of other problems, most often associated with immune dysfunction or deficiency (4). Early identification of these conditions, potentially via routine exome sequencing, may be of huge benefit to individual patients, preventing mismanagement and enabling potentially curative treatments (8).

Conclusion

From these data we hypothesise that phenotypic presentation in some cases of IBD may be influenced by heterozygous mutations in genes previously associated with severe monogenic IBD. Further work is needed to functionally examine potentially pathogenic variants in genes associated with monogenic forms of IBD in other large cohorts.

Acknowledgements

The authors are very grateful to all participants and their families. We thank Liz Blake for assisting with pediatric recruitment, Nikki Graham and Steven Booth for technical assistance in DNA laboratory in Human Genetics & Genomic Medicine, University of Southampton, the NIHR & the Southampton Centre for Biomedical Research (SCBR). This work is supported by The Crohn's in Childhood Research Association (CIRCA) and GA is supported by The Gerald Kerkut Charitable Trust. JJA is funded by a University of Southampton National Institute of Health Research Academic Clinical Fellowship.

Figures

Figure 1 Variant filter steps. A total of 574 variants in all 51 known monogenic IBD genes were extracted, of which 219 synonymous variants were discarded due to their low likelihood to impact protein function. Of the remaining 355 variants, 214 variants were removed as these were observed in in any zygosity state within the local control cohort of exomes (n=315), 1 variant was discounted due to a MAF > 0.05 (1000 Genome Project) and 28 splicing variants with a MaxEnt score < 3 were removed. Of the 112 mutations remaining, 45 variants were removed due to their low conservation across species (Phylop < 0.90). A total of 67 mutations remained of which four of these were allocated into the 'Pathogenic' category, two into the 'Likely Pathogenic' category and 61 into the 'Benign' category of the ACMG guidelines.

Figure 2. Patient 5 family pedigree. Segregation analysis for FERMT1 variant c.1577G>A and c.812G>A. c.1577G>A is in heterozygous status in proband and mother while c.812G>A is in homozygous status in proband and heterozygous status in mother and father.

Figure 3. Patient 6 family pedigree. Segregation analysis for *XIAP* variant c.1408A>T. The variant is present in heterozygous and hemizygous status in the mother and in the proband respectively.

Table 1 Genes associated with Monogenic IBD, adapted from Uhlig et al (8)

Table 2 Southampton PIBD cohort demographics

Table 3 Pathogenic, Likely Pathogenic and 'second hit' variants identified in genes known to cause monogenic IBD

Table 4 Clinical details of patients with 'Pathogenic' and 'Likely Pathogenic' variants

References

1. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. Lancet. 2007;369(9573):1627--40.

2. Baumgart DC, Sandborn WJ. Crohn's disease. Lancet. 2012;380(9853):1590--605.

3. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15(3):382-92.

4. Uhlig HH. Monogenic diseases associated with intestinal inflammation: implications for the understanding of inflammatory bowel disease. Gut. 2013;62(12):1795-805.

5. Ashton JJ, Wiskin AE, Ennis S, et al. Rising incidence of paediatric inflammatory bowel disease (PIBD) in Wessex, Southern England. Arch Dis Child. 2014;99(7):659-64.

6. Henderson P, Hansen R, Cameron FL et al. Rising incidence of pediatric inflammatory bowel disease in Scotland. Inflamm Bowel Dis. 2012;18:999--1005.

7. Ashton JJ, Coelho T, Ennis S, et al. Presenting Phenotype of Paediatric Inflammatory Bowel Disease (PIBD) in Wessex, Southern England 2010-13. Acta Paediatr. 2015.

8. Uhlig HH, Schwerd T, Koletzko S, et al. The diagnostic approach to monogenic very early onset inflammatory bowel disease. Gastroenterology. 2014;147(5):990-1007.e3.

9. Blaydon DC, Biancheri P, Di WL, et al. Inflammatory skin and bowel disease linked to ADAM17 deletion. N Engl J Med. 2011;365(16):1502-8.

10. Freeman EB, Köglmeier J, Martinez AE, et al. Gastrointestinal complications of epidermolysis bullosa in children. Br J Dermatol. 2008;158(6):1308-14.

11. Shah N, Freeman E, Martinez A, et al. Histopathological features of gastrointestinal mucosal biopsy specimens in children with epidermolysis bullosa. J Clin Pathol. 2007;60(7):843-4.

12. Agarwal S, Mayer L. Pathogenesis and treatment of gastrointestinal disease in antibody deficiency syndromes. J Allergy Clin Immunol. 2009;124(4):658-64.

13. Li Q, Lee CH, Peters LA, et al. Variants in TRIM22 that Affect NOD2 Signaling Are Associated With Very Early Onset Inflammatory Bowel Disease. Gastroenterology. 2016.

14. Christodoulou K, Wiskin AE, Gibson J, et al. Next generation exome sequencing of paediatric inflammatory bowel disease patients identifies rare and novel variants in candidate genes. Gut. 2013;62:977--84.

15. Kelsen JR, Dawany N, Moran CJ, et al. Exome Sequencing Analysis Reveals Variants in Primary Immunodeficiency Genes in Patients With Very Early Onset Inflammatory Bowel Disease. Gastroenterology. 2015;149(6):1415-24.

16. Kelsen JR, Dawany N, Martinez A, et al. A de novo whole gene deletion of XIAP detected by exome sequencing analysis in very early onset inflammatory bowel disease: a case report. BMC Gastroenterol. 2015;15(1):160.

17. Fiskerstrand T, Arshad N, Haukanes BI, et al. Familial diarrhea syndrome caused by an activating GUCY2C mutation. N Engl J Med. 2012;366(17):1586-95.

18. Dinwiddie DL, Bracken JM, Bass JA, et al. Molecular diagnosis of infantile onset inflammatory bowel disease by exome sequencing. Genomics. 2013;102(5-6):442-7.

19. Mao H, Yang W, Lee PP, et al. Exome sequencing identifies novel compound heterozygous mutations of IL-10 receptor 1 in neonatal-onset Crohn's disease. Genes Immun. 2012;13(5):437-42.

20. 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(6):795-806.

21. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988;16(3):1215.

22. Andreoletti G, Ashton JJ, Coelho T, et al. Exome Analysis of Patients with Concurrent Pediatric Inflammatory Bowel Disease and Autoimmune Disease. Inflamm Bowel Dis. 2015.

23. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.

24. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078--9.

25. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998;8(3):186--94.

26. Ewing B, Hillier KL, Wendl MC, et al. Base-Calling of Automated Sequencer Traces UsingPhred. I. Accuracy Assessment. Genome Res. 1998;8(3):175--85.

27. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

28. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308--11.

29. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56--65.

30. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010;327(5961):78--81.

31. Pengelly RJ, Gibson J, Andreoletti G, et al. A SNP profiling panel for sample tracking in whole-exome sequencing studies. Genome Med. 2013;5(9):89.

32. Fuentes Fajardo KV, Adams D, Mason CE, et al. Detecting false-positive signals in exome sequencing. Hum Mutat. 2012;33(4):609-13.

33. 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(5):405-24.

34. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Human Genetics, 2014;133(1), 1–9.

35. Dang N, Murrell DF. Mutation analysis and characterization of COL7A1 mutations in dystrophic epidermolysis bullosa. Exp Dermatol. 2008;17(7):553-68.

36. O'Driscoll M, Cerosaletti KM, Girard PM, et al. DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. Mol Cell. 2001;8(6):1175-85.

37. Kammermeier J, Drury S, James CT, et al. Targeted gene panel sequencing in children with very early onset inflammatory bowel disease--evaluation and prospective analysis. J Med Genet. 2014;51(11):748-55.

38. Siegel DH, Ashton GH, Penagos HG, et al. Loss of kindlin-1, a human homolog of the Caenorhabditis elegans actin-extracellular-matrix linker protein UNC-112, causes Kindler syndrome. Am J Hum Genet. 2003;73(1):174-87.

39. Techanukul T, Sethuraman G, Zlotogorski A, et al. Novel and recurrent FERMT1 gene mutations in Kindler syndrome. Acta Derm Venereol. 2011;91(3):267-70.

40. Pachlopnik Schmid J, Canioni D, Moshous D, et al. Clinical similarities and differences of patients with X-linked lymphoproliferative syndrome type 1 (XLP-1/SAP deficiency) versus type 2 (XLP-2/XIAP deficiency). Blood. 2011;117(5):1522-9.

41. Tsuma Y, Imamura T, Ichise E, et al. Successful treatment of idiopathic colitis related to XIAP deficiency with allo-HSCT using reduced-intensity conditioning. Pediatr Transplant. 2015;19(1):E25-8.

42. Aguilar C, Lenoir C, Lambert N, et al. Characterization of Crohn disease in X-linked inhibitor of apoptosis-deficient male patients and female symptomatic carriers. J Allergy Clin Immunol. 2014;134(5):1131-41.e9.

43. Zeissig Y, Petersen BS, Milutinovic S, et al. XIAP variants in male Crohn's disease. Gut. 2015;64(1):66-76.

44. Lai-Cheong JE, Parsons M, Tanaka A, et al. Loss-of-function FERMT1 mutations in kindler syndrome implicate a role for fermitin family homolog-1 in integrin activation. Am J Pathol. 2009;175(4):1431-41.

45. Has C, Wessagowit V, Pascucci M, et al. Molecular basis of Kindler syndrome in Italy: novel and recurrent Alu/Alu recombination, splice site, nonsense, and frameshift mutations in the KIND1 gene. J Invest Dermatol. 2006;126(8):1776-83.

46. Fabre A, Charroux B, Martinez-Vinson C, et al. SKIV2L mutations cause syndromic diarrhea, or trichohepatoenteric syndrome. Am J Hum Genet. 2012;90(4):689-92.

47. Egritas O, Dalgic B, Onder M. Tricho-hepato-enteric syndrome presenting with mild colitis. Eur J Pediatr. 2009;168(8):933-5.

48. Kugathasan S, Baldassano RN, Bradfield JP, et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. Nat Genet. 2008;40(10):1211-5.

49. Bianco AM, Girardelli M, Tommasini A. Genetics of inflammatory bowel disease from multifactorial to monogenic forms. World J Gastroenterol. 2015;21(43):12296-310.

50. Hutton SM, Spritz RA. A comprehensive genetic study of autosomal recessive ocular albinism in Caucasian patients. Invest Ophthalmol Vis Sci. 2008;49(3):868-72.

51. Wang L, Baudhuin LM, Boardman LA, et al. MYH mutations in patients with attenuated and classic polyposis and with young-onset colorectal cancer without polyps. Gastroenterology. 2004;127(1):9-16.

52. Klein C, Lohmann-Hedrich K, Rogaeva E, et al. Deciphering the role of heterozygous mutations in genes associated with parkinsonism. Lancet Neurol. 2007;6(7):652-62.

53. Zaki PA, Collinson JM, Toraiwa J, et al. Penetrance of eye defects in mice heterozygous for mutation of Gli3 is enhanced by heterozygous mutation of Pax6. BMC Dev Biol. 2006;6:46.

54. Hizarcioglu-Gulsen H, Saltik-Temizel IN, Demir H, et al. Intractable diarrhea of infancy: 10 years of experience. J Pediatr Gastroenterol Nutr. 2014;59(5):571-6.

55. Bianco AM, Zanin V, Girardelli M, et al. A common genetic background could explain early-onset Crohn's disease. 4 ed. Med Hypotheses. 2012. p. 520-2.

56. Whittock NV, Ashton GH, Mohammedi R, et al. Comparative mutation detection screening of the type VII collagen gene (COL7A1) using the protein truncation test, fluorescent chemical cleavage of mismatch, and conformation sensitive gel electrophoresis. J Invest Dermatol. 1999;113(4):673-86.

57. Roos D, Kuhns DB, Maddalena A, et al. Hematologically important mutations: the autosomal recessive forms of chronic granulomatous disease (second update). Blood Cells Mol Dis. 2010;44(4):291-9.

58. Schäppi MG, Smith VV, Goldblatt D, et al. Colitis in chronic granulomatous disease. Arch Dis Child. 2001;84(2):147-51.

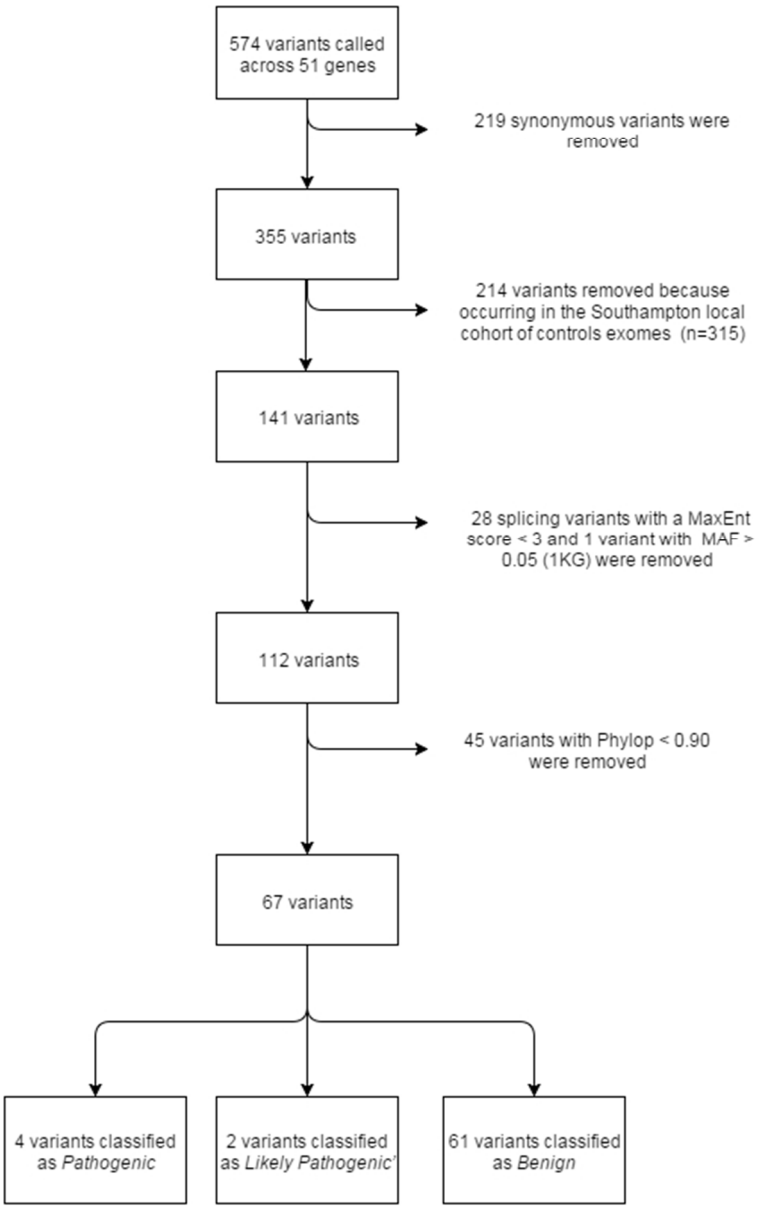

Figure 1 Variant filter steps. A total of 574 variants in all 51 known monogenic IBD genes were extracted, of which 219 synonymous variants were discarded due to their low likelihood to impact protein function. Of the remaining 355 variants, 214 variants were removed as these were observed in in any zygosity state within the local control cohort of exomes (n=315), 1 variant was discounted due to a MAF > 0.05 (1000 Genome Project) and 28 splicing variants with a MaxEnt score < 3 were removed. Of the 112 mutations remaining, 45 variants were removed due to their low conservation across species (Phylop < 0.90). A total of 67 mutations remained of which four of these were allocated into the 'Pathogenic' category, two into the 'Likely Pathogenic' category and 61 into the 'Benign' category of the ACMG guidelines.
135x209mm (96 x 96 DPI)

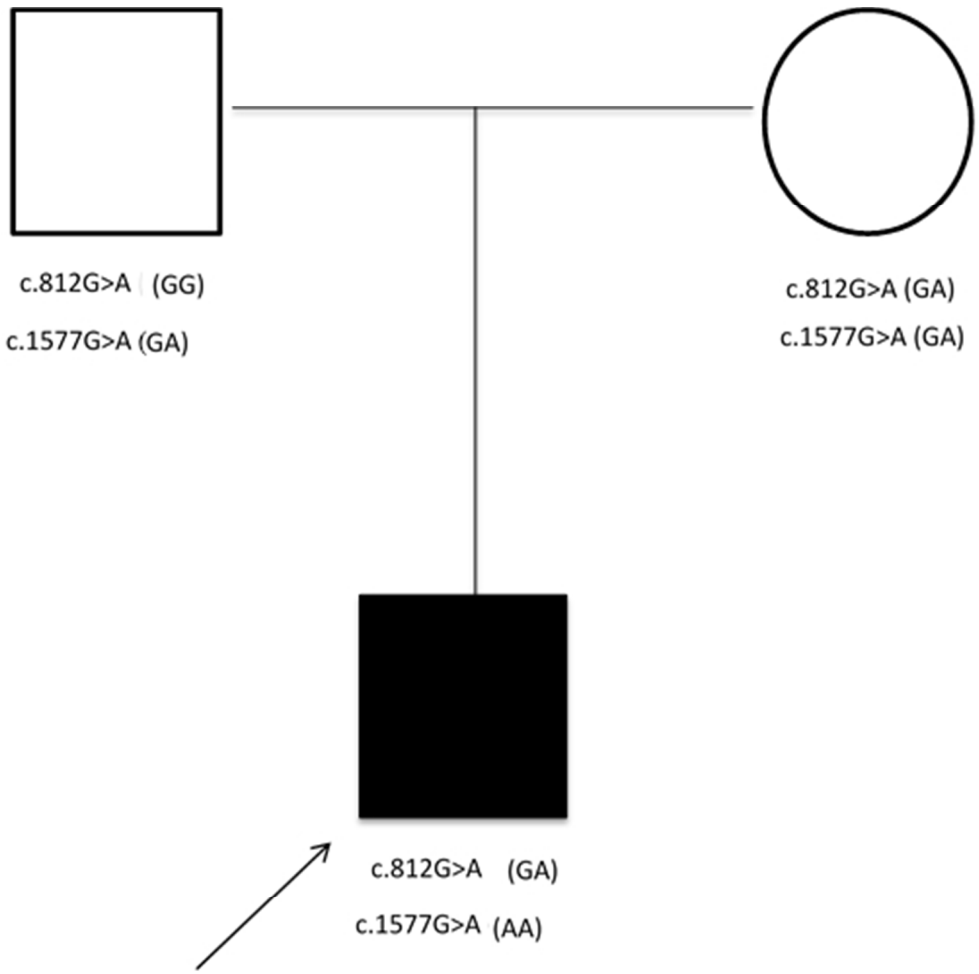

Figure 2. Patient 5 family pedigree. Segregation analysis for FERMT1 variant c.1577G>A and c.812G>A. c.1577G>A is in heterozygous status in proband and mother while c.812G>A is in homozygous status in proband and heterozygous status in mother and father.
135x134mm (96 x 96 DPI)

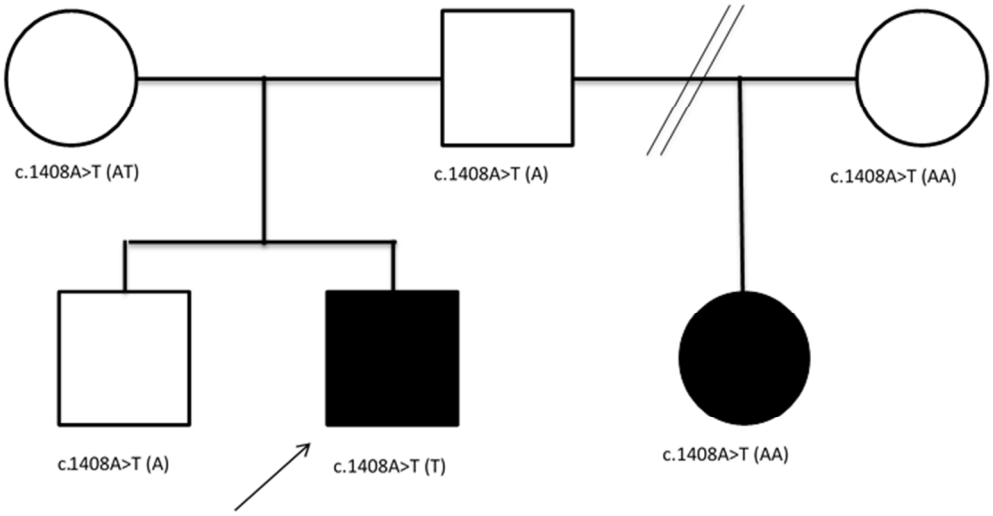

Figure 3. Patient 6 family pedigree. Segregation analysis for XIAP variant c.1408A>T. The variant is present in heterozygous and hemizygous status in the mother and in the proband respectively.
192x99mm (96 x 96 DPI)

Table 1 ***Genes associated with Monogenic IBD, adapted from Uhlig et al (8)***

Gene	*Associated condition*	*Inheritance*	Agilent V5 % gene coverage (51 patients)	Agilent V4 % gene coverage (96 patients)
ADA	*Severe combined immunodeficiency*	*AR*	100.00	97.85
ADAM17	*ADAM17 deficiency*	*AR*	100.00	79.74
AICDA	*Hyper IgM syndrome*	*AR*	93.27	30.92
BTK	*Agammaglobulinaemia*	*X*	93.76	87.21
CD3y	*Severe combined immunodeficiency*	*AR*	49.35	49.35
CD40LG	*Hyper IgM syndrome*	*X*	49.35	49.35
COL7A1	*Dystrophic bullosa*	*AR*	100.00	59.52
CYBA	*Chronic granulomatous disease*	*AR*	100.00	99.17
CYBB	*Chronic granulomatous disease*	*X*	100.00	96.68
DCLRE1C	*Omenn syndrome*	*AR*	100.00	46.34
DKC1	*Hoyeraal-Hreidarsson syndrome*	*X*	65.01	55.30
DOCK8	*Hyper IgE syndrome*	*AR*	100.00	78.77
FERMT1	*Kindler syndrome*	*AR*	99.29	88.98
FOXP3	*IPEX*	*X*	85.03	54.46
G6PC3	*Congenital neutropenia*	*AR*	93.03	63.06
GUCY2C	*Familial diarrhoea*	*AD*	68.80	71.77
HPS1	*Hermansky-Pudlak 1*	*AR*	100.00	89.35
HPS4	*Hermansky-Pudlak 4*	*AR*	91.68	60.22
HPS6	*Hermansky-Pudlak 6*	*AR*	99.30	71.34
ICOS	*CVID 1*	*AR*	100.00	98.80
IKBKG	*X-linked ectodermal immunodeficiency*	*X*	100.00	35.11
IL10	*IL-10 signalling defects*	*AR*	92.88	51.38
IL10RA	*IL-10 signalling defects*	*AR*	100.00	55.56
IL10RB	*IL-10 signalling defects*	*AR*	97.04	66.65
IL21	*IL-21 deficiency*	*AR*	100.00	100.00
IL2RA	*IPEX-like*	*AR*	100.00	38.77
IL2RG	*Severe combined immunodeficiency*	*X*	100.00	90.55
ITGB2	*Leukocyte adhesion deficiency type 1*	*AR*	86.80	81.29
LIG4	*Severe combined immunodeficiency*	*AR*	86.99	67.37
LRBA	*CVID8*	*AR*	99.10	90.65

MASP2	*MASP deficiency*	*AR*	100.00	93.76
MEFV	*Familial Mediterranean fever*	*AR*	89.56	75.59
MVK	*Mevalonate kinase deficiency*	*AR*	100.00	62.99
NCF1	*Chronic granulomatous disease*	*AR*	23.61	0.00
NCF2	*Chronic granulomatous disease*	*AR*	96.96	83.77
NCF4	*Chronic granulomatous disease*	*AR*	100.00	100.00
PIK3R1	*Agammaglobulinaemia*	*AR*	96.19	39.86
PLCG2	*Phospholipase C-y2 defects*	*AD*	93.33	93.68
RAG2	*Severe combined immunodeficiency*	*AR*	74.73	61.29
RTEL1	*Hoyeraal-Hreidarsson syndrome*	*AR*	87.14	83.26
SH2D1A	*X-linked lymphoproliferative syndrome type 1*	*X*	100.00	39.13
SKIV2L	*Trichohepatoenteric syndrome*	*AR*	16.66	16.16
SLC37A4	*Glycogen storage disease type 1b*	*AR*	53.53	54.86
STAT1	*IPEX-like*	*AD*	91.69	61.99
STXBP2	*Familial haemophagtocytic lymphohistiocytosis type 5*	*AR*	100.00	100.00
TRIM22	*Granumolatous colitis with severe perianal disease*	*AR*	90.50	57.90
TTC37	*Trichohepatoenteric syndrome*	*AR*	95.85	85.94
TTC7A	*TTC7A deficiency*	*AR*	100.00	57.68
WAS	*WAS*	*X*	100.00	95.85
XIAP	*X-linked lymphoproliferative syndrome type 2*	*X*	92.72	15.51
ZAP70	*Severe combined immunodeficiency*	*AR*	93.20	88.49

AR: autosomal recessive; X: X-linked

Table 1 Southampton PIBD cohort demographics

	Crohn's Disease	Ulcerative colitis	IBDU	Total IBD
Number of patients	87	37	23	147
Median Age	12.24	10.04	12.30	12.24
Female no. (%)	34 (39.09)	16 (43.24)	14 (60.87)	64 (43.54%)
Mean age of Onset (SD)	11.30 (3.53)	9.38 (3.98)	11.17 (3.54)	11.04 (3.80)

Table 1 ***Pathogenic, Likely Pathogenic and 'second hit' variants identified in genes known to cause monogenic IBD***

Patient	sex	Zygosity	Gene	Mode on inheritance	Chromosme	Location hg19	Variant type	Variant info	phylop	1-sift	,polyphen2,	mutationtaster,	gerp++	MaxEnt	dbSNP135	Frequency in 1KG	Frequency in EVS	Frequency in ExAc	ACHG annotation	Occurrence in controls (n=315)	Occurrence in cases (n=147)
1	F	1	COL7A1	AR	3	48611694	sp	COL7A1:NM_000094:exon80:c.6501+1G>C	.	.	.	.	.	8.27	.	Not seen	Not seen	0.000008245	pathogenic	Not seen	1
1	F	1	COL7A1	AR	3	48625266	sn	COL7A1:NM_000094:exon21:c.A2817G:p.P939P,	.	.	.	.	.	.	rs1264194	0.66	0.708256	0.6799	benign	237	138
2	F	1	SKIV2L	AR	6	31928119	sp	SKIV2L:NM_006929:exon4:c.354+5G>A	.	.	.	.	.	3.69	-	Not seen	Not seen	Not seen	pathogenic	Not seen	1
2	F	2	SKIV2L	AR	6	31929014	ns	SKIV2L:NM_006929:exon8:c.A640C:p.M214L,	0.222828	.	.	.	1.8	.	rs437179	0.78	0.711163	0.7699	benign	269	135
2	F	2	SKIV2L	AR	6	31936668	sn	SKIV2L:NM_006929:exon26:c.T3201C:p.Y1067Y,	.	.	.	.	.	.	rs410851	0.81	0.710303	0.7826	benign	257	135
3	F	1	LIG4	AR	13	108861177	sg	LIG4:NM_002312:exon2:c.C2440T:p.R814X,LIG4:NM_001098268:exon2:c.C2440T:p.R814X,LIG4:NM_206937:exon3:c.C2440T:p.R814X,	0.991331	0.903176	0.734134	1	4.25	.	rs104894419	0.0005	0.000233	0.00008237	pathogenic	Not seen	1
4	F	1	CYBA	AR	16	88713161	sp	CYBA:NM_000101:exon5:c.287+2T>C	.	.	.	.	.	7.75	rs747774702	Not seen	Not seen	0.00004987	pathogenic	Not seen	1
4	F	2	CYBA	AR	16	88709828	ns	CYBA:NM_000101:exon6:c.T521C:p.V174A,	0.000535	.	.	0.000001	-7.6	.	rs1049254	0.71	0.628315	0.6892	benign	135	122
4	F	2	CYBA	AR	16	88717386	sn	CYBA:NM_000101:exon1:c.A36G:p.E12E,	.	.	.	.	.	.	rs8053867	1	1	0.9991	benign	229	146
5	M	1	FERMT1	AR	20	6088216	ns	FERMT1:NM_017671:exon6:c.G812A:p.R271Q,	0.998967	0.99	1	0.999994	5.34	.	rs144791466	Not seen	0.000116	0.00004207	pathogenic	Not seen	1
5	M	2	FERMT1	AR	20	6064710	sn	FERMT1:NM_017671:exon13:c.T1695C:p.F565F,	.	.	.	.	.	.	rs753927	0.44	0.362442	0.3683	benign	187	92
5	M	2	FERMT1	AR	20	6065729	ns	FERMT1:NM_017671:exon12:c.G1577A:p.R526K,	0.963443	.	.	0.000272	5.01	.	rs2232074	0.48	0.051628	0.3977	benign	194	93
5	M	2	FERMT1	AR	20	6065731	sn	FERMT1:NM_017671:exon12:c.A1575G:p.K525K,	.	.	.	.	.	.	rs2232073	0.46	0.053953	0.3978	benign	194	93
5	M	1	FERMT1	AR	20	6069723	sn	FERMT1:NM_017671:exon10:c.C1153T:p.L385L,	.	.	.	.	.	.	rs35413391	0.04	0.076512	0.05306	benign	35	24
5	M	1	FERMT1	AR	20	6093116	sp	FERMT1:NM_017671:exon5:c.532+8T>C	.	.	.	.	.	1.20	rs41308641	0.07	0.129302	0.0956	benign	69	33
5	M	1	FERMT1	AR	20	6096695	sp	FERMT1:NM_017671:exon4:c.152-4G>A	.	.	.	.	.	1.72	rs2295435	0.4	0.43	0.4489	benign	194	98
5	M	1	FERMT1	AR	20	6100088	sn	FERMT1:NM_017671:exon2:c.T114C:p.H38H,	.	.	.	.	.	.	rs10373	0.51	0.532093	0.5243	benign	238	114
6	M	hemizygous	XIAP	X	X	123040945	ns	XIAP:NM_001167:exon7:c.A1408T:p.T470S,XIAP:NM_001204401:exon7:c.A1408T:p.T470S,	0.998574	0.94	0.004	0.274905	4.11	.	rs143165174	Not seen	0.000595	0.0004818	pathogenic	Not seen	1

One (1) denotes heterozygous state; two (2) denotes homozygous stat; AR: autosomal recessive; X: X-linked; ns: nonsynonymous, sn, synonymous; sp: splicing; benign variants are shaded in grey.

Phylop- Phylogenetic tests of lineage, 1-sift- predicts whether an amino acid substitution affects protein function, PolyPhen2- predicts impact of an amino acid substitution on protein structure and function, Mutationtaster- evaluates the pathogenic potential of DNA sequence alterations, Gerp- Genomic evolutionary rate profiling, MaxEnt- splice site scoring system

Table 1 Clinical details of patients with 'Pathogenic' and 'Likely Pathogenic' variants

Patient ID	Mutation	Novel or Known	ACMG guidelines- Pathogenic or Likely Pathogenic	Previously identified causative variant and functional impact	Phenotype associated with previous gene variant	Disease	Age at diagnosis (years)	Gender	Paris classification at diagnosis	Other clinical features	Clinical course since diagnosis	Family history
1	**COL7A1** (c.6501+1G>C)	Known	P	c.6501+1G>C (53)	**Hallopeau-Siemens recessive dystrophic epidermolysis bullosa** when homozygous- severe skin and digestive tract blistering (35, 56)	CD	12	F	L3	Autoimmune hypothyroidism diagnosed age 7 (Anti-thyroid peroxidase antibodies 2864iu/ml (<75iu/ml) Mouth ulcers	Mild course over 2 years follow-up	-
2	**SKIV2L** (c354+5G>A)	Novel	LP	c.355-2A>C (37)	**Trichohepatoenteric syndrome**- intractable diarrhoea, hair/facial abnormalities presenting in infancy (46)	UC	5	F	E4	No additional features at diagnosis	Turbulent with frequent relapses and prolonged steroid dependency over 11 year follow-up	Paternal ulcerative colitis
3	**LIG4 (**c.C2440T:p.R814X)	Known	P	c.C2440T p.R814X (39)	**Lig4 syndrome**- immunodeficiency, skin abnormalities (photosensitivity, psoriasis), protracted diarrhoea (36)	IBDU	12	F	NA	Vitiligo diagnosed age 6, presented with diarrhoea	Mild course over 5 year follow-up	Maternal great grandfather suffered from ulcerative colitis
4	**CYBA (**c.287+2T>C)	Known	P	c.287+2T>C (54)	**Chronic granulomatous disease**- recurrent infections, Crohn's-like colitis, perianal disease, granuloma are not always present on biopsy (58)	CD	16	F	L3 + structuring disease (B2)	Extremely severe stricturing disease requiring surgery on presentation to remove terminal ileal stricture. Granuloma on histology and extensive granulation tissue on resected specimen	Subsequent right hemicolectomy 1 year after diagnosis Progressed to anti-TNF therapy quickly and now dependant after 4 years follow-up.	-
5	**FERMT1** (c.G812A:p.R271Q)	Novel	LP	c.811C>T p.R271X (38)	**Kindler's syndrome**- blistering skin disease (38)	UC	9	M	E2	Severe oral pemphigus (blistering skin disease) diagnosed age 12 by immunofluorescence	Disease controlled with azathioprine after initial frequent relapses. 9 year follow-up	-
6	**XIAP (**c.A1408T:p.T470S)	Known	P	c.A1408T p.T470S (40)	**X-linked lymphoproliferative disease type 2**- dysgammaglobulinemia (can be low) and lymphoma. IBD-type presents with perianal disease (40)	CD	4	M	L1 + L4	Severe perianal disease presenting with abscess and fistula age 3. Mouth ulcers IgA deficiency (<0.07) but other immune work-up normal including neutrophil burst and ANCA	Ongoing perianal disease with subsequent fissures and recurrent fistulae. Turbulent course after 3 years follow-up	Brother has mouth ulcers and has been investigated for IBD aged 7 years. No diagnosis at time of writing. Half-sister (paternal) has severe Crohn's disease

P: pathogenic; LP: likely pathogenic; UC, ulcerative colitis; CD, Crohn's disease; IBDU, inflammatory bowel disease unclassified.