ORIGINAL ARTICLE

Genetic analysis of four consanguineous multiplex families with inflammatory bowel disease

Noam Ben-Yosef [1,2,*], Matthew Frampton[1], Elena R. Schiff[3], Saleh Daher[2], Fadi Abu Baker[4], Rifaat Safadi[2], Eran Israeli[5], Anthony W. Segal[1,#] and Adam P. Levine[1,6,*,#]

[1]Centre for Molecular Medicine, Division of Medicine, University College London, London, UK, [2]Institute of Gastroenterology and Liver disease, Hadassah Medical Center, Jerusalem, Israel, [3]Institute of Ophthalmology, Moorfields Eye Hospital, University College London, London, UK, [4]Instiue of Gastroenterology and Hepatology, Hillel Yaffe Medical Center, Hadera, Israel, [5]Institute of Gastroenterology and Liver disease, E. Wolfson Medical Center, Holon, Israel; [6]Department of Pathology, University College London, London, UK

*Corresponding authors. Adam P. Levine, Department of Pathology, UCL Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London, WC1E 6DD, UK. Email: a.levine@ucl.ac.uk. Noam Ben-Yosef, Department of Internal Medicine, Hadassah University Hospital, Mount Scopus, Jerusalem 91999, Israel. Tel: 972-2-5844521; Email: nbenyosef@gmail.com
#Joint senior authors.

Abstract

Background: Family studies support a genetic predisposition to inflammatory bowel diseases (IBD), but known genetic variants only partially explain the disease heritability. Families with multiple affected individuals potentially harbour rare and high-impact causal variants. Long regions of homozygosity due to recent inbreeding may increase the risk of individuals bearing homozygous loss-of-function variants. This study aimed to identify rare and homozygous genetic variants contributing to IBD.
Methods: Four families with known consanguinity and multiple cases of IBD were recruited. In a family-specific analysis, we utilised homozygosity mapping complemented by whole-exome sequencing.
Results: We detected a single region of homozygosity shared by Crohn's disease cases from a family of Druze ancestry, spanning 2.6 Mb containing the *NOD2* gene. Whole-exome sequencing did not identify any potentially damaging variants within the region, suggesting that non-coding variation may be involved. In addition, affected individuals in the families harboured several rare and potentially damaging homozygous variants in genes with a role in autophagy and innate immunity including *LRRK1*, *WHAMM*, *DENND3*, and *C5*.
Conclusion: This study examined the potential contribution of rare, high-impact homozygous variants in consanguineous families with IBD. While the analysis was not designed to achieve statistical significance, our findings highlight genes or loci that warrant further research. Non-coding variants affecting *NOD2* may be of importance in Druze patients with Crohn's disease.

Key words: inflammatory bowel disease; family study; genetics; homozygosity

Introduction

Inflammatory bowel diseases (IBDs) comprise Crohn's disease (CD) and ulcerative colitis (UC). The pathogenesis of these diseases remains incompletely understood; this may reflect heterogeneity in the underlying causal pathways as a variety of factors may contribute to patients exhibiting similar phenotypes. The heritability of IBD illustrates this complexity. Early studies of families of IBD patients and twin pairs indicated a genetic risk to disease aetiology [1–5]. The discovery of damaging coding variants in the *NOD2* gene associated with CD was the first step in understanding the molecular basis of the disease and also highlighted the role of the innate immune system in CD pathogenesis [6].

Subsequent genome-wide association studies (GWASs) of large cohorts of IBD patients and controls, including those from diverse ethnicities, have identified a growing number of loci and genes associated with IBD, and highlight three notable concepts: (i) the genetic background is complex and polygenic; >200 risk loci have been detected implicating a wide variety of genes (in linkage disequilibrium (LD) with the associated polymorphisms) in CD, UC, or both [7–10]. In fact, more risk loci have been associated with IBD than with any other common complex disease [8]; (ii) most of the associated polymorphisms are common (minor allele frequency (MAF) >5%) and are hence shared by healthy controls (indicating low penetrance) and contribute only modest effect sizes with odds ratios (ORs) typically <1.3 [11]; (iii) even considering the entire spectrum of associated polymorphisms, they account for only 26% (CD) and 19% (UC) of disease variance [12]. This observation, also shared with other complex diseases, is referred to as the 'missing heritability'.

The heritability gap may be accounted for by several factors including the contribution of environmental exposures (overestimation of the heritability), epistatic effects, epigenetic modifications, or the presence of rare variants with a large effect size [13, 14]. Rare variants are difficult to detect and would demand very large cohorts [15] or the resequencing of candidate genes in population-specific cohorts [16, 17]. Alternatively, a family-designed study complemented by deep sequencing is another approach to detect rare and potentially causal variants. This strategy has been employed in several studies of multiplex families with IBD [18–21] and other comparable complex diseases [22–24].

Inbreeding or recent consanguinity leading to long stretches of homozygosity are known to contribute to Mendelian recessive disease, but may also play a role in complex polygenic diseases [25]. Examples include coronary artery disease [26], schizophrenia [27], Alzheimer's disease [28], and rheumatoid arthritis [29], although evidence is inconclusive. IBD homozygous risk variants have a higher impact than heterozygous variants. This has been shown for *NOD2* [30] and *ATG16L1* [31]. In the case of *NOD2* risk variants, a homozygous variant exceeds the risk simulated by a simple additive effect [30].

Based on the aforementioned concepts, we attempted to identify homozygous and rare variants that potentially contribute to IBD pathogenesis. As such variants are difficult to find in unrelated individuals in outbred populations, we focused on consanguineous families from populations practising endogamy. We report the genetic analysis of four consanguineous families with multiple cases of adult-onset IBD combining genome-wide single-nucleotide polymorphism (SNP) homozygosity mapping (HM) with whole-exome sequencing (WES).

Methods

Recruitment

Individuals with IBD and at least one affected first-degree relative were identified from an internal database of patients attending the IBD Unit at Hadassah Medical Center, Israel. Additional patients were referred from other specialist centres in Israel and invited to participate or volunteered following advertisement of the study by the Israel Foundation for Crohn's Disease and Ulcerative Colitis. Further patients were intentionally sought from clinics that treat large numbers of individuals from the Arab Muslim (AM) and Druze populations, as these populations have consanguinity rates that exceed 40% [32]. Interviews were initially conducted by telephone with subsequent visits to the family if there were confirmed affected first-degree relatives.

The study protocol was approved by the Institutional Review Board of Hadassah Medical Center (0557–13) and the Israeli Ministry of Health (042–2015). Consenting participant identities were coded at sample collection and the participant identity-code match was known only to the recruiting physician.

Family members were questioned regarding family background, consanguineous relations, and the presence of IBD or other immune-mediated diseases in the extended family. Diagnosis and phenotypic details were ascertained by review of medical files. The absence of disease in unaffected family members was based on specific questioning with an emphasis on chronic gastrointestinal symptoms. Saliva samples were obtained and DNA isolated according to standard procedures [33].

Genetic-analysis overview

Four families with affected offspring from consanguineous matings were identified from the recruited cohort that comprised a total of 60 multiplex families. We performed SNP microarray genotyping and subsequent HM, complemented by WES. For downstream variant prioritisation, a series of filters were applied based on population frequency, *in silico* prediction of pathogenicity, gene function and expression, and previous association with IBD.

This study focused on private and rare variants within families and was not powered or designed to achieve statistical significance. The analysis was family-specific, as the diverse background of the families did not suggest the existence of a shared risk variant.

Family pedigrees

Pedigrees were drawn with HaploPainter [34] and modified to protect the anonymity of participants without altering the distribution of individuals and sex of cases.

Genotyping and quality control

All individuals recruited from the four consanguineous families (n = 38) were genotyped using the Illumina Infinium HumanCoreExome-24 Bead Chip (Illumina, San Diego, CA) and processed with the Illumina GenomeStudio software. Quality control was undertaken using PLINK: samples were removed if missing >5% of SNPs and then SNPs were removed if they were missing in >5% of samples or deviated significantly from Hardy–Weinberg Equilibrium (Fisher's exact test *P*-value $<5 \times 10^{-8}$) [35].

The sex of each sample was verified by the homozygosity rate of common variants (MAF >0.05) on the X chromosome (excluding the pseudoautosomal regions). Pedigree structure and relatedness between samples were verified by kinship coefficient calculations using KING [36].

Homozygosity mapping

Regions of homozygosity (ROH) were detected using HomozygosityMapper [37]. Common (MAF >0.05 for the combined gnomAD (https://gnomad.broadinstitute.org/) population) SNPs were used as input. Default software parameters were employed enabling the 'require genetic homogeneity' feature, thus limiting the detection of ROH to those shared by all cases and of identical genotype. Analyses were performed with and without the exclusion of long ROH in unaffected family controls, allowing for the possibility of incomplete penetrance.

A minimum threshold for ROH length of 1 Mb was defined; this represents ROH caused by recent consanguinity as implied by the pedigree (expected to contain several mega bases), as well as homozygous segments resulting from a more distant history of inbreeding, as suggested by the reported history of the families [38]. A minimum threshold of 50 continuous identical SNPs was required to exclude regions identical by state rather than identical by descent and large regions with low SNP density (Supplementary Methods).

Protein-coding genes within the ROH coordinates (human genome build GRCh37) were identified using the Ensembl BioMart tool [39]. The presence of IBD GWAS loci within ROH was examined using the top SNP reported in the loci [11].

HM analysis was complemented with manual inspection of all rare variants identified through WES (described below) that were located within each ROH. The WES data were also used to confirm the identification of ROH, with the absence of heterozygous variants within these regions.

Whole-exome sequencing

Exome capture and library preparation were performed using the BGI 59 Exome Enrichment Kit (BGI, China). One sample failed library preparation quality control. Paired-end sequence reads (150 bp) were generated using the Illumina HiSeq 2000 system (Illumina, San Diego, USA).

Read alignment and variant calling

Sequence reads were aligned to the human reference genome hg19 using Novoalign (version 3.02.08). Samples were processed jointly with 5,285 samples comprising the UCL Exome Sequence Consortium, a local collection of exomes from a variety of cohorts. Variants were called using the Genome Analysis Tool Kit (GATK) [40] according to best practices [41] with local realignment around InDels, followed by joint variant calling and variant quality score recalibration (VQSR).

Several quality-control filters were applied to the variants. First, we required a VQSR truth tranche ≤99.5% for SNPs and ≤99.0% for InDels. Next, we applied filters to individual genotypes and changed those failing to 'no-calls': Genotype Quality should be ≥30, for heterozygous calls the read depth of the alternative allele should not diverge significantly from 50:50 (chi-squared test $P < 0.001$), and for heterozygous or alternate homozygous calls the read depth of the alternate allele should be ≥3. Finally, we only retained variants with a no-call rate <0.25.

Variant annotation, filtering, and prioritization

Variants were interpreted and annotated using the Ensembl Variant Effect Predictor tool (VEP) [42] for the predicted impact on Ensembl gene transcripts and population frequencies in gnomAD [43]. Missense variants were annotated with their predicted effect on protein function *in silico* using PolyPhen-2 [44], SIFT [45] (with VEP), and CADD score [46]. The Druze and Arab Middle-Eastern populations are not represented in gnomAD; variant frequencies from the Greater Middle East Variome Project (GMEV) [47] were thus included. The GMEV consists of WES-derived variants from >1,000 unrelated subjects of Middle-Eastern descent, enabling variants common amongst such populations to be identified and excluded [47] (Supplementary Methods).

Variants were defined as 'damaging' if they were either a frameshift, start/stop altering, splice acceptor/donor variants, or missense variants predicted to be deleterious by either PolyPhen-2, SIFT, or CADD (CADD score >20). Three classes of population allele frequency were defined: 'uncommon' as 1% < MAF < 5%; 'rare' as 0.1% < MAF < 1%; and 'very rare' as MAF <0.1%. For the Druze and AM families, these thresholds had to be satisfied in both the gnomAD combined population and the GMEV. Similarly, for the Ashkenazi Jewish (AJ) family, these thresholds had to be satisfied in both in the gnomAD combined and AJ-specific populations.

To maximize the sensitivity to detect homozygous variants of potential interest, we utilised a non-stringent threshold for predicted deleteriousness (retaining variants predicted to be deleterious by any predictive model) or population frequency (retaining uncommon variants). We note that, under Hardy–Weinberg Equilibrium, even uncommon variants should be exceptionally scarce in the homozygous state.

Following annotation and filtering, homozygous 'damaging' and 'uncommon' (or of lower frequency) variants present in the proband of each family were identified. Each of these variants was further examined separately considering the segregation within cases in the family and absence in unaffected family controls. Furthermore, the encoded protein's function (in UniProt [48]), expression profile (as per BioGPS [49]), and prior association with IBD (described below) were examined. Finally, we queried the IBD Exomes Browser [50] for each specific variant to determine whether the gene is enriched for variants affecting IBD risk.

IBD-associated genes

To focus our search for candidate variants, a set of genes with known IBD association was defined. This included 251 genes reported by Mirkov *et al.* [11] from loci associated with IBD by GWAS, 67 genes reported in cases of monogenic phenotypes of very-early-onset IBD [51], and additional genes from a literature search for association studies or direct-sequencing studies. After removing duplicates, a list of 366 IBD-associated genes was generated (Supplementary Table 1A). Our aim was to achieve a permissive and comprehensive list, which we refer to as the 'extensive' IBD-association list.

A second list of 20 'top genes' that have been confidently implicated in IBD risk by fine-mapping or with functional data [15] was also generated (Supplementary Table 1B). The enrichment of variants within genes from the 'extensive' gene list and the 'top genes' list was examined. This strategy was separately employed for all variants and for the homozygous variants. To avoid bias due to pedigree structure, the number of homozygous variants was only compared between siblings, as they are
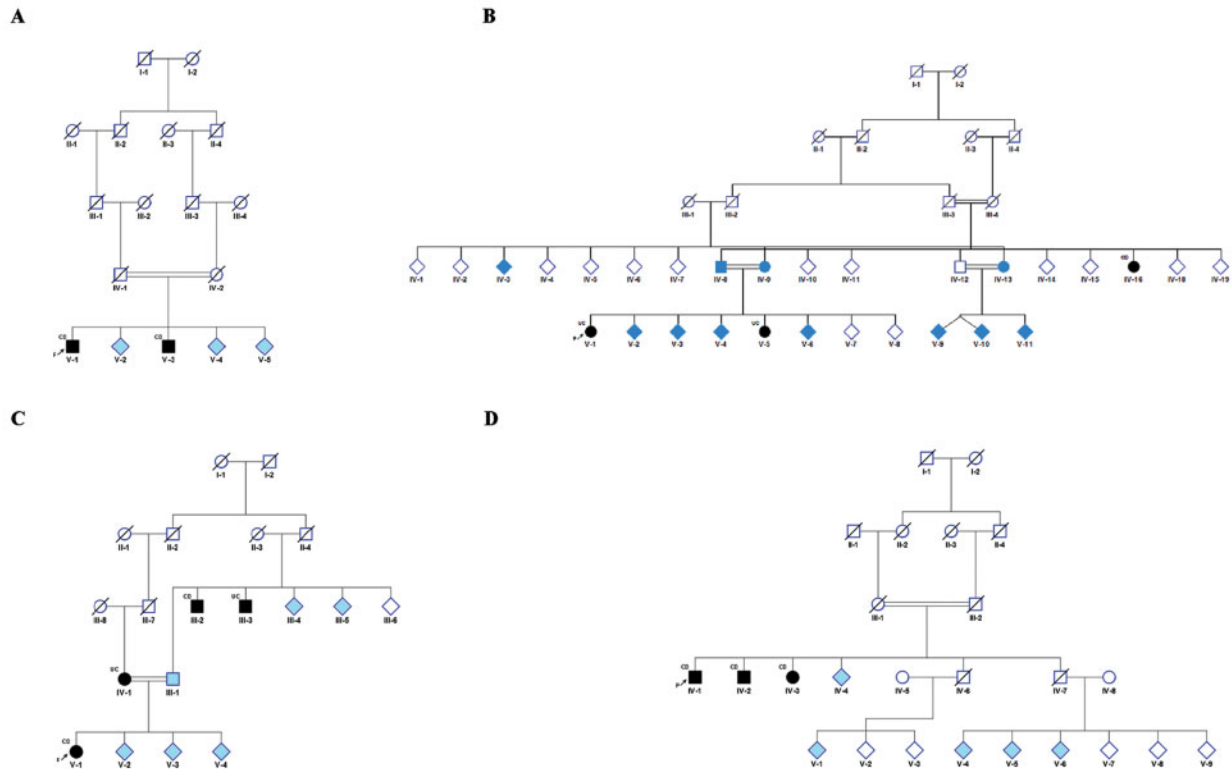
Figure 1. Pedigrees of the four consanguineous families with inflammatory bowel disease (IBD). (A) Arab Muslim family with Crohn's disease (CD) as the major phenotype. (B) Arab Muslim family with ulcerative colitis (UC) as the major phenotype. (C) Druze family. (D) Ashkenazi Jewish family. IBD cases are coloured in black. Unaffected family controls who were also exome sequenced are coloured in light blue.

expected to share an equal fraction of their genome in the homozygous state. Finally, cases and family controls were screened for a set of coding functional variants directly implicated in IBD pathogenesis [52–54] (Supplementary Table 1C).

Results

Description of families

A total of 38 individuals from four consanguineous families participated in the study. This comprised 8 CD cases, 4 UC cases, and 26 unaffected family members. Pedigrees for the families are provided in Figure 1. The case with the most severe phenotype at the time of study recruitment was designated as the proband in each family. Two families of AM ancestry expressed either CD or UC as the major phenotype. The other two families were of Druze and AJ descent. For purposes of clarity and ease of reading, the families are designated as AM-CD, AM-UC, DR, and AJ, respectively.

Families AM-CD and AM-UC are each members of a different tribe (*chamulla*) tracing their ancestry back 200–300 years. Today, each tribe numbers a few thousand members and marriage is generally, although not exclusively, within the tribe. DR have a unique social and demographic structure, with customs that strongly favour marriage within the village, and mating outside the religion is forbidden [55]. AJ are a genetically distinct population, estimated to have arisen from ~350 individuals 25–32 generations ago [56].

In family AJ (Figure 1D), two siblings died from ischemic heart disease diagnosed when >60 years of age and had no prior history of gastrointestinal diagnoses or symptoms. Four descendants of these individuals were available for sequencing;

if any descendant was homozygous for the reference allele, then homozygosity of the parent for the candidate variant (alternate allele) was inferred to not be possible.

The average age of disease onset amongst the cohort was 42.9 years for CD and 26.5 years for UC, but varied substantially between families, with the AJ family having a relatively later age of onset (average 56.3 years). Additional clinical information is provided in Table 1.

Data generation

Genome-wide SNP genotyping
Members of the four consanguineous families were genotyped with a SNP microarray ($n = 38$). Two individuals failed to pass genotyping quality-control criteria (missing genotypes >5%); both were cases from family AM-CD and therefore this family was not included in the HM analysis. Following quality-control filtering and removal of SNPs with duplicate IDs, 251,716 markers across the autosomes were available for HM, covering 2,790 Mb of the autosome (Supplementary Methods).

Whole-exome sequencing
WES analysis was performed on 37 individuals from the four families following successful library preparation. Following quality control, 95.4% of the exome was covered at 5× and 87.6% at 20×, with an average read depth of 89.1×. A total of 32,272 variants were common to the WES and SNP-microarray genotyping for which the call concordance rate was >0.99. The average number of exome-wide (coding and splice site only) sequence variants in the autosomes per individual was 19,550.

Table 1. Clinical information for cases with inflammatory bowel disease

Family	Patient ID	Relation to proband	Gender	Diagnosis	Age at diagnosis	Location	Behaviour (CD)
AJ	IV-1	Proband	Male	CD	59	Ileal	Stricturing and penetrating
	IV-2	Sibling	Male	CD	51	Ileal	Stricturing
	IV-3	Sibling	Female	CD	59	Ileal	Non-stricturing, non-penetrating
DR	V-1	Proband	Female	CD	22	Ileocolonic with upper gastrointestinal involvement	Non-stricturing, non-penetrating
	III-2	Uncle	Male	CD	45	Ileocolonic	Non-stricturing, non-penetrating
	IV-1	Mother	Female	UC	21	Left-sided/proctitis	–
	III-3	Uncle	Male	UC	42	Left-sided/proctitis	–
AM-CD	V-1	Proband	Male	CD	32	Ileocolonic	Stricturing
	V-3	Sibling	Male	CD	33	Ileal	Non-stricturing, non-penetrating
AM-UC	V-1	Proband	Female	UC	28	Left-sided	–
	V-5	Sibling	Female	UC	15	Proctitis	–
	IV-16	Aunt	Female	CD	42	Ileal	Non-stricturing, non-penetrating

AJ, Ashkenazi Jewish family; DR, Druze family; AM, Arab Muslim family; CD, Crohn's disease; UC, ulcerative colitis.

Patient IDs refer to the pedigrees in Figure 1.

Homozygosity mapping

ROH shared by affected family members with concordant IBD subtypes, following manual inspection of the region and confirmation of identity by descent, are presented in **Table 2**. No ROH were shared among the UC cases from family DR.

Together, these ROH contain 703 protein-coding genes (∼3.6% of all protein-coding genes covered by HomozygosityMapper; a complete list of these genes is provided in Supplementary Table 2). They are significantly enriched in genes from the top 20 IBD-associated genes ('top genes' list [15], 3/20 genes; *CARD9*, *SMAD3*, and *NOD2*; OR 4.7, $P = 0.03$), but not with broader sets of genes previously associated with IBD (10/366 IBD-associated genes from our 'extensive' list and 4/242 GWAS loci [11] are contained in these ROH). Only four rare variants predicted to be damaging were located within these ROH, all in family AM-UC (**Table 2**).

The two affected siblings from family AM-UC shared four ROH spanning a total of 55.8 Mb (2% of the covered genome). This was 10-fold longer than the total shared ROH in other families—an expected finding given the pedigree structure in family DR and the requirement for overlapping ROH for three cases in family AJ. Combining HM in family AM-UC with the affected aunt (diagnosed with CD, Figure 1B) did not identify any ROH passing our threshold of 1 Mb (the longest shared region is 0.87 Mb long containing 51 identical homozygous SNPs and all other stretches are shorter than 0.5 Mb). Similarly, in family DR, combining HM for all four affected members with any IBD subtype identified only one ROH longer than 1 Mb located on chromosome 11 near the centromere, which was also shared with four unaffected members of the family. A manual inspection of this region reveals that it is not a true ROH; it encompassed only 49 continuous SNPs and includes 2 heterozygous SNP calls that were incorrectly considered by the software to be erroneous genotypes as they were flanked on both sides by a series of identical homozygous SNPs. This illustrates the complexity of SNP-microarray-based homozygosity-mapping analysis in highly inbred families, favouring a stringent threshold of continuous identical SNPs to define true autozygosity, especially in genomic regions with low SNP-marker density.

The two CD cases from family DR (V-1 and III-2) shared a single ROH located on chromosome 16 (Figure 2A) spanning 2.6 Mb containing the *NOD2* gene. Overlapping ROH were not observed

for any other member of the family (Figure 2B). This ROH also contains *BRD7* and *HEATR3*, both associated with CD (either independently or due to their linkage disequeilibrium (LD) with *NOD2*) [57]. Interestingly, the three well-established CD associated *NOD2* risk variants (fs1007insC, R702W, and G908R) were homozygous wild type in these CD cases. In fact, the entire region contained only four intronic variants and four coding synonymous variants, all common (MAF >0.05). Two of these are synonymous variants in *NOD2* (S178 and R587, in LD) that are significantly associated with CD *protection* (OR 0.77, $P < 1 \times 10^{-16}$ in the IBD Exome Browser). The robustness of this finding was verified by repeating the analysis with variable parameters, implicating only one additional ROH on chromosome 2, which is also shared by five other members of the family and contains no protein-coding genes.

Individuals in family DR were found to have a greater fraction of their genome within ROH than expected based on the pedigree alone, suggestive of historical inbreeding. Indeed, long ROH were also detected in individual III-2 and his siblings (generation III in Figure 1C, Supplementary Methods). Therefore, to evaluate the probability that a specific region on chromosome 16 containing *NOD2* would be identically homozygous between the CD cases in this family, a hypothetical pedigree was constructed in which individual III-2 is the offspring of a first-cousin mating. Meiotic events including recombination were simulated by gene dropping for the entire autosome in the pedigree 10,000 times (Supplementary Methods). The probability that both affected individuals of relevance were identically homozygous for a specific 1 cM region on chromosome 16 (at the approximate position of *NOD2*) was estimated as $P = 0.0021$.

Exome-wide variant analysis

Focusing on the homozygous variants, each individual had an average of 6,749 for any (regardless of allele frequency), 39.7 for uncommon (MAF <0.05), 11.8 for rare (MAF <0.01), and 3.8 for rare and damaging variants. Examining only offspring of observed consanguineous matings, the corresponding averages were higher, at 60.0, 19.6, and 7.0, respectively. An overview of the homozygous variants observed in these families is provided in Supplementary Table 3.

No CD case carried any of the known *NOD2* risk variants. All three CD cases in family AJ were homozygous for the common

Table 2. Regions of homozygosity (ROH) detected by homozygosity mapping

Family	Chr	Start[a]	End[a]	Length (Mb)	Number of SNPs	Overlapping ROH in family controls	Presence of IBD GWAS locus[b]	Known IBD-associated genes located in region	Number of protein-coding genes[c]	Number of uncommon variants[d]	Rare variants predicted to be damaging[d]
AM-UC	1	37,013,031	46,037,357	9.0	772	V-3	No IBD GWAS loci	None	123	7	0
	8	137,523,046	142,950,775	5.4	612	V-3	No IBD GWAS loci	None	14	2	DENND3
	9	124,704,304	141,022,295	16.3	1,539	V-2, V-9, V-10	rs10781499, OR 1.19 (IBD), P-value 4.38E-56, RAF 0.5	CARD9, SDCCAG3, PMPCA	281	18	CERCAM, WDR38
	15	54,343,188	79,365,780	25.0	2,154	None	rs17293632, OR 1.11 (IBD), P-value 2.71E-20, RAF 0.14	SMAD3	220	16	AAGAB
DR: CD cases	15	51,663,597	52,878,719	1.2	57	IV-9, IV-13, V-2	No IBD GWAS loci	None	14	3	0
	2	220,753,783	222,044,974	1.3	141	III-1, IV-1, V-2, V-3, V-4	No IBD GWAS loci	None	0	0	0
AJ	16	48,846,170	51,468,005	2.6	359	None	rs2066844, OR 2.003 (IBD), P-value 2.27E-217, RAF 0.06	HEATR3, ADCY7, BRD7, NOD2	13	0	0
	2	183,467,456	190,452,586	7.0	451	IV-4	rs144344067, OR 1.12 (IBD), P-value 1.29E-08, RAF 0.89	ITGAV, FAM171B	21	5	0
	4	53,383,129	54,396,077	1.0	55	None	No IBD GWAS loci	None	6	0	0
	12	22,890,730	26,139,522	3.2	361	None	No IBD GWAS loci	None	11	0	0

AM, Arab Muslim family; DR, Druze family; AJ, Ashkenazi Jewish family; CD, Crohn's disease; UC, ulcerative colitis; Chr, chromosome; SNP, single-nucleotide polymorphism; IBD, inflammatory bowel disease; GWAS, genome-wide association study; OR, odds ratio; RAF, risk-allele frequency. Minimum threshold for ROH identification is length >1 Mb and ≥50 SNPs.

ROH shared between cases in each family.

[a]ROH positions refer to GRCh37.

[b]Evaluated by the position of the SNP with the smallest P-value in an IBD GWAS locus. SNP rsID, OR, P-value for association, and RAF in the European population are provided, adapted from Ref. [11] and originally reported in Refs [8–10, 52].

[c]Protein-coding genes within the region were identified using the BioMart tool available from Ensembl. A complete list of genes in each ROH is available in Supplementary Table 2.

[d]Minor-allele frequency <0.05 for uncommon variants and <0.01 for rare variants.
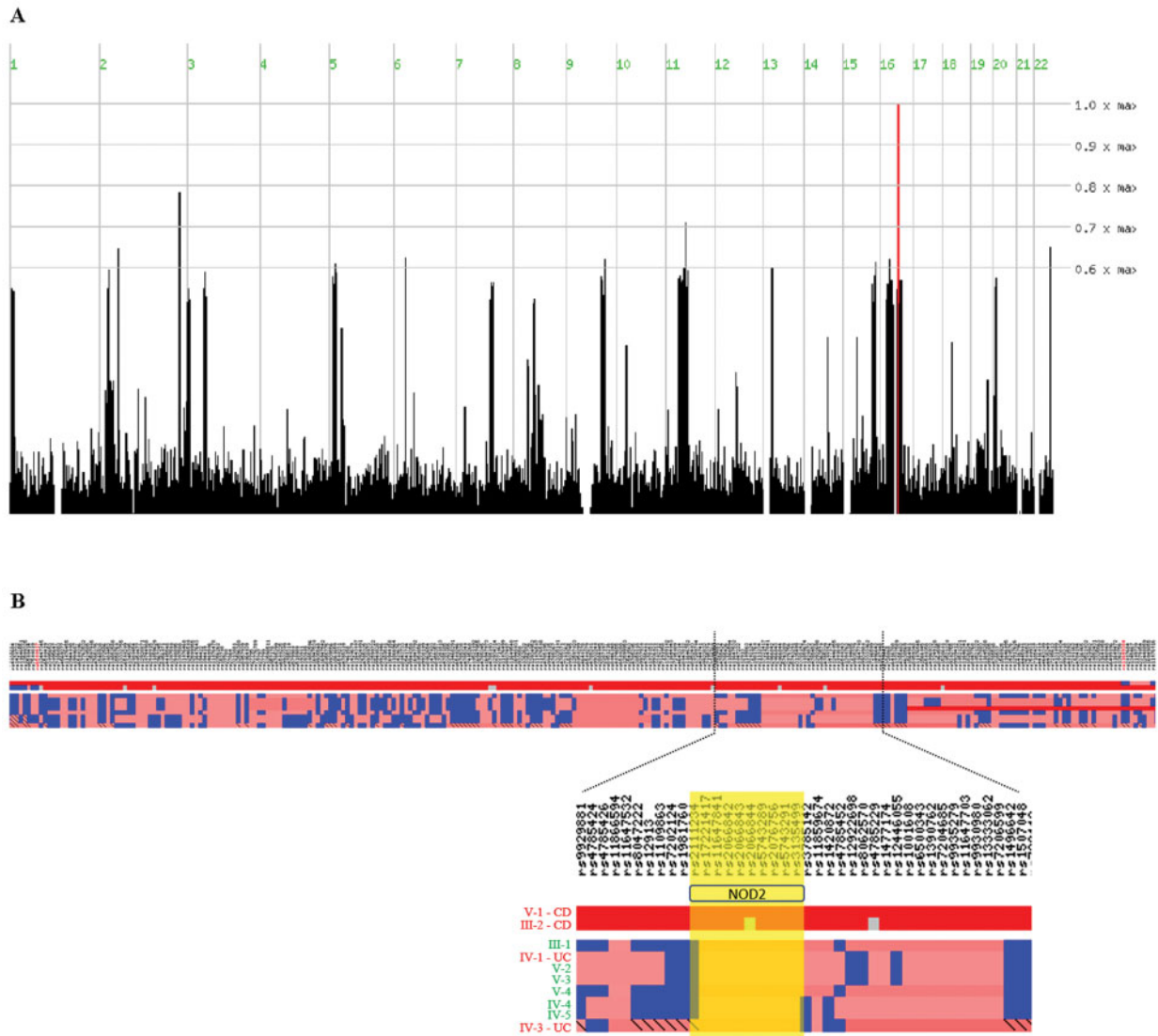
Figure 2. Region of homozygosity on chromosome 16 in the Druze family. (A) Whole-genome view of the region of homozygosity (ROH) shared by Crohn's disease cases (V-1 and III-2) in the Druze family. The height of the bars corresponds to the number of continuous homozygous markers shared by both cases and the width corresponds to the length of the ROH. The region with the highest score (red bar) is located on chromosome 16 and contains 359 single-nucleotide polymorphism (SNP) markers, spanning 2.6 Mb and surrounding *NOD2*. (B) Genotype view for region of homozygosity on chromosome 16. Each column represents a SNP marker and each row an individual in the family; the two CD cases are at the top. The blue colour represents a heterozygous call for the marker and red represents the homozygous calls (darker shades of red reflect the number of continuous homozygous calls; the grey colour indicates a 'no-call' for that marker). The position of *NOD2* is indicated by the yellow box. Examination of the exome sequence confirms that the two CD cases are indeed homozygous across and surrounding *NOD2*, but other members of the family are heterozygous.

T300A risk allele (OR 1.2[8]) in *ATG16L1*, as were five unaffected controls in this family. No other case was homozygous for any of the functional variants directly implicated in IBD pathogenesis [52–54]. The distribution of potentially damaging homozygous variants in IBD-associated genes is outlined in Supplementary Table 4.

The probands in each family had 9–18 (average 13.5) homozygous, uncommon, and damaging variants. A description of these 54 variants is provided in Supplementary Table 5. For further examination, variants were prioritised if they were homozygous in both affected siblings in family AM-CD or family AM-UC or at least two of three affected siblings in family AJ. In family DR, the complex pedigree structure and case relatedness did not favour a prioritisation strategy so all variants were initially examined, unless they were also common in unaffected family

controls. We further excluded variants in unknown genes or genes that have a specific function or expression profile that does not suggest involvement in IBD pathogenesis, leaving 19 variants. The prioritized variants are described in Table 3 and with additional detail in Supplementary Table 6.

In family AM-CD, all four prioritised variants were present in a 12-Mb region on chromosome 10. A manual inspection of all variants in this region confirmed that it is part of a long ROH shared by the two affected brothers, but not their unaffected siblings. In families DR, AM-CD, and AM-UC, incorporating the GMEV as a reference for allele frequency altered the MAF class in 4/23 variants (3 changed from 'rare' to 'uncommon', 1 from 'very rare to rare').

As expected, all variants identified on the basis of their presence within a ROH identified by HM were independently

Table 3. Prioritized homozygous variants

Family	Chr	rsID	Gene	Consequence	Amino-acid change	SIFT	PolyPhen	CADD score	gnomAD MAF	Population-specific MAF[a]	Homozygous in controls[b]
AM-CD	10	rs56226109	FGFR2	Missense	S57L	Tolerated	Benign	22.4	0.0049	0.0065	0/3
	10	rs143892520	DMBT1	Missense	D560E	Deleterious	Unknown	12.66	0.015	–	0/3
AM-UC	8	–	DENND3	Missense	R487S	Deleterious	Possibly damaging	18.83	–	–	1/10
	9	rs5794	PTGS1	Missense	V481I	Deleterious	Benign	23.3	0.0072	0.0146	3/10
	9	–	WDR38	Missense	R51W	Deleterious	Possibly damaging	28.4	–	–	3/10
	9	rs146651928	CERCAM	Missense	Q187E	Deleterious	Probably damaging	22.9	0.0006	0.0005	3/10
	15	–	AAGAB	Missense	D312N	Tolerated	Possibly damaging	25.8	–	–	0/10
DR	2	rs148167737	MAP4K3	Missense	A410T	Tolerated	Benign	23.9	0.0008	–	1/6
	2	rs377671536	MAP4K4	Missense	Q515E	Tolerated (LC)	Possibly damaging	23.3	0.0002	0.0005	0/6
	9	rs3739740	NIPSNAP3B	Missense	K154E	Deleterious	Possibly damaging	27.7	0.0267	0.0347	1/6
	15	rs185529473	WHAMM	Missense	R725W	Deleterious	Possibly damaging	29.5	4.47E-05	–	0/6
	15	rs61752778	ADAMTSL3	Missense	P821S	Tolerated	Probably damaging	24.4	0.0163	0.0091	0/6
	15	rs61731243	AKAP13	Missense	E1106G	Deleterious (LC)	Benign	16.55	0.0415	0.0267	0/6
	15	rs55798315	LRRK1	Missense	P543S	Deleterious	Probably damaging	29.8	0.0044	0.0171	1/6
AJ	4	rs112033303	COQ2	Stop gained	R22X	–	–	23	0.0166	0.0304	0–1/3
	2	rs138440701	IL36B	Missense	I110T	Deleterious (LC)	Benign	3.083	0.013	0.0222	0–1/3
	9	rs34552775	C5	Missense	L354M	Deleterious	Possibly damaging	25	0.0054	0.0044	0–1/3
	19	rs75841596	PALM3	Missense	D604N	Deleterious	Possibly damaging	26.8	0.0245	0.0319	0/3
	19	rs75251420	MYO9B	Missense	V1700M	Deleterious	Benign	23.9	0.0014	0.0024	0/3

AM, Arab Muslim family; DR, Druze family; AJ, Ashkenazi Jewish family; CD, Crohn's disease; UC, ulcerative colitis; Chr, chromosome; LC, low confidence; gnomAD, Genome Aggregation Database; MAF, minor allele frequency (missing MAF: non-existent in data set).

Family-specific variants identified by whole-exome sequencing. Variants were prioritised if they satisfied all of the following: homozygous, predicted to be damaging, allele frequency <0.05, segregated in the affected cases, and could not be excluded based on an unrelated specific function or gene expression.

[a]Using MAF in the Greater Middle East Variome (for families AM-CD, AM-UC, and DR) or the gnomAD Ashkenazi Jewish population (for family AJ).

[b]In the AJ family, estimation for homozygosity for the deceased unaffected siblings was based on whole-exome sequence data from their descendants.

detected by WES filtering and prioritisation. In contrast, three homozygous variants derived from the WES and shared by all cases (one in family AJ and two in family AM-UC) were not identified with HM. One was located on the X chromosome and would not be detected, as only the autosome was analysed by HM. The other two were located in very short ROHs, far below our threshold for identification. In family AJ, 9/10 variants shared by two affected siblings would have been detected, with HM targeting two cases separately instead of all three together.

Discussion

Research efforts aiming to elucidate the genetic susceptibility to IBD have focused on common variants with modest effect size derived from large-scale GWASs, alongside attempts to unravel rare and highly penetrant variants in population-specific cohorts or family-focused studies. Familial cases have the potential to harbour rare causal variants that would be difficult to identify in cohorts of unrelated outbred cases and controls [13]. High-impact variants are likely to induce substantially increased risk when both alleles are pathogenic. In CD, this is exemplified by the *NOD2* frameshift variant, for which a homozygous variant is associated with greater disease risk [30], earlier onset [58], and more severe phenotype [59].

In order to target potentially causal and rare homozygous variants, we recruited multiplex families from selected homogenous populations. Despite focusing on populations with high consanguinity rates [32] in the AM and Druze, we only identified three multiplex consanguineous families with IBD. This likely reflects the lower background disease prevalence in the AM and Druze in Israel [60, 61].

Our analysis strategy combined HM and WES. The most intriguing finding derived from HM is a 2.6-Mb ROH shared by both CD cases in family DR, but not any other member of the family. This region harbours 13 protein-coding genes, including *NOD2*. Interestingly, the entire region included only four coding variants, all of which are common synonymous variants, ultimately leading to our impression that this ROH may contain a non-coding sequence variant that alters *NOD2* function or expression. The presence of known *NOD2* risk variants is significantly lower in the non-Jewish than in the Jewish CD cases in Israel, raising the possibility that novel *NOD2* variants determine disease susceptibility in the non-Jewish population [62]. We suggest that further study of this locus, including examination for non-coding variants, particularly amongst the Druze population, may be warranted. Apart from the abovementioned ROH, we did not detect an enrichment of IBD-associated genes within ROH shared by cases using our extensive list of 366 genes.

WES identified a total of 54 uncommon and damaging variants that were homozygous in the probands. Only four of these variants are located in genes previously implicated to have a role in IBD (*DMBT1*, *LRRK1*, *MUC1*, and *MYO9B*). Additional prioritisation steps narrowed the list to 19 variants (2–7 per family) that potentially contribute to disease pathogenesis in a family-specific manner. While this study was not designed to achieve statistical significance or experimental validation for their role in IBD aetiology, we suggest that several variants residing in genes with a role in autophagy or innate immunity warrant additional consideration in future studies:

i. *LRRK1* is essential for autophagosome and lysosome fusion through its regulation of the activation-deactivation cycle of Rab7 [63]. A locus containing the *LRRK1* gene has been associated with CD [64]. Missense variants in *LRRK2*, a close paralogue, have been causally implicated in CD [54].

ii. *WHAMM* is an actin nucleation-promoting factor with a key role in autophagosome formation. *WHAMM* directs the activity of the Arp2/3 complex for autophagosome biogenesis through an actin-comet tail motility mechanism, and *WHAMM* expression correlates with the number and size of autophagosomes [65, 66].

iii. *DENND3* activates Rab12, which facilitates the trafficking of autophagosomes towards the lysosomes [67]. The *p. Arg487Ser* variant does not exist in any population in gnomAD or GMEV, and is located within the 'linker region' required to link *DENND3* to Rab12.

iv. *C5* is a terminal complement factor common to all complement pathways. C5 is cleaved to C5a, a potent chemoattractant and anaphylatoxin, and C5b, which initiates the formation of and contributes to the membrane attack complex [68]. The candidate variant *p. Leu354Met* is located within C5b. Rare variants in the C5b chain are associated with recurrent bacterial infections [68]. Earlier studies suggest association of CD with impaired complement function [69]. Reduced immunity and clearance of bacterial intrusions may lead to the chronic bowel inflammation in CD [70, 71].

Two unresolved questions regarding the genetic basis of familial IBD and other complex polygenic diseases with late onset underlie the intention and design of this study. (i) Is the increased familial risk a result of private high-impact variants or the accumulation of many common low-effect polymorphisms? (selected arguments reviewed in Ref. [72]). (ii) Does increased homozygosity across the genome contribute to disease risk? Studies of complex polygenic diseases have provided both positive [26, 27, 73, 74] and negative [75, 76] results. Even if homozygosity increases disease risk, it may not necessarily reflect the presence of rare loss-of-function variants residing on both alleles, which was the focus of our study, but could derive from the combined effects of many common risk alleles. To our knowledge, only one study has attempted to determine the contribution of consanguinity to IBD risk and it did not reach definitive conclusions; however, no genetic data were available and pedigree information was incomplete [77].

One methodological issue arising from this study regards the added benefit of combining HM with WES, as all homozygous variants should be detected with WES alone. HM can implicate regions of interest if they harbour genes associated with the studied phenotype and allow careful reconsideration of variants that were filtered by the WES analysis pipeline (e.g. if they were not predicted to be damaging according to the study definition). In addition, WES filtering for uncommon variants would not have detected our most interesting finding from HM—the ROH on chromosome 16 in family DR.

We note two important limitations of this study. First, we did not replicate the identified variants in an unrelated case–control cohort. However, this would be challenging given the rarity of the variants prioritised. Additionally, validation would require using similar populations, and we are unaware of the presence of IBD case–control cohorts with exome data from the Druze and AM populations. Second, exome-focused analysis covers only a small fraction of the genome. We have chosen this strategy not only because of reduced costs and feasibility for systematic bioinformatic annotation, but mainly because high-impact penetrant variants are more likely to reside within the protein-coding sequence [52].

In summary, we have performed a comprehensive genetic analysis of four multiplex and consanguineous families in an attempt to identify rare, high-impact, homozygous IBD-associated variants. We have identified a ROH spanning 2.6 Mb surrounding the *NOD2* gene and shared by two CD cases, lacking potentially damaging coding variants. This region should be explored in the future in studies targeting the unique Druze population. We have suggested four genes, namely *LRRK1*, *WHAMM*, *DENND3*, and *C5*, which are associated with autophagy and/or innate immunity and contain rare and damaging variants that may be aetiologically involved with IBD in these families and also warrant further research.

Supplementary Data

Supplementary data is available at *Gastroenterology Report* online.

Authors' contributions

N.B.Y.: study design, participant recruitment, data collection and interpretation, data analysis, drafting of the manuscript; M.F. and N.P.: data analysis; A.P.L. and E.R.S.: participant recruitment, data analysis; S.D., F.A.B, and F.S.: participant recruitment; E.I.: participant recruitment, interpretation of data; A.W.S.: study design and supervision. All authors critically revised the manuscript and approved the final version to be submitted.

Funding

Acknowledgements

The authors kindly acknowledge all study participants and clinical collaborators.

Conflicts of interest

None declared.

References

1. Orholm M, Munkholm P, Langholz E *et al.* Familial occurrence of inflammatory bowel disease. *N Engl J Med* 1991;**324**:84–8.
2. Yang H, McElree C, Roth MP *et al.* Familial empirical risks for inflammatory bowel disease: differences between Jews and non-Jews. *Gut* 1993;**34**:517–24.
3. Peeters M, Nevens H, Baert F *et al.* Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. *Gastroenterology* 1996;**111**:597–603.
4. Orholm M, Binder V, Sørensen TI *et al.* Concordance of inflammatory bowel disease among Danish twins: results of a nationwide study. *Scand J Gastroenterol* 2000;**35**:1075–81.
5. Schiff ER, Frampton M, Semplici F *et al.* A new look at familial risk of inflammatory bowel disease in the Ashkenazi Jewish population. *Dig Dis Sci* 2018;**63**:3049–57.
6. Ogura Y, Bonen DK, Inohara N *et al.* A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease. *Nature* 2001;**411**:603–6.
7. Barrett JC, Hansoul S, Nicolae DL; the NIDDK IBD Genetics Consortium *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008;**40**:955–62.
8. Jostins L, Ripke S, Weersma RK; The International IBD Genetics Consortium (IIBDGC) *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;**491**:119–24.
9. Liu JZ, van Sommeren S, Huang H; International Multiple Sclerosis Genetics Consortium *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015;**47**:979–86.
10. de Lange KM, Moutsianas L, Lee JC *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* 2017;**49**:256–61.
11. Mirkov MU, Verstockt B, Cleynen I. Genetics of inflammatory bowel disease: beyond NOD2. *Lancet Gastroenterol Hepatol* 2017;**2**:224–34.
12. Chen GB, Lee SH, Brion MJ; the International IBD Genetics Consortium *et al.* Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. *Hum Mol Genet* 2014;**23**:4710–20.
13. Manolio TA, Collins FS, Cox NJ *et al.* Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747–53.
14. Verstockt B, Smith KG, Lee JC. Genome-wide association studies in Crohn's disease: past, present and future. *Clin Transl Immunol* 2018;**7**:e1001.
15. Luo Y, De Lange KM, Jostins L *et al.* Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nat Genet* 2017;**49**:186–92.
16. Beaudoin M, Goyette P, Boucher G, Quebec IBD Genetics Consortium *et al.* Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. *PLoS Genet* 2013;**9**:e1003723.
17. Rivas MA, Beaudoin M, Gardet A; National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC) *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011;**43**:1066–73.
18. Stittrich AB, Ashworth J, Shi M *et al.* Genomic architecture of inflammatory bowel disease in five families with multiple affected individuals. *Hum Genome Var* 2016;**3**:15060.
19. Levine AP, Pontikos N, Schiff ER *et al.* Genetic complexity of Crohn's disease in two large Ashkenazi Jewish families. *Gastroenterology* 2016;**151**:698–709.
20. Sasaki MM, Skol AD, Hungate EA *et al.* Whole-exome sequence analysis implicates rare Il17REL variants in familial and sporadic inflammatory bowel disease. *Inflamm Bowel Dis* 2016;**22**:20–7.
21. Schiff ER, Frampton M, Ben-Yosef N *et al.* Rare coding variant analysis in a large cohort of Ashkenazi Jewish families with inflammatory bowel disease. *Hum Genet* 2018;**137**:723–34.
22. Timms AE, Dorschner MO, Wechsler J *et al.* Support for the N-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families. *JAMA Psychiatry* 2013;**70**:582–90.
23. Farlow JL, Lin H, Sauerbeck L; FIA Study Investigators *et al.* Lessons learned from whole exome sequencing in multiplex families affected by a complex genetic disorder, intracranial aneurysm. *PLoS One* 2015;**10**:e0121104.

24. Kohli MA, Cukier HN, Hamilton-Nelson KL *et al*. Segregation of a rare TTC3 variant in an extended family with late-onset Alzheimer disease. *Neurol Genet* 2016;**2**:e41.

25. Erzurumluoglu AM, Shihab HA, Rodriguez S *et al*. Importance of genetic studies in consanguineous populations for the characterization of novel human gene functions. *Ann Hum Genet* 2016;**80**:187–96.

26. Christofidou P, Nelson CP, Nikpay M *et al*. Runs of homozygosity: association with coronary artery disease and gene expression in monocytes and macrophages. *Am J Hum Genet* 2015;**97**:228–37.

27. Keller MC, Simonson MA, Ripke S; The Schizophrenia Psychiatric Genome-Wide Association Study Consortium *et al*. Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *PLoS Genet* 2012;**8**:e1002656–11.

28. Nalls MA, Guerreiro RJ, Simon-Sanchez J *et al*. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics* 2009;**10**:183–90.

29. Yang HC, Chang LC, Liang YJ *et al*. A genome-wide homozygosity association study identifies runs of homozygosity associated with rheumatoid arthritis in the human major histocompatibility complex. *PLoS One* 2012;**7**:e34840.

30. Mathew CG, Lewis CM. Genetics of inflammatory bowel disease: progress and prospects. *Hum Mol Genet* 2004;**13**:R161–8.

31. Okazaki T, Wang MH, Rawsthorne P *et al*. Contributions of IBD5, IL23R, ATG16L1, and NOD2 to Crohn's disease risk in a population-based case-control study: evidence of gene-gene interactions. *Inflamm Bowel Dis* 2008;**14**:1528–41.

32. Vardi-Saliternik R, Friedlander Y, Cohen T. Consanguinity in a population sample of Israeli Muslim Arabs, Christian Arabs and Druze. *Ann Hum Biol* 2002;**29**:422–31.

33. Quinque D, Kittler R, Kayser M *et al*. Evaluation of saliva as a source of human DNA for population and association studies. *Anal Biochem* 2006;**353**:272–7.

34. Thiele H, Nurnberg P. HaploPainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics* 2005;**21**:1730–2.

35. Purcell S, Neale B, Todd-Brown K *et al*. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75.

36. Manichaikul A, Mychaleckyj JC, Rich SS *et al*. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010;**26**:2867–73.

37. Seelow D, Schuelke M, Hildebrandt F *et al*. HomozygosityMapper: an interactive approach to homozygosity mapping. *Nucleic Acids Res* 2009;**37**:W593–9.

38. Pemberton TJ, Absher D, Feldman MW *et al*. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 2012;**91**:275–92.

39. Yates A, Akanni W, Amode MR *et al*. Ensembl 2016. *Nucleic Acids Res* 2016;**44**:D710–6.

40. McKenna A, Hanna M, Banks E *et al*. The genome analysis toolkit: a map reduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;**20**:1297–303.

41. DePristo MA, Banks E, Poplin R *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;**43**:491–8.

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

43. Lek M, Karczewski KJ, Minikel EV; Exome Aggregation Consortium *et al*. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;**536**:285–91.

44. Adzhubei IA, Schmidt S, Peshkin L *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–9.

45. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;**4**:1073–81.

46. Kircher M, Witten DM, Jain P *et al*. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;**46**:310–5.

47. Scott EM, Halees A, Itan Y; Greater Middle East Variome Consortium *et al*. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet* 2016;**48**:1071–6.

48. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;**47**:D506–15.

49. Wu C, Orozco C, Boyer J *et al*. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 2009;**10**:R130.

50. Rivas MA, Koskela J, Huang H *et al*. Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population. *bioRxiv* 2016;077180.

51. Uhlig HH. Monogenic diseases associated with intestinal inflammation: implications for the understanding of inflammatory bowel disease. *Gut* 2013;**62**:1795–805.

52. Huang H, Fang M, Jostins L; International Inflammatory Bowel Disease Genetics Consortium *et al*. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* 2017;**547**:173–8.

53. Uniken Venema WT, Voskuil MD, Dijkstra G *et al*. The genetic background of inflammatory bowel disease: from correlation to causality. *J Pathol* 2017;**241**:146–58.

54. Hui KY, Fernandez-Hernandez H, Hu J *et al*. Functional variants in the *LRRK2* gene confer shared effects on risk for Crohn's disease and Parkinson's disease. *Sci Transl Med* 2018;**10**:eaai7795.

55. Shlush LI, Behar DM, Yudkovsky G *et al*. The Druze: a population genetic refugium of the near east. *PLoS One* 2008;**3**:e2105.

56. Carmi S, Hui KY, Kochav E *et al*. Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat Commun* 2014;**5**:4835.

57. Zhang W, Hui KY, Gusev A *et al*. Extended haplotype association study in Crohn's disease identifies a novel, Ashkenazi Jewish-specific missense mutation in the NF-κB pathway gene, HEATR3. *Genes Immun* 2013;**14**:310–6.

58. Seiderer J, Schnitzler F, Brand S *et al*. Homozygosity for the CARD15 frameshift mutation 1007fs is predictive of early onset of Crohn's disease with ileal stenosis, entero-enteral fistulas, and frequent need for surgical intervention with high risk of re-stenosis. *Scand J Gastroenterol* 2006;**41**:1421–32.

59. Jürgens M, Brand S, Laubender RP *et al*. The presence of fistulas and NOD2 homozygosity strongly predict intestinal stenosis in Crohn's disease independent of the IL23R genotype. *J Gastroenterol* 2010;**45**:721–31.

60. Zvidi I, Boltin D, Niv Y *et al*. The incidence and prevalence of inflammatory bowel disease in the Jewish and Arab populations of Israel. *Isr Med Assoc J* 2019;**21**:194–7.

61. Daher S, Katz L, Khoury T *et al*. P792 Prevalence of inflammatory bowel diseases in Israeli Jewish and Druze adolescents, a longitudinal comparative study. *J Crohn's Colitis* 2018;**12**:S512–S512.

62. Karban A, Atia O, Leitersdorf E *et al*. The relation between NOD2/CARD15 mutations and the prevalence and phenotypic

heterogeneity of Crohn's disease: lessons from the Israeli Arab Crohn's Disease Cohort. *Dig Dis Sci* 2005;**50**:1692–7.

63. Toyofuku T, Morimoto K, Sasawatari S *et al*. Leucine-rich repeat kinase 1 regulates autophagy through turning on TBC1D2-dependent rab7 inactivation. *Mol Cell Biol* 2015;**35**:3044–58.

64. Hong M, Ye BD, Yang S-K *et al*. Immunochip meta-analysis of inflammatory bowel disease identifies three novel loci and four novel associations in previously reported loci. *J Crohns Colitis* 2018;**12**:730–41.

65. Kast DJ, Zajac AL, Holzbaur ELF *et al*. WHAMM directs the Arp2/3 complex to the ER for autophagosome biogenesis through an actin comet tail mechanism. *Curr Biol* 2015;**25**:1791–7.

66. Zientara-Rytter K, Subramani S. Role of actin in shaping autophagosomes. *Autophagy* 2016;**12**:2512–5.

67. Xu J, Kozlov G, McPherson PS *et al*. A PH-like domain of the Rab12 guanine nucleotide exchange factor DENND3 binds actin and is required for autophagy. *J Biol Chem* 2018;**293**:4566–74.

68. Pfarr N, Prawitt D, Kirschfink M *et al*. Linking C5 deficiency to an exonic splicing enhancer mutation. *J Immunol* 2005;**174**:4172–7.

69. Elmgreen J, Both H, Binder V. Familial occurrence of complement dysfunction in Crohn's disease: correlation with intestinal symptoms and hypercatabolism of complement. *Gut* 1985;**26**:151–7.

70. Smith AM, Rahman FZ, Hayee B *et al*. Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn's disease. *J Exp Med* 2009;**206**:1883–97.

71. Segal AW. Making sense of the cause of Crohn's—a new look at an old disease. *F1000Res* 2016;**5**:2510.

72. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet* 2012;**13**:135–45.

73. Giacopuzzi E, Gennarelli M, Minelli A *et al*. Exome sequencing in schizophrenic patients with high levels of homozygosity identifies novel and extremely rare mutations in the GABA/glutamatergic pathways. *PLoS One* 2017;**12**:e0182778–12.

74. Mukherjee S, Guha S, Ikeda M *et al*. Excess of homozygosity in the major histocompatibility complex in schizophrenia. *Hum Mol Genet* 2014;**23**:6088–95.

75. Johnson EC, Bjelland DW, Howrigan DP; Schizophrenia Working Group of the Psychiatric Genomics Consortium *et al*. No reliable association between runs of homozygosity and schizophrenia in a well-powered replication study. *PLoS Genet* 2016;**12**:e1006343–20.

76. Okada Y, Diogo D, Greenberg JD *et al*. Integration of sequence data from a consanguineous family with genetic data from an outbred population identifies PLB1 as a candidate rheumatoid arthritis risk gene. *PLoS One* 2014;**9**:e87645–12.

77. Mouzan ME, Al-Mofarreh M, Assiri A *et al*. Consanguinity and inflammatory bowel diseases: Is there a relation? *J Pediatr Gastroenterol Nutr* 2013;**56**:182–5.