Somatic mosaicism and common genetic variation contribute to the risk of very-early-onset inflammatory bowel disease

Eva Gonçalves Serra et al.

Very-early-onset inflammatory bowel disease (VEO-IBD) is a heterogeneous phenotype associated with a spectrum of rare Mendelian disorders. Here, we perform whole-exome-sequencing and genome-wide genotyping in 145 patients (median age-at-diagnosis of 3.5 years), in whom no Mendelian disorders were clinically suspected. In five patients we detect a primary immunodeficiency or enteropathy, with clinical consequences (XIAP, CYBA, SH2D1A, PCSK1). We also present a case study of a VEO-IBD patient with a mosaic de novo, pathogenic allele in CYBB. The mutation is present in ~70% of phagocytes and sufficient to result in defective bacterial handling but not life-threatening infections. Finally, we show that VEO-IBD patients have, on average, higher IBD polygenic risk scores than population controls (99 patients and 18,780 controls; $P < 4 \times 10^{-10}$), and replicate this finding in an independent cohort of VEO-IBD cases and controls (117 patients and 2,603 controls; $P < 5 \times 10^{-10}$). This discovery indicates that a polygenic component operates in VEO-IBD pathogenesis.

---

*A full list of authors and their affiliations appears at the end of the paper.*
Inflammatory bowel disease (IBD) represents a heterogeneous group of disorders characterized by a dysregulated immune response toward commensal gut bacteria leading to chronic relapsing intestinal inflammation. Crohn’s disease (CD) and ulcerative colitis (UC), the two common forms of IBD, affect around 0.5% of individuals of European descent with lower, but rising, prevalence in other parts of the world.

Though IBD can occur at any age, the peak age of onset for CD and UC is between 20–30 years and 30–40 years, respectively. About 20% of patients develop IBD before the age of 17 years (pediatric-onset-IBD) and around 14/100,000 children have an IBD onset under the age of six (very-eary-onset IBD). Onset of IBD before the age of 2 years (infantile-IBD) or even within the first month of life (neonatal-IBD) is exceptionally rare. Patients with VEO-IBD present with a higher rate of pancolitis, and subgroups present with severely fistulising disease, resistance to conventional immunosuppressive treatments and immune defects associated with increased lethality.

Several Mendelian disorders present with IBD and IBD-like intestinal inflammation and have an onset during infancy or within the first 6 years of life. Biallelic loss-of-function variants in the IL10 signaling pathway (IL10, IL10RA, and IL10RB) are fully penetrant for VEO-IBD, while genetic variants underling several other Mendelian disorders show an incomplete penetrance of the IBD phenotype (for simplicity called monogenic IBD). The majority of these conditions are autosomal recessive or X-linked inherited primary immunodeficiencies (PIDs). A simple differential diagnosis based on phenotypic associations is often difficult and a sequential work-up based on candidate genes is time consuming. Whole-exome-sequencing (WES) is increasingly used to screen for causal mutations and such studies have identified pathogenic variants in a proportion of VEO-IBD patients, in genes such as IL10, IL10RA, IL10RB, XIAF, TTC7A, and TTC27.

Identifying the underlying disease-causing variants in VEO-IBD patients is important because it can directly influence patient management and inform on the appropriate treatment strategy, e.g., early haematopoietic stem cell transplantation (HSCT) in patients with haematopoietic defects caused by IL10, FOXP3, or XIAF mutations.

Rare monogenic forms of IBD are a stark contrast to the polygenic nature of pediatric and adult-onset IBD. Genome-wide association studies (GWAS) have identified more than 240 loci associated with IBD, the majority of which are driven by common variants (minor allele frequency (MAF) > 5%) of small effect (increasing risk by 1.1–1.3 fold), together explaining 13 and 8% of the variance in disease liability for CD and UC, respectively. The majority of patients included in these GWAS have adult-onset disease, and very few (<1%) have VEO-IBD. Genome-wide association studies focusing on pediatric-onset cases identified risk loci that, at the time, were not associated with adult-onset disease but all have since been robustly associated in adult-IBD cohorts. More recently, three studies focusing on patients with an age at disease onset greater than 6 years, reported a weak, but statistically significant, negative relationship between polygenic risk score and the age at CD and UC diagnoses.

The role of polygenicity in VEO-IBD remains unknown and it is possible that, while rare monogenic variants underlie disease in a proportion of VEO-IBD patients, an exceptionally high burden of common IBD-susceptibility alleles may also contribute to VEO-IBD risk. Such a hypothesis has not been previously investigated in VEO-IBD cohorts.

Here, we use WES and genome-wide SNP arrays to better understand the genetic architecture of VEO-IBD in a multi-center cohort of 145 patients with a median age-at-diagnosis of 3.5 years and a severe disease course, indicated by previous surgery or need for biological therapy. We include patients in whom a Mendelian disorder is expected due to clinical presentation, or were mutation negative following screening of specific VEO-IBD genes selected based on patient presentation (e.g., IL10, IL10RB, or IL10RA defects in patients with IBD onset in the first three months of life). The cohort is therefore potentially enriched for cases harboring undiscovered monogenic causes of VEO-IBD, or alternative causal mechanisms. We investigate the extent to which mutations in 67 known monogenic IBD genes account for disease in this selected cohort and search for novel monogenic causes exome-wide. Moving beyond rare variation, we use genome-wide SNP data to evaluate the role of common CD and UC-susceptibility alleles in the pathogenesis of VEO-IBD. By generating polygenic risk scores (PRS) based on the effect-size estimates of SNPs significantly associated with adult-onset CD and UC and replication in independent VEO-IBD and control cohorts, we investigate whether VEO-IBD children harbor a higher load of such alleles when compared to a large collection of adult-onset IBD cases or population controls (see Supplementary Fig. 1 for an overview of the study workflow).

Results

Cohort sequencing and quality control. A cohort of 145 VEO-IBD individuals and 4436 population controls were exome sequenced at a mean coverage of 69x and 53x, respectively. Following sample and variant-based quality control (see Methods section), 145 VEO-IBD cases and 3969 controls with equivalent sequencing-based QC metrics (Supplementary Fig. 2) remained for analysis, with an average of ~40,000 variants called per exome and 94% of genes covered at a mean depth of 30x or above (Supplementary Fig. 3). There were two well defined ancestry-matched groups within our dataset: 104 cases and 3787 controls defined as being of European descent and 21 cases and 68 controls defined as South Asian (the remaining samples did not cluster with a clearly defined ancestry group) (see Methods section, Supplementary Fig. 4).

Somatic mosaicism: non-Mendelian inheritance of VEO-IBD. The initial screening for pathogenic variants in established monogenic IBD genes identified a nonsense mutation in CYBB (p.W380X) in a 31-year-old male patient of European descent with infantile-onset of granulomatous colitis, perianal abscesses and hidradenitis suppurativa (Fig. 1a, f; Table 1). A detailed clinical summary of the case is presented in Supplementary Note 1 and Supplementary Fig. 5. Loss-of-function mutations in CYBB are known to cause chronic granulomatous disease (CGD). However, our patient had no history of invasive infections, a potential fatal hallmark of CGD unless patients are closely monitored and rapidly treated. We found no carriers of the mutation in more than ~156,000 whole-exome sequences from ethnically-diverse population controls (including ExAC and gnomAD).

CYBB resides on the X-chromosome and thus we were surprised to notice that only 122 out of 174 sequence reads (~70%) covering the site of the mutation carried the nonsense allele (A) (the other reads carried the wild-type allele—G). Investigating the common genetic variation across the X chromosome, we found no evidence of Klinefelter Syndrome (47 chromosomes, XXY) or partial X-chromosomal duplication that could explain this observation. We confirmed the mutation via Sanger sequencing undertaken in an independent clinical genetics laboratory. We also Sanger-sequenced DNA from the patient’s mother and sister and established that neither are carriers of the nonsense mutation (Fig. 1b), suggesting that the mutation is a de novo event of hemizygous mosaicism. Accordingly, 70% of neutrophils lacked CYBB protein expression, further supporting a hemizygous mosaicism (Fig. 1c). Functional
validation experiments confirmed the mosaicism, with 70% of neutrophils and monocytes showing completely absent NADPH-oxidase activity as seen in CGD patients, and 30% of cells showing a normal profile (Fig. 1d, see Methods section). This degree of mosaicism affected bacterial handling capacity since monocyte-derived macrophages from the patient had ~30% more colonies at baseline but did not respond to muramyl dipeptide (MDP) in the gentamicin protection assay (Fig. 1e; see Methods section), a finding suggestive of a bacterial handling defect. Our results suggest that the proportion of cells with complete loss-of-function of CYBB affects the phenotypic presentation, i.e., 30% wild-type neutrophils is sufficient to prevent life-threatening infections but 30% fully-functioning macrophages is insufficient to prevent intestinal inflammation as a result of ineffective bacterial killing and inflammatory cytokine production.

To further understand the origin and the developmental time at which the mutation arose during embryogenesis, we sequenced DNA from several flow cytometry assay (FACS) sorted immune cell subsets and used DNA from buccal swabs, containing epithelial cells, and hair follicles from the patient and compared those with wild type cells (Fig. 1f, g; see Methods section, Supplementary Fig. 6). In all immune cell subsets of the patient...
Fig. 1 Analysis of CYBB mosaicism in a male patient. a Pedigree structure for the family of the male patient with the mosaic hemizygous mutation in CYBB (chrX:37,663,371A/G, p.W380X). b Sanger sequencing of the chrX:37,663,371 CYBB mutation site in the patient and unaffected relatives (sister and mother). c p91-phox protein expression (the gene product encoded by CYBB) analysed by flow cytometry assay (FACS). Control is a healthy donor. d Measurement of oxidative burst in neutrophils and monocyes using the dihydrodorhamidine-1,2,3 (DHR) assay. Obtained from the patient and a healthy donor (control). e Defective bacterial handling in monocyte derived macrophages with the CYBB mosaicism. Intracellular survival of Salmonella typhimurium was quantified using the agar plate technique. Results show three technical replicates. Obtained from the patient and a healthy donor (control). f, g Quantification of mutant read proportion at chrX:37,663,371 using the IGV browser. PBMCs were sorted into immune cell subsets (Supplementary Figs. 6B, C) and compared with buccal swabs and hair follicles, as well as with healthy donor immune cells and a HEK293T cell line as technical control. h FACS sorting strategy for DHR-high and DHR-low populations following DHR staining and PMA stimulation (Supplementary Fig. 6A). i Quantification of mutant reads at chrX:37,663,371 following sorting based on DHR for control DHR-high, patient DHR-high, and patient DHR-low neutrophils (Supplementary Fig. 6A). j Gentamicin protection assay on neutrophils for control DHR-high, patient DHR-high, and patient DHR-low populations (Supplementary Fig. 6A).

Briefly, neutrophils were infected at a MOI 1:10 for 45 min with Salmonella enterica serovar typhimurium and subsequently treated with gentamicin for 45 min. Neutrophils were then lysed and plated on LB agar plates for CFU counting on the following day. **p < 0.001, Mann-Whitney U-test.

Primary immunodeficiencies as a cause of VEO-IBD. We identified four patients with an underlying primary immunodeficiency. Among the genes screened (Supplementary Table 2) there were two hemizygous nonsense alleles in XIAP (p.R222X) and SH2D1A (p.R75X) in two patients (Table 1) of European and African descent, respectively, and a homozygous missense change in CYBA (p.S118R), shared by two South Asian siblings (Table 1). These changes reflect conserved amino acids (GERP scores = 3.8, 2.7, 4.5, respectively), high CADD scores (37, 38, 25, respectively), and there are no recessive carriers of these alleles in 156,000 ethnically diverse population sequences (including ExAC and gnomAD), indicating that these mutational events are exceptionally rare in the population (Table 1). A list of likely benign variants, or variants of uncertain significance identified in the 67 screened monogenic IBD genes can be found in the Supplementary Note 2. The majority of the 67 genes were sequenced at a depth (mean = 67x) comparable to the rest of the exome (mean = 69x, Supplementary Fig. 3). In keeping with previous reports13, two of the genes (NCFI and IKBKGD) had poor coverage (<10x) following WES, and thus pathogenic variants in these genes may have been missed.

Mutations in XIAP have been associated with X-linked lymphoproliferative syndrome 2 (XLP2, MIM: 308240). The SH2D1A nonsense allele (p.R75X) detected in our patient truncates the protein in the middle of the SH2 domain, the critical region for signal transduction32. The truncation in SH2D1A was confirmed by FACS, showing absent binding of a C-terminal detection antibody (Fig. 2b). The patient died due to severe EBV infection and liver failure, a severe phenotype previously described in XLP1 patients33.

Mutations in CYBA, which encodes p22phox, cause chronic granulomatous disease (CGD, MIM: 233690). The p.S118R variant observed in both our patients is located within the putative transmembrane-spanning domain of the protein, where the majority of missense pathogenic variants have been found34–36. Functional impairment was confirmed using the dihydrodorhamidine-1,2,3 (DHR) assay using several stimuli (formylpeptide, E. coli and PMA; Fig. 2c), which showed impaired superoxide production in neutrophils from both siblings. A similar homozygous mutation in CYBA (p.S118R) has been described previously in a patient with CGD37.

Exome-wide screening for recessive loss-of-function variants. In addition to screening for genetic defects that have previously been described in patients with IBD-like intestinal inflammation, we searched exome-wide, for homozygous, or potential compound heterozygous, or hemizygous essential loss-of-function variants in our VEO-IBD cohort. This analysis revealed a homozygous nonsense variant in PCSK1 (p.R391X), which was absent from gnomAD and affected a highly conserved nucleotide (GERP score of 5.4). PCSK1 (Proprotein Convertase Subtilisin/Kexin type 1) encodes the proprotein convertase enzyme which cleaves prohormones and defects in the gene have been linked to an endocrinopathy syndrome characterized by diarrhea but rarely intestinal inflammation. The patient harboring the PCSK1 variant was of Asian ancestry and presented with indeterminate mild colitis before the age of one when recruited to the study. Interestingly, the initial intestinal inflammation did not progress but the phenotype changed over time. After recruitment and submission of the DNA sample for sequencing, the phenotype evolved towards growth delay, excessive weight gain, and endocrine disorders including diabetes and hypothyroidism, hypogonadism, cryptorchism, cortisol deficiency, and chronic lung disease. Whereas the initial phenotype was uncharacteristic, the subsequent syndromic findings are fully explained by PCSK1 deficiency. This finding highlights the value of next generation sequencing as a predictive diagnostic tool, as well as the need to take phenotype progression into account. No other likely essential loss-of-function variants were identified in our cohort.
Testing for a rare variant burden in pathways and genesets. We next searched for a rare variant enrichment across multiple related genes, such as those that reside in the same biological pathway (see Methods section and Supplementary Table 3). This approach offers additional statistical power compared to individual gene tests, as a larger number of variants are collapsed across a larger testing unit. In total we tested 195 different biological genesets, 186 of which represent the whole set of KEGG pathways available in the KEGG pathway database. No geneset or pathway showed a significant burden of rare variants in patients versus controls (N = 3855 INTERVAL samples) after correction for multiple testing (PLINK/SEQ burden test statistic P < 3.2 × 10⁻⁵, Supplementary Fig. 8).

No evidence of a rare variant burden in PID loci in VEO-IBD. A previous WES study of 125 VEO-IBD children and 145 healthy controls reported an over-representation of rare, damaging variation (AF < 0.1%) in PID-associated genes in VEO-IBD patients (P < 1 × 10⁻⁴). We found no evidence (PLINK/SEQ burden test statistic P = 0.7) of such an enrichment in our VEO-IBD cohort when compared against a larger set of control sequences (N = 3855 INTERVAL samples; Supplementary Fig. 8) and while controlling for potential confounding factors (such as ancestry and sequencing depth) between cases and controls (see Methods section).

A polygenic component operates in VEO-IBD risk. We next searched for a polygenic component underlying the disease using our genome-wide genotyping data. We calculated a PRS for each VEO-IBD COLORS patient of European ancestry (N = 99) by weighting their risk allele count at each disease-associated SNP (N_C = 147; N_UC = 119) by the estimated effect size of the risk allele (see Methods section). The CD and UC risk polygenic risk scores were significantly greater in our VEO-IBD cases compared to a large cohort of population controls (N = 18,780) (Student’s t-test P = 3.97 × 10⁻¹⁰ and P = 1.23 × 10⁻¹⁰ for CD and UC, respectively) (Fig. 3). We did not detect a significant difference between the risk scores of our VEO-IBD cases and a cohort of 13,896 predominantly adult-onset IBD cases ascertained by the UK IBD Genetics Consortium (Student’s t-test P = 0.457 and P = 0.661 for CD and UC, respectively) (Fig. 3). We can conclude that if there is a difference in mean PRS between VEO-IBD and adult-onset disease then it must be smaller than that seen between adult-onset IBD cases and controls (Cohen’s d = 0.85 and 0.64 for CD and UC, respectively) because our sample sizes would provide 100% power to detect such an effect. However, with our sample sizes we cannot rule out the possibility of a smaller difference in mean PRS between VEO-IBD and adult-onset IBD cases (Cohen’s d = 0.2) because 310 VEO-IBD cases would be required to provide 80% power to detect such an effect, given our sample size of

| Table 1 Pathogenic variants identified in VEO-IBD patients upon screening of known IBD-associated Mendelian disorder genes. |
|---|
| Gene | Position | Variant | GTs | GERP | CADD | GM | IT | Ethnicity | Phenotype |
|---|---|---|---|---|---|---|---|---|---|
| XAP | X:123602976 X:123590407 | ENST0000036919:7 ENST0000036502:4 | Hem | 3.8 | 37 | 0 | 0 | EU | CD-like phenotype with a severe fistulizing perianal phenotype |
| SF22A | X:123651407 | ENST0000036260:2 R7F | Hem | 2.7 | 38 | 0 | 0 | African | Acute EBV infection and liver failure |
| CYBA | 16:82712540 | ENST00000361623:8 | Hom | 4.5 | 1 | 0 | 0 | South Asian | Granulomas and a non-stricturing, non-penetrating CD-like pathology |
| CYBB | X:37666337 | ENST00000367851:4 | Hem | 5.6 | 40 | 1 | 0 | EU | CD (perianal disease), Hidradenitis suppurativa |
| XIAP | X:37663371 | ENST00000378588:4 | Hom | 4.5 | 1 | 0 | 0 | EU | Acute EBV infection and liver failure |

Each row represents a variant in a conserved site (GERP > 2), predicted damaging by in-silico tools, identified in VEO-IBD cases. Patient genotypes (GTs) are listed (Hem: hemizygous, if male; Hom: homozygous for the alternative allele). The number of gnomAD (GM) and INTERVAL (IT) samples that harbored similar variants in that gene (i.e., nonsense alleles) with the same genotype as our patients are also listed. Patient ADD refers to age at diagnosis (in years). Ethnic origin of patients as confirmed via vCAFA (see Methods section). GERP scores in table refer to C-scaled scores. CD and UC disease. CD, Crohn’s disease. UC, ulcerative colitis (for variant severity and minor allele frequency; see Methods for definitions and the Methods). No individual gene achieved exome-wide significance (Fisher’s exact test P < 1.7 × 10⁻⁵ after correction for multiple testing) irrespective of the variant inclusion criteria (Supplementary Fig. 7).

Searching for VEO-IBD genes: gene-based analysis. To identify previously unreported genes involved in VEO-IBD, we searched exome-wide for genes with a significant difference in the burden of rare variants in our cases versus a large cohort of exome-sequenced controls (N = 3855 INTERVAL samples; see Methods section). This approach has the advantage of allowing variants across the penetrance spectrum to contribute to the association test (see Methods section). Nine different exome-wide, gene-based screens were conducted using different variant inclusion criteria (for variant severity and minor allele frequency; see Methods for definitions and the Methods). No individual gene achieved exome-wide significance (Fisher’s exact test P < 1.7 × 10⁻⁵ after correction for multiple testing) irrespective of the variant inclusion criteria (Supplementary Fig. 7).
adult-onset patients and an α of 0.01 (Supplementary Fig. 9). Our PRS results were not driven by the few loci known to be associated with age-at-diagnosis of IBD (NOD2, HLA, and MST1) and similar findings were obtained when restricting the CD polygenic burden test to VEO-IBD cases defined in the COLORS cohort as CD or CD plus undeterminable IBD (IBDu), and when restricting the UC polygenic burden test to VEO-IBD cases defined as UC or UC plus IBDu (Supplementary Fig. 10).

To validate these results we generated polygenic risk scores across another cohort of European-descent VEO-IBD cases (Toronto SickKids VEO-IBD, N = 117) and population controls (NIDDK, N = 2603) with existing genotype data (see Methods).

**Fig. 2 Functional validation of pathogenic variants identified in monogenic IBD genes.** a Defective MDP response in a patient with hemizygous XIAP p.R222X. MDP (muramyl dipeptide) induced intracellular TNF response was determined using FACS. b Absent SAP staining (gene product of SH2D1A) as indicated by C-terminal antibody in a patient with hemizygous SH2D1A p.R75X. Measured with fluorescence-activated cell sorting (FACS). c Defecting ROS production in neutrophils from patients with homozygous CYBA p.S118N variants. Dihydrorhodamine-1,2,3 (DHR) flow cytometry assay (FACS) was performed to measure NADPH oxidase activity in response to PMA, E. coli particles and formylpeptide.
and AMP criteria, which requires the variant to be absent or have a measurable effect on protein function. Specifically, HSCT was initiated in the patient diagnosed with XIAP deficiency. This patient experienced multiple gastrointestinal operations and wound healing problems but has not yet developed EBV-triggered immunopathology (e.g., hemophagocytic lymphohistiocytosis). Routine infection screening to detect this potentially fatal complication that commonly develops in XLP2 patients will now be part of his clinical follow-up. Furthermore, the two siblings with pathogenic variants in the NADPH oxidase genes (CYBA) were referred to an immunology clinic and stem cell transplantation is currently being evaluated. The genetic diagnosis of SH2D1A deficiency was confirmed by the XLP1-specific extraintestinal complications of EBV-driven disease and liver failure. The importance of early genetic screening in VEO-IBD children to allow assessment before full manifestation of complications, or need for surgery, was highlighted by the lethal outcome for the patient with SH2D1A deficiency (XLP1, p.R75X) while WES was being undertaken in this patient.

The patients with XIAP, SH2D1A, and CYBA defects and the patient with CYBB mosaicism illustrate the opportunities and challenges with genetic diagnosis in patients with extreme phenotypes of IBD, where rare genetic variants can cause both immunodeficiency, such as invasive infections, and tissue inflammation. In light of the variable manifestation of the phenotype (some patients with IBD have not presented with the classical immunodeficiency phenotypes) the intestinal inflammation in these patients might either be regarded as a manifestation of the

Fig. 3 Distribution of CD and UC risk scores in VEO-IBD, CD, UC cases and healthy controls. The CD score was calculated using 147 CD risk alleles and the UC score using 119 UC risk alleles. Both scores were generated for a discovery cohort comprising 99 VEO-IBD cases (VEO-COLORS), 7578 CD cases, 6318 UC cases, 18,780 UK population controls (from the UKIBDGC), all of European ancestry. The replication cohort comprised 117 VEO-IBD cases (VEO-Toronto) and 2603 population controls (from the NIDDK Genetics Consortium). The CD and UC risk scores did not significantly differ between the two VEO-IBD cohorts (CD P = 0.98; UC P = 0.64). The Student’s t-test was used in group comparisons.
immunodeficiency (such as XIAP deficiency, XLPI, or chronic granulomatous disease) or as a form of monogenic IBD.

Somatic mosaicism in CYBB has previously been reported in two patients with clinical presentation of chronic granulomatous disease, including liver and perianal abscesses and lymphadenitis40. Yamada et al. identified the mosaic events following functional tests of neutrophil function warranted by the presentation of CGD. Our study identified the somatic mutation in CYBB via whole-exome sequencing and showcases how this approach can be used to find pathogenic mutations that are not indicated by the clinical presentation of a patient. The two patients reported by Yamada et al. only had 0 and 1.6% wild-type PBMCs (but 1.8 and 18.8% wild-type cells from buccal swabs) and were diagnosed with CGD. Our patient had 30% wild-type PBMCs and showed no symptoms of CGD, suggesting that having around ~30% functional phagocytes is sufficient to prevent these life-threatening infections (but >2% is insufficient). This is an important insight into the potential utility of gene therapy for treatment of CGD; correcting the genetic sequence in around 30% of phagocytes could be sufficient to prevent serial life-threatening infections, but unlikely to reduce the risk of intestinal inflammation. The mosaic patient herein reported adds to an expanding spectrum of NADPH oxidase deactivation. The mosaic patient herein reported adds to an expanding infections, but unlikely to reduce the risk of intestinal in

Treatment of CGD; correcting the genetic sequence in around 30% of patients showed no symptoms of CGD, suggesting that having around 1.8 and 18.8% wild-type cells from buccal swabs) and were compared to patients with type 1 diabetes43. However, for two patients with clinical presentation of chronic granulomatous disease) or as a form of monogenic IBD.

It remains to be seen if the extent to which the polygenic background influences VEO-IBD susceptibility varies depending on the monogenic “foreground”. For example, the polygenic burden of IBD risk alleles may be of little phenotypic consequence in the presence of fully penetrant VEO-IBD mutations (such as those seen in the IL10 signaling pathway). This model would match that seen in patients with diabetes, where individuals with monogenic MODY have lower polygenic risk score compared to patients with type 1 diabetes43. However, for patients with a monogenic VEO-IBD diagnosis that demonstrates incomplete penetrance for VEO-IBD, the polygenic burden of IBD risk could play a more significant role in defining the phenotype. The recent observation that patients with chronic granulomatous disease (CGD) and intestinal inflammation have a greater burden of classical IBD risk variants compared to CGD patients without intestinal inflammation44 supports this general hypothesis. We show that, at least for individuals with VEO-IBD and no current genetic Mendelian diagnosis, common genetic variants associated with susceptibility to Crohn’s disease and ulcerative colitis do play a role in disease. It has recently been shown that there is a weak, but statistically significant, inverse correlation between age of IBD onset and burden of IBD associated risk alleles27–29. We found no evidence to support the hypothesis that VEO-IBD is due to an increased burden of common IBD susceptibility alleles relative to adult-onset disease, but larger sample sizes than ours are required to powerfully search for small differences in PRS. This model applies in familial hypercholesterolemia, a dominantly-inherited disorder where individuals with very high LDL-cholesterol and no known monogenic cause of disease have a particularly high burden of common cholesterol-increasing alleles45. The recent identification of common genetic variants that are associated with Crohn’s disease prognosis but not susceptibility also suggests that the search for disease age-at-onset loci should be performed genome-wide and not just restricted to known IBD susceptibility loci27.

In summary, our data show that primary immunodeficiencies caused by rare genetic variants can be found in some VEO-IBD patients even if no Mendelian disease was clinically suspected, suggesting that genetic screening is relevant across this entire patient group. We implicate cellular mosaicism with Mendelian disorder-associated variants as a possible mechanism underlying VEO-IBD, as highlighted by our case study. Finally, we show that whatever factors are driving an early age at disease onset in individuals without a conclusive Mendelian diagnosis, in the majority of patients they do so on a polygenic background similar to classical IBD.

Methods

Patient and controls samples. The study was approved by the North Staffordshire Research Ethics Committee (REC: 09/H1204/30; subject COLORS in IBD) and local ethics committees at the study sites. All patients, or their parents, gave written informed consent before enrollment. The somatic mosaic patient consented to open access publication of a detailed case report including genetic, clinical, laboratory data as well as endoscopic and histological images. The cohort consists of 146 VEO-IBD (91 female) – either without a previous genetic diagnosis, recruited as part of the COLORS IN IBD project (COLsrcs of early Onset Rare disorders). Samples were referred from participating centers in the United Kingdom, Switzerland, Poland and Germany. All patients had a confirmed diagnosis of IBD by standard methods, including endoscopic, radiologic, laboratory, and clinical evaluation (ESPGHAN guidelines46). Phenotypic status was based on the Paris Classification47. Patients were selected according to age-at-diagnosis (<7 years, age of symptom onset <6 years) and the severity of the IBD phenotype, as indicated by need for surgery and/or therapy progression to biologics or immunomodulators. When a clinical diagnosis of a known Mendelian disease was suspected (e.g., IL10, IL10RB, < 18.8% wild-type cells from buccal swabs) and were compared to patients with type 1 diabetes43. However, for two patients with clinical presentation of chronic granulomatous disease) or as a form of monogenic IBD.

The recent identification of common genetic variants that are associated with Crohn’s disease and ulcerative colitis do play a role in disease. It has recently been shown that there is a weak, but statistically significant, inverse correlation between age of IBD onset and burden of IBD associated risk alleles27–29. We found no evidence to support the hypothesis that VEO-IBD is due to an increased burden of common IBD susceptibility alleles relative to adult-onset disease, but larger sample sizes than ours are required to powerfully search for small differences in PRS. This model applies in familial hypercholesterolemia, a dominantly-inherited disorder where individuals with very high LDL-cholesterol and no known monogenic cause of disease have a particularly high burden of common cholesterol-increasing alleles45. The recent identification of common genetic variants that are associated with Crohn’s disease prognosis but not susceptibility also suggests that the search for disease age-at-onset loci should be performed genome-wide and not just restricted to known IBD susceptibility loci27.

In summary, our data show that primary immunodeficiencies caused by rare genetic variants can be found in some VEO-IBD patients even if no Mendelian disease was clinically suspected, suggesting that genetic screening is relevant across this entire patient group. We implicate cellular mosaicism with Mendelian disorder-associated variants as a possible mechanism underlying VEO-IBD, as highlighted by our case study. Finally, we show that whatever factors are driving an early age at disease onset in individuals without a conclusive Mendelian diagnosis, in the majority of patients they do so on a polygenic background similar to classical IBD.

Methods

Patient and controls samples. The study was approved by the North Staffordshire Research Ethics Committee (REC: 09/H1204/30; subject COLORS in IBD) and local ethics committees at the study sites. All patients, or their parents, gave written informed consent before enrollment. The somatic mosaic patient consented to open access publication of a detailed case report including genetic, clinical, laboratory data as well as endoscopic and histological images. The cohort consists of 146 VEO-IBD (91 female) – either without a previous genetic diagnosis, recruited as part of the COLORS IN IBD project (COLsrcs of early Onset Rare disorders). Samples were referred from participating centers in the United Kingdom, Switzerland, Poland and Germany. All patients had a confirmed diagnosis of IBD by standard methods, including endoscopic, radiologic, laboratory, and clinical evaluation (ESPGHAN guidelines46). Phenotypic status was based on the Paris Classification47. Patients were selected according to age-at-diagnosis (<7 years, age of symptom onset <6 years) and the severity of the IBD phenotype, as indicated by need for surgery and/or therapy progression to biologics or immunomodulators. When a clinical diagnosis of a known Mendelian disease was suspected (e.g., IL10, IL10RB, 

Sample QC. Cross-sample contamination was evaluated using VerifyBAMID (version 1.1.0). No case samples showed evidence of contamination, however 113 controls had a FREEMIX fraction >3% and were thus excluded from the study. Samples with a mean genotype quality (GQ) < 85.4 (representing 3 s.d. from the mean), a depth < 40, a missing genotype rate >0.2%, or controls with a close familial relationship (Pihat > 0.125) were removed from further analysis. A total of 468 poor quality samples (one case and 467 controls) were excluded from subsequent analyses.

Ancestry analysis. Cases and controls were assigned to ancestry-matched groups based on principal components inferred from the 2504 individuals in the 1000 Genomes Project (1KG) phase 3 data44. Overall, 104 cases and 3787 controls were defined as Europeans, and 21 cases and 68 controls as South Asians. The remaining 20 cases and 115 controls did not cluster with a clearly defined ancestral group. In total, 145 of 146 cases and 3969 of 4436 controls remained after QC (Supplementary Fig. 4).

Variant QC. Individual variants were evaluated using Variant Quality Score Recalibrator (VQSR) using the recommended training sets50 and a 99.9% sensitivity tranche. Individual genotypes with GQ < 20 or depth < 8 were set to missing, and with a close familial relationship (Pihat > 0.125) were removed from further analysis. A total of 468 poor quality samples (one case and 467 controls) were excluded from subsequent analyses.

Ancestry analysis. Cases and controls were assigned to ancestry-matched groups based on principal components inferred from the 2504 individuals in the 1000 Genomes Project (1KG) phase 3 data54. Overall, 104 cases and 3787 controls were defined as Europeans, and 21 cases and 68 controls as South Asians. The remaining 20 cases and 115 controls did not cluster with a clearly defined ancestral group. In total, 145 of 146 cases and 3969 of 4436 controls remained after QC (Supplementary Fig. 4).

Variant QC. Individual variants were evaluated using Variant Quality Score Recalibrator (VQSR) using the recommended training sets50 and a 99.9% sensitivity tranche. Individual genotypes with GQ < 20 or depth < 8 were set to missing52. Variants with more than 10% missing genotypes in either cases or controls were also removed. In total, 1,267,058 variants passed QC.

Variant annotation. The exome data was annotated using several resources: dbSNP v137 rsIDs and allele frequencies (AFs) from 1KG phase 1 (N = 2818)53. NHLBI GO Exome Sequencing Project 6500i (ESP, N = 6500i54, the UK10K low-coverage study (UK10K WGS, N = 3781), UK10K WES samples (UK10K WES,
NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-14275-y
ARTICLE

N = 4975)35, the Exome Aggregation Consortium n0.3 (ExAC, N = 60,706)36 and the Genome Aggregation Database (gnomAD, N = 141,356)36. Functional annota-
tions were added using the Ensembl Variant Effect Predictor (VEP) version 95
according to Genecode v19 coding transcripts, using the most severe consequence
on the gene. Deletions/insertions scores for missense variants were inferred using the
SIFT27 and PolyPhen-216 algorithms. Conserved amino acids were identified using
GERP67.

Rare variants were defined as those that were absent or had AFs <1% in
reference datasets. We defined two broad categories of variants, based on predicted
functional impact at the protein level: (1) Functional: transcript ablation, stop
insertion/deletion, frameshift, indegree, frameshift, splice donor/acceptor, porphyrin, and
missense variants and (2) Disruptive: nonsense, frameshift and splice acceptor/donor variants or missense variants with a
PolyPhen-2 and SIFT pathogenicity predictions of “possibly damaging” or “deleterious” (or greater) and a
GERP score ≥ 2.

Screening of known IBD-associated Mendelian disorder genes. A list of 67
genes known to be associated with Mendelian disorders with IBD-like inflammato-
ry phenotypes was ascertained from the literature28,58. The mode of inheritance
and associated disorders for these genes are listed in Supplementary Table 2. We
screened these genes for rare variants consistent with the established
mode of inheritance. For variants on the X chromosome, we also assessed possible
instances of non-classical inheritance, such as non-random X-inactivation events in
females, or somatic mosaicism in males. To assess the pathogenicity of the variants
identified and to adhere to current best practices57,58, we checked that the patient’s
template data that reported for the Mendelian disorder, and associated disorders for these genes are listed in Supplementary Table 2. We
screened these genes for rare variants consistent with the established
mode of inheritance. For variants on the X chromosome, we also assessed possible
instances of non-classical inheritance, such as non-random X-inactivation events in
females, or somatic mosaicism in males. To assess the pathogenicity of the variants
identified and to adhere to current best practices57,58, we checked that the patient’s

Validation of polygenic risk scores in VEO-IBD. To validate the risk scores
obtained in our VEO-IBD COLORS cohort, we used a cohort of 117 VEO-IBD
patients from the SickKids Toronto IBD cohort, and a cohort of population con-

NATURE COMMUNICATIONS

| https://doi.org/10.1038/s41467-019-14275-y

where xi is the number of risk alleles at SNP, carried by individual i, and ORi is the
allelic odds ratio at SNP, as estimated by Liu et al. or De Lange et al.21,22. Risk scores
were compared between groups using the Student’s t-test and assuming a sig-
ificance threshold of P < 8 × 10⁻³, which accounts for the six pairwise compar-
isons between VEO-IBD cases, IBD cases and controls, in CD and UC. Power
estimations were performed using the R package pwr. Cohen’s d was estimated using
the R package psr.

DHR FACS to measure NADPH activity. The neutrophil oxidative burst assay to
detect reactive oxygen species by DHR FACS assay was performed using standard
methods38. Firstly, EDTA blood was incubated with DHR-123 (Life Technolo-
gies, D23806) at 2.5 μg/mL for 15 min at 37 °C followed by PMA (Sigma, P1585) at
100 ng/mL stimulation and FACS staining. DHR response was measured in FSC/
fluorescence observed in the

Genome-wide genotyping of VEO-IBD cases. VEO-IBD cases were genotyped using
the Illumina Infinium Core Exome v2-12 chip. Poorly genotyped SNPs,
deidentified in exac with a missing genotype rate >5% of high-impact
variants were removed. The European-descent VEO-IBD samples with autosomal heterozygosity rates >3 s.d. from the
mean were excluded. Data were phased using SHAPEIT v244 and imputation performed using IMPUTE v276 with 5010 haplotypes from the 1kg phase 3 data24. Genotypes with an INFO score below 0.9 were excluded from the dataset.

Polymorphic risk score calculation. We assessed the contribution of IBD loci
associated with adult-onset IBD to VEO-IBD32. This analysis made use of the
genoeration-wide genotyping and imputation data, rather than our exome-sequences,

CHROMOSOMES | (2020) 11:995 | https://doi.org/10.1038/s41467-019-14275-y | www.nature.com/naturecommunications

NATURE COMMUNICATIONS
from the patient were sorted into a DHR-low and DHR-high population (BD FACs Aria III). Sorting performed with an 85 micron nozzle.

Miseq analysis of flow-sorted neutrophils. The proportion of the flow-sorted cells carrying the mutation was determined by NGS in a targeted 3-PCR approach. The primers were designed using the Eurofins Genomics PCR Primer Design Tool or provided by Illumina Inc. and synthesized by Sigma-Aldrich (Suffolk, United Kingdom). The PlatinumTaq DNA Polymerase High Fidelity Kit (11304-011, Invitrogen) was used according to the instructions of the manufacturer, with a final reaction volume of 0.2 µL, 25 cycles of annealing temperature (T_a) of 60°C, unless otherwise stated. A successful PCR was confirmed after each reaction by detecting an appropriately sized band in a 2% agarose gel. The first primer pair (F: AACTCACCCTTCTCCAAACCATC; R: ACTTGGGCTTGACCCCTAC) amplified a 513 bp fragment surrounding the variant of interest from 40 ng of genomic DNA. The second primer pair (F: ACATCTCTTCTCAACGTGCTCGTGTGCTAATTTGTTACATACGAGATC; R: TCTCGGATTTCGTCAAGCAGGCTTCTGACATGCTTTTACACTGACATT) used 0.5 µL of the resulting amplicon to amplify a 258 bp fragment composed of 192 bp of the template surrounded by the Illumina Adapter 5' and 3' overhangs. The diluted PCR products were then directly used for each additional cycle of PCR using Phusion High Fidelity DNA Polymerase kit and (#M0530L, Beckman Colter) at 98°C for 30s, 72°C for 30s, and 70°C for 30s. The proportion of the

Gentamicin protection assay. The neutrophil gentamicin protection assay was adapted from Riffelmacher et al., 201770. Briefly, neutrophils were sorted into RPMI11640 (Sigma) and 10% FCS (Sigma). Monocyte derived macrophages were generated from PBMCs as previously described51. Forty-four thousand neutrophils or macrophages were infected in a 9-well round-bottom plate or flat-bottom plate, respectively, at a MOI 1:10 using Salmonella enterica serovar Typhimurium for 45 min or 1 h, respectively. Cells were treated with 100 µg/ml gentamicin (Sigma) for a further 45 min for neutrophils, or 2 h for macrophages. Cells were then lysed in 1% Triton X-100 (Sigma) in H2O. Lysates were plated on LB agar plates using the track method and CFU were quantified the following day.

MDP induced intracellular TNF response. Quantification of intracellular TNF in monocytes was performed as previously described51,71. Briefly, freshly isolated PBMC were rested overnight in RPMI 10% FCS. Unstimulated cells were compared with MDP and LPS (lipopolysaccharide) stimulated cells. Intracellular TNF (MAB11, ebioscience) was detected in CD14+HLADR+monocytes (CD14 (M5E2) and HLA-DR (L243), both BioLegend). Viable cells were detected using live-dead staining (Fixable Viability Dye, ebioscience).

SAP expression. FACS was used to quantify SAP protein expression encoded by SHZD1A as previously described27. In brief, blood cells were incubated with mouse anti-human CD3 (BD Bioscience, Clone SK7, cat 354767), mouse anti-human CD8-PerCP (BD Bioscience, Clone SK1, cat 354774), and mouse anti-human CD4-PE (BD Bioscience, Clone MY31, cat 354810). Samples were then fixed, washed, and permeabilized. The anti-SAP antibody (Stratech Scientific Biosciences; clone 1C9, cat H00004068-M01) or isotype control antibody (IgG1 isotype control; BD Biosciences 349040). Samples were again washed and stained with anti-IgG1-FITC (Dako; F0479) before FACS analysis.

Reporting summary. Further information on research design is available in the Nature Reporting Summary linked to this article.

Data availability. Sequencing and genotyping data that supports this study have been deposited to the European Genome-phenome Archive (EGA) under the accession code EGAS00001005513 and EGAS00001009294, respectively. All other data are contained in the paper and its supplementary information or available upon request.

Received: 12 April 2018; Accepted: 20 December 2019; Published online: 21 February 2020

References

1. Khair, B., Gardert, A. & Xavier, R. J. Genetics and pathogenesis of inflammatory bowel disease. Nature 474, 307–317 (2011).
2. Peolquin, J. M., Goel, G., Villablanca, E. J. & Xavier, R. J. Mechanisms of pediatric inflammatory bowel disease. Annu. Rev. Immunol. 34, 31–64 (2016).
3. Molodecky, N. A. et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 142, 46–54.e12 (2012).
4. Cosnes, J., Gowerrouxse, C., Seksik, P. & Cortot, A. Epidemiology and natural history of inflammatory bowel diseases. Gastroenterology 140, 1785–1794 (2011).
5. Benchimol, E. I. et al. Epidemiology of pediatric inflammatory bowel disease. Inflamm. Bowel Dis. 17, 423–439 (2011).
6. Ruenmele, F. M. et al. Characteristics of inflammatory bowel disease with onset during the first year of life. J. Pediatr. Gastroenterol. Nutr. 43, 603–609 (2011).
7. Thapar, N., Shah, N., Ramsay, A. D., Lindley, K. J. & Mills, P. J. Long-term outcome of intractable ulcerating enterocolitis of infancy. J. Pediatr. Gastroenterol. Nutr. 40, 582–588 (2005).
8. Uhlig, H. H. Monogenic diseases associated with intestinal inflammation: implications for the understanding of inflammatory bowel disease. Gut 62, 1795–1805 (2013).
9. Uhlig, H. H. et al. The diagnostic approach to monogenic very early onset inflammatory bowel disease. Gastroenterology 147, 990–1007.e3 (2014).
10. Glocker, E.-O. et al. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. N. Engl. J. Med. 361, 2033–2045 (2009).
11. Glocker, E.-O. et al. Infant colitis—it’s in the genes. Lancet 376, 1277 (2010).

ARTICLE

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-14275-y | www.nature.com/naturecommunications
12. Avitzur, Y. et al. Mutations in tetratricopeptide repeat domain 7A result in a severe form of very early onset inflammatory bowel disease. *Gastroenterology* **140**, 1023–1039 (2011).

13. Kammermeier, J. et al. Targeted gene panel sequencing in children with very early onset inflammatory bowel disease: evaluation and prospective analysis. *J. Med. Genet.* **51**, 748–755 (2014).

14. Dinwiddie, D. L. et al. Molecular diagnosis of infantile onset inflammatory bowel disease by whole sequencing. *Nat. Genet.* **42**, 102–104 (2013).

15. Worthy, E. A. et al. Making a definitive diagnostic: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* **13**, 255–262 (2011).

16. Denson, L. A. et al. Clinical and genomic correlates of neutrophil reactive oxygen species production in pediatric patients With Crohn’s disease. *Gastroenterology* **140**, 2097–2110 (2011).

17. Amininejad, L. et al. Analysis of genes associated with monogenic primary immunodeficiency identifies rare variants in XIAP in patients with Crohn’s disease. *Gastroenterology* **154**, 2165–2177 (2018).

18. Barzaghi, F., Passerini, L. & Racchetta, R. Immune dysregulation, polyendocrinopathy, enteropathy, x-linked syndrome: a paradigm of immunodeficiency with autoimmunity. *Front. Immunol.* **3**, 211 (2012).

19. Ono, S. et al. Hematopoietic stem cell transplantation for XIAP deficiency in Japan. *J. Clin. Immunol.* **37**, 85–91 (2017).

20. Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979 (2015).

21. de Lange, K. M. et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–262 (2017).

22. Luo, Y. et al. Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nat. Genet.* **49**, 186 (2017).

23. Kugathasan, S. et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979 (2015).

24. de Lange, K. M. et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 1118–1125 (2010).

25. Cutler, D. J., Zwick, M. E., Okou, D. T., Prahlad, S. & Walters, T. G. S. Dissecting allele architecture of early onset IBD using high-density genotyping. *PLOS ONE* **10**, 1–12 (2015).

26. Ananthakrishnan, A. N. et al. Differential effect of genetic burden on disease manifestations, management and outcome of the disease. *Blood* **117**, 53–62 (2011).

27. Rae, J. et al. Molecular analysis of 9 new families with chronic granulomatous disease caused by mutations in CYBA, the gene encoding p22(phox). *Blood* **96**, 1106–1112 (2000).

28. Roos, D. & de Boer, M. Molecular diagnosis of chronic granulomatous disease. *Clin. Exp. Immunol.* **175**, 139–149 (2014).

29. Dhillon, S. S. et al. Variants in nicotinamide adenine dinucleotide phosphate oxidase complex components determine susceptibility to very early onset inflammatory bowel disease. *Gastroenterology* **147**, 680–689 (2014).

30. Dinauer, M. C., Pierce, Ea, Bruns, Ga. P., Cumutte, J. T. & Orkin, S. H. Human neutrophil cytochrome b light chain (p22-phox). *J. Clin. Invest.* **86**, 1729–1737 (1990).

31. Kelsen, J. R. et al. Exome sequencing analysis reveals variants in primary immunodeficiency genes in patients with very early onset inflammatory bowel disease. *Gastroenterology* **149**, 1415–1425 (2015).

32. Richards, 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.* **17**, 405–424 (2015).

33. Yamada, M. et al. Somatic mosaicism in two unrelated patients with X-linked chronic granulomatous disease characterized by the presence of a small 15q21.3 deletion. *J. Genet.* **97**, 110–115 (2012).

34. van de Geer, A. et al. Inherited p40phox deficiency differs from classic chronic granulomatous disease. *J. Clin. Invest.* **128**, 3957–3975 (2018).

35. Battersby, A. C., Cale, C. M., Goldblatt, D. & Genney, A. R. Clinical manifestations of disease in X-linked carriers of chronic granulomatous disease. *J. Clin. Immunol.* **33**, 1276–1284 (2013).

36. Patel, K. A. et al. Type 1 diabetes genetic risk score: a novel tool to distinguish patients with polygenic and monogenic familial hypercholesterolemia: a case-control study. *Lancet* **381**, 1293–1301 (2013).

37. Levine, A. et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J. Pediatr. Gastroenterol. Nutr.* **58**, 795–806 (2014).

38. Levine, A. et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflamm. Bowel Dis.* **17**, 1314–1321 (2011).

39. Moore, C. et al. Recruitment and representativeness of blood donors in the INTERVAL randomised trial assessing varying inter-donation intervals. *Trials* **17**, 1–12 (2016).

40. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

41. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498 (2011).

42. Aulton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

43. Carson, A. R. et al. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMJ Bioinformatic*. **15**, 1–15 (2014).

44. Project, G., Asia, E., Africa, S., Figs, S. & Tables, S. An integrated map of genetic variation from 1092 human genomes. *Nature* **135**, 8–9 (2012).

45. Tennesen, J. et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).

46. Walter, K. et al. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).

47. McLaren, W. et al. Deriving the consequences of genomic variants with the ensemble API and SNP effect predictor. *Bioinformatics* **26**, 2069–2070 (2010).

48. Ng, P. C. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).

49. Adzhubei, I., Schmidt, S., Peshkin, L., Ramensky, V. & Sunyaev, S. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).

50. Davydov, E. V. et al. Identifying a high fraction of the human genome to be critical for normal neutrophil differentiation. *Science* **337**, 806–810 (2014).

51. Munoz, N. M. & Leff, A. R. Highly purified functional CD8 T cells from human peripheral blood by negative immunomagnetic selection. *J. Clin. Immunol.* **1573** (2010).

52. Purcell, S. M. et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).

53. Anderson, C. A. et al. Data quality control in genetic case-control association studies. *Nat. Protoc.* **5**, 1564–1573 (2010).

54. O’Connell, J. et al. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.* **10**, e1004324 (2014).

55. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).

56. Chen, Y. & Junger, W. G. Measurement of oxidative burst in neutrophils. *Methods Mol. Biol.* **844**, 115–124 (2012).

57. Munoz, N. M. & Leff, A. R. Highly purified selective isolation of eosinophils from human peripheral blood by negative immunomagnetic selection. *Nat. Protoc.* **1**, 2620–2627 (2006).

58. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

59. Rimmer, A. et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).

60. Hafner, T. et al. Autophagy-dependent generation of free fatty acids is critical for normal neutrophil differentiation. *Immunity* **47**, 466–480.e5 (2017).
Acknowledgements

We thank all individuals who contributed samples to the study. We acknowledge Satish Keshav who sadly died during the revision process of this manuscript and who was instrumental to setup the Oxford IBD cohort study. The COLORS in IBD project was supported by an ESPGHAN collaboration network grant and Crohn’s and Colitis UK (M/11/01). The genotyping and sequencing of the COLORS in IBD cohort was funded by the Wellcome Trust (098051; 093885/Z/10/2). Adult-onset IBD case collections were supported by Crohn’s and Colitis UK and sequencing was co-funded by the Wellcome Trust (098051) and the Medical Research Council, UK (MR/R00341X/1). We thank the investigators of the NIDDK IBD Genetics Consortium for sharing previously unpublished genotype data for a large cohort of population controls, used in our polygenic risk analyses. We acknowledge support of the Wellcome Trust (E.G.S., L.M., F.J.M., J.C.B. and C.A.A.; 098051), the Crohn’s & Colitis Foundation of America (H.H.U.), the Leona M. and Harry B. Helmsley Charitable Trust (H.H.U., A.M., S.S.), the Deutsche Forschungsgemeinschaft (T.S.), the Swiss National Science Foundation (C.P.B.), the Medical Research Council (A.A., T.A.F., DCW), the Medical Research Foundation (D.C.W.) and the Ovarian Cancer Action (A.A.). T.A.B. is supported by a Radcliffe Department of Medicine/MRC Scholars Program Studentship. A.M.M. is funded by a CIHR Grant and the Tier 1 Canada Research Chair in Pediatric IBD. The NIDDK IBD Genetics Consortium is supported by National Institute of Diabetes and Digestive and Kidney Diseases grants. We acknowledge the contribution of the Oxford Biomedical Research Center, which is supported by the National Institute for Health Research. We thank the INTERVAL Study for contributing sequencing data. INTERVAL was supported by: NHST, NIHR, NIHR BioResource, NIHR (Cambridge Biomedical Research Center at the Cambridge University Hospitals NHS Foundation Trust) [4], NIHR BTRU in Donor Health and Genomics, UK MRC, BHF, Wellcome Trust, and the ESRC. J.D. is funded by the NIHR [Senior Investigator Award]. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Author contributions

E.G.S., L.M., F.J.M., R.W., H.H.U. and C.A.A. analysed the sequence and genotype data. T.S., S.P., A.C., K.G., T.B., T.F., E.M.C., J.C.M. and A.A. performed functional validation experiments. E.C.S., T.S., J.K., N.C., C.P., A.R., R.R., F.B., M.A., R.H., M.Z., K.F., C.B., S.T., J.S., M.P., N.T., A.W., C.M., J.S., P.S., D.J.R., I.B., A.G., A.M., N.S., A.E., S.S., D.W. and H.H.U. contributed to recruitment of patient or control groups. E.G.S., H.H.U. and C.A.A. wrote the manuscript. All authors read and approved the final version of the manuscript. H.H.U. and C.A.A. jointly supervised and coordinated the project.

Competing interests

The authors have made the following disclosures: C.A.A. has received consultancy fees from Genomics plc and Illumina. H.H.U. received collaborative research support or consultancy fees from Eli Lilly, UCB Pharma, Celgene, Regeneron, Boehringer Ingelheim, Pfizer, and AbbVie. SPT has been adviser to, in receipt of educational or research grants from, or invited lecturer for AbbVie, Amgen, Asahi, Biogen, Boehringer Ingelheim, BMS, Cosmo, Elan, Enterome, Ferrimg, FPRT Bio, Genentech/Roche, Genzyme, Glenmark, GW Pharmaceuticals, Immunocore, Immunometabolism, Janssen, Johnson & Johnson, Lilly, Merck, Novartis, Novo Nordisk, Ocea, Pfizer, Shire, Santarus, SigmoidPharma, Synthon, Takeda, Tiltotts, Topivert, Trino Therapeutics with Wellcome Trust, UCB Pharma, Vertex, VHSquared, Vifor, Warner Chilcott, and Zeria. The remaining authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-14275-y.

Correspondence and requests for materials should be addressed to H.H.U. or C.A.A.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020, corrected publication 2022
COLORS in IBD group investigators

Marlen Zurek28, Caterina Strisciuglio29, Mamoun Elawad30 & Bernice Lo30

28St. Georg Hospital, Leipzig, Germany. 29Department of Translational Medical Sciences, Section of Pediatrics, University of Naples, Naples, Italy.
30Sidra Medical and Research Center, Doha, Qatar

Oxford IBD cohort study investigators

Carolina Arancibia-Carcamo2, Adam Bailey2, Ellie Barnes2, Elizabeth Louise Bird-Lieberman2, Oliver Brain2, Barbara Braden2, Jane Collier2, James East2, Lucy Howarth7, Satish Keshav2, Paul Kleereman2, Simon Leedham2, Rebecca Palmer2, Fiona Powrie2 & Alison Simmons2

INTERVAL Study

Matthew Walker15, Zoe Tolkien15, Stephen Kaptoge15, David Allen18, Susan Mehenny31, Jonathan Mant15, Emanuele Di Angelantonio15 & Simon G. Thompson15

31NHS Blood and Transplant, Longley Lane, Sheffield, UK

Swiss IBD cohort investigators

Bahtiyar Yilmaz32,33, Pascal Juillerat32,33, Markus Geuking32, Reiner Wiest32, Andrew J. Macpherson32,33, Francisco Damian Bravo33, Lukas Brügger33, Ove Carstens33, Ulrike Graf Bigler33, Benjamin Heimgartner33, Monica Rusticeanu33, Sybille Schmid (-Uebelhart)33, Bruno Strebel33, Aurora Tatu33, Radu Tutuian33, Reiner Wiest35, Ove Øyås34, Charlotte Ramon34, Jörg Stelling34, Yannick Franc35, Nicolas Fournier35, Valerie E.H. Pittet35, Bernard Burnand35, Mara Egger35, Yannick Franc35, Delphine Golay35, Astrid Marot35, Leilla Musso35, Valérie Pittet35, Jean-Benoit Rossel35, Viviane Seematter35, Joachim Sommer35, Rachel Vuilliamy35, Pierre Michetti36,37, Michel H. Maillard36,37, Céline Keller36, Michel H. Maillard36,37, Andreas Nydegger36,37, Alain Schoepfer37, Eva Archanioti37, Jessica Ezri37, Montserrat Fraga37, Alain Schoepfer37, Christoph Müller38, Gerhard Rogler39, Luc Biedermann39, Mirjam Blattmann39, Sabine Burk39, Barbara Dora39, Michael Fried39, Benjamin Misselwitz39, Beat Müllhaupt39, Nicole Obialo39, Daniel Pohl39, Nadia Raschle39, Gerhard Rogler39, Michael Scharl39, Stephan Vavricka39, Roland von Känel39, Jonas Zeit39, Karim Abdelrahman40, Gentiana Ademi41, Jan Borovicka41, Stephan Brand41, Remus Frei41, Johannes Haarer41, Christina Knellwolf (-Grieger)41, Claudia Krieger(-Grübel)41, Patrizia Künzler41, Christa Meyenberger41,
Pamela Meyer41, Nina Röhrich41, Mikael Sawatzki41, Martin Schelling41, Gian-Marco Semadeni41, Michael Sulz41, Dorothee Zimmermann41, Patrick Aepli42, Dominique H. Criblez42, Cyrill Hess42, Jean-Pierre Richterich42, Johannes Spalinger42, Dominic Staudemann42, Andreas Stulz42, Stefanie Wöhrl42, Amman Thomas43, Claudia Anderegg44, Henrik Köhler44, Rachel Kusche44, Anca-Teodora Antonino45, Eviano Arrigoni46, José M. Bengoa46, Sophie Cunningham46, Philippe de Saussure46, Laurent Girard46, Diana Bakker de Jong47, Polat Bastürk47, Simon Brunner47, Lukas Degen47, Petr Hruz47, Carolina Khalid-de Bakker47, Jan Niess47, Bruno Balsiger48, Janine Haldemann48, Gaby Saner48, Frank Seibold48, Peter Bauerfeind49, Andrea Becucci50, Dominique Belli50, Janek Binek51, Peter Hengstler51, Stephan Boehm52, Tujana Boldanov53, Patrick Bühr12, Rebekka Koller12, Vanessa Rueger12, Arne Senning12, Emanuel Burrr54, Sophie Buyse55, Dahlia-Thao Cao56, Fabrizia D’Angelo57, Joakim Delarive58, Christopher Doering59, Roxane Hessler59, Claudia Preissler60, Ronald Rentsch61, Branislav Risti62, Marc Alain Rit63, Michael Steuerwald63, Jürg Vögtlin63, Markus Sagmeister64, Bernhard Sauter65, Susanne Schibli66, Christiane Sokollik66, Johannes Spalinger66, Hugo Schlauri67, Jean-François Schnegg68, Mariam Seirafi69, Holger Spangenberger70, Philippe Stadler71, Peter Staub72, Volker Stenz73, Michela Tempia-Caliera74, Joël Thorens75, Kaspar Truninger76, Patrick Urfer77, Francesco Viani78, Dominique Vouillamo79, Silvan Zander80 & Tina Wyli81

UK IBD Genetics Consortium

L. Jostins82, N.A. Kennedy83, T. Ahmad83, C.A. Lamb84, C. Edwards85, A. Har86, C. Hawk87, J.C. Mansfield88, C. Mowat89, W.G. Newman90, A. Simmons2, M. Tremelling91, J.C. Lee14, N.J. Prescott92, C.G. Mathew92 & C.W. Lees93

82The Wellcome Trust Centre for Human Genetics, Oxford, UK. 83Precision Medicine Exeter, University of Exeter, Exeter, UK. 84Institute of Cellular Medicine, Newcastle University, Newcastle Upon Tyne, UK. 85Department of Gastroenterology, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. 86Department of Gastroenterology, Royal Hallamshire Hospital, Sheffield, UK. 87Nottingham Digestive Diseases Centre, Queens Medical Centre, Nottingham, UK. 88Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, UK. 89Department of Medicine, Ninewells Hospital and Medical School, Dundee, UK. 90Genetic Medicine, Manchester Academic Health Science Centre, Manchester, UK. 91Gastroenterology & General Medicine, Norfolk and Norwich University Hospital, Norwich, UK. 92Department of Medical and Molecular Genetics, King’s College London School of Medicine, Guy’s Hospital London, London, UK. 93Gastrointestinal Unit, Western General Hospital University of Edinburgh, Edinburgh, UK.
