ORIGINAL ARTICLE

Next generation exome sequencing of paediatric inflammatory bowel disease patients identifies rare and novel variants in candidate genes

Katja Christodoulou,[1] Anthony E Wiskin,[2] Jane Gibson,[1] William Tapper,[1] Claire Willis,[2] Nadeem A Afzal,[3] Rosanna Upstill-Goddard,[1] John W Holloway,[4] Michael A Simpson,[5] R Mark Beattie,[3] Andrew Collins,[1] Sarah Ennis[1]

► Additional materials are published online only. To view these files please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2011-301833).

For numbered affiliations see end of article.

Correspondence to
Dr Sarah Ennis, Genetic Epidemiology and Genomic Informatics Group, Human Genetics, Faculty of Medicine, University of Southampton, Duthie Building (Mailpoint 808), Southampton General Hospital, Southampton SO16 6YD, UK; s.ennis@soton.ac.uk

KC and AEW contributed equally to this study.

Revised 27 March 2012
Accepted 1 April 2012
Published Online First 28 April 2012

ABSTRACT

Background Multiple genes have been implicated by association studies in altering inflammatory bowel disease (IBD) predisposition. Paediatric patients often manifest more extensive disease and a particularly severe disease course. It is likely that genetic predisposition plays a more substantial role in this group.

Objective To identify the spectrum of rare and novel variation in known IBD susceptibility genes using exome sequencing analysis in eight individual cases of childhood onset severe disease.

Design DNA samples from the eight patients underwent targeted exome capture and sequencing. Data were processed through an analytical pipeline to align sequence reads, conduct quality checks, and identify and annotate variants where patient sequence differed from the reference sequence. For each patient, the entire complement of rare variation within strongly associated candidate genes was catalogued.

Results Across the panel of 169 known IBD susceptibility genes, approximately 300 variants in 104 genes were found. Excluding splicing and *HLA*-class variants, 58 variants across 39 of these genes were classified as rare, with an alternative allele frequency of <5%, of which 17 were novel. Only two patients with early onset Crohn's disease exhibited rare deleterious variations within *NOD2*: the previously described R702W variant was the sole *NOD2* variant in one patient, while the second patient also carried the L1007 frameshift insertion. Both patients harboured other potentially damaging mutations in the *GSDMB*, *ERAP2* and *SEC16A* genes. The two patients severely affected with ulcerative colitis exhibited a distinct profile: both carried potentially detrimental variation in the *BACH2* and *IL10* genes not seen in other patients.

Conclusion For each of the eight individuals studied, all non-synonymous, truncating and frameshift mutations across all known IBD genes were identified. A unique profile of rare and potentially damaging variants was evident for each patient with this complex disease.

Significance of this study

What is already known on this subject?

- Genome-wide association studies have implicated numerous candidate genes for inflammatory bowel disease (IBD), but evidence of causality for specific variants is largely absent. Furthermore, by design, genome-wide association studies are limited to the study of common variants and overlook the functionally detrimental variation imposed by rare/novel mutation.
- Exome analysis is fully informative for the spectrum of variation within the protein coding sequence of genes. It has been used to successfully identify disease causing variants in Mendelian disorders, but its potential to identify the missing heritability in complex diseases such as paediatric IBD has not yet been realised.

What are the new findings?

- This study examines genetic variants from the perspective of the patient rather than the gene—for each paediatric case a profile of deleterious variation is determined across a comprehensive panel of known IBD genes.
- Paediatric IBD patients carry a wide spectrum of low frequency variants within candidate IBD genes.
- In silico analyses indicate a substantial proportion of these mutations are potentially deleterious.
- Consistent with complex inheritance, this small subset of patients with severe IBD exhibit a varied profile of mutation with limited sharing of specific variants across the set of eight exomes.

INTRODUCTION

Ulcerative colitis (UC) and Crohn's disease (CD) are the two main clinical phenotypes of inflammatory bowel disease (IBD), both resulting in chronic and relapsing inflammation. The incidence of IBD in the paediatric population of the UK is 5.2 per 100 000 children per year, with breakdown figures of 3.1 for CD, 1.4 for UC and 0.6 for IBD unclassified (IBDU).[1] While the precise aetiology and pathogenesis is complex and incompletely understood, it is widely accepted that IBD occurs as the result of a dysregulated mucosal immune response to commensal gut flora in the genetically susceptible host.[2] Familial aggregation of disease implies a strong genetic component,[3] although

Significance of this study

How might it impact on clinical practice in the foreseeable future?

- Functional studies are required to confirm in silico assessment of variation impact on biology.
- Even mutations confirmed to confer susceptibility must be considered among the full profile of disease predisposing variation present in any individual.
- As the cost of next generation sequencing falls and the number of mutation profiles increases, there is clear potential for genetic characterisation of IBD phenotypic sub-types facilitating targeted therapeutic intervention/personalised medicine.

environmental factors may play a greater role in ulcerative colitis.[4]

Over recent years, genome-wide association studies (GWAS) have been applied with huge success to identify common genes involved in both CD and UC. Genes with replicated evidence for strong association suggest that pathways involving disruption of the innate and adaptive immune system, compromised epithelial barrier function and impaired autophagy play a significant role in disease.[2] However, despite the identification of over one hundred unique genes in IBD susceptibility, these common variants in combination account for less than a quarter of the genetic risk.[5–7] The source of this missing heritability is the subject of much debate with various explanations: overestimates of original heritability statistics; underpowered GWAS studies (in terms of sample size and single nucleotide polymorphism (SNP) coverage) to detect common variants associated with decreasing effect sizes; poorly investigated epistatic and gene–environment interactions; and rare variation.[8]

Rare variants form the group of infrequent mutations that occur in <5% of the population. A large proportion of variants in this class occur at a much lower frequency (<0.1%), and many thousands are likely to be specific to ethnic groups, isolates, families or even individuals. Nevertheless, this class of variation harbours multiple penetrant disease mutations conferring medium to high risk. Rare variants escape detection by GWAS. *BRCA1* and *BRCA2* are examples of familial breast cancer genes that harbour many high risk variants but go undetected by GWAS. This is consequent to each of the disease causing mutations being shared by only a fraction of the patient group and so no common SNP can act as a proxy or 'tag' to flag the gene as causal. It is entirely plausible that a proportion of IBD and other complex disease heritability unaccounted for by common variation lies within higher risk rare variants. Furthermore, many of these mutations may lie within genes already implicated by association studies.

Exome sequencing determines each letter of the genetic code at nearly all coding regions or exons in the genome (the 'exome'), thereby generating the complete profile of coding variation. It has already proved its success in identifying causal mutations in an ever growing list of both recessive and dominant rare Mendelian disorders whereby sequencing of a small number of unrelated cases has been used to identify disease causing variants.[9] One such case reported exome sequencing undertaken in a male child presenting at 15 months with intractable IBD; exome sequencing was used to successfully identify a causal mutation in the *XIAP* gene (X-linked inhibition of apoptosis gene) for which the child was hemizygous. After haematopoietic progenitor cell transplant treatment, as recommended for XIAP deficiency, the IBD resolved, suggesting that the Crohn's-like illness seen in this patient was driven by this single mutation.[10]

As next generation sequencing technology advances, it becomes increasingly affordable. Nevertheless, while costs remain in the region of several hundred pounds per sample, targeted analyses of those patient groups most likely to yield positive results is prudent. Prioritisation of cases with strong family history and/or patients representing the phenotypic 'extreme' of common traits is a useful strategy.[11] One such example of an 'extreme' phenotype is paediatric disease in which onset is particularly early. Genetic susceptibility is thought to play a more important role in the aetiology of early-onset IBD than in late-onset IBD.[12] This is supported by a higher rate of positive family history of IBD in patients with a younger age at diagnosis compared to the older age group, suggesting that an earlier presentation may be due to a higher burden of disease-causing mutations in the genomes of these affected children compared to those in whom disease manifests later in life.[13] In addition, environmental confounding factors such as smoking are less likely to be exerting an influence on disease in paediatric cohorts. It has also been suggested that early-onset disease may in itself be a more aggressive phenotype; in CD, earlier age at diagnosis is associated with a greater need for surgery and increased small bowel disease.[12–14]

Two of the most comprehensive association studies investigating IBD have used adult cohorts, but a recent GWAS of 3246 early-onset IBD cases successfully identified five new loci associated with childhood susceptibility as well as replicating loci previously implicated in adult-onset disease.[15] Early-onset disease genes have also been located using linkage analysis and candidate gene sequencing approaches undertaken in two unrelated consanguineous families.[16] Despite distinct clinical and histopathological features of the CD and UC phenotypes, an estimated 30% of IBD-related loci are shared between both phenotypes.[2] It is likely that further study of rare variation across implicated genes may uncover more commonality.

The application of exome sequencing to complex diseases is fraught with analytical difficulty; finding disease causing variants among the many innocent variants present in the genome has been likened to finding 'needles in stacks of needles'.[17] Targeting analyses to subsets of genes in patients with extreme phenotype is a practical approach to examining genetic influence in disease. In this study we apply next generation sequence technology to paediatric IBD (PIBD). The study is focused on a small cohort of eight paediatric patients with markedly early onset/severe disease. Patients are representative of the spectrum of IBD presentation, and limiting the study to this modest number makes data interpretable on a case-by-case basis. We focus on a comprehensive panel of known causal genes and for each patient describe their individual burden of rare and novel damaging variation.

MATERIALS AND METHODS

Recruitment of paediatric IBD cohort of patients

Children included in this study were selected from the 'Genetics of Paediatric IBD' cohort between October 2010 and October 2011. This cohort was recruited through tertiary referral paediatric IBD clinics at the University Hospital Southampton Foundation Trust. This hospital is the regional centre for paediatric gastroenterology, providing a tertiary paediatric

gastroenterology and endoscopy service for the Wessex region, and draws on a patient population of 3.5 million. The service has a rolling database of over 300 paediatric IBD cases and approximately 50–70 patients are diagnosed each year. All children had a diagnosis of IBD and were aged between 5 and 18 years at time of recruitment, although their diagnosis may have been made at an earlier age. Diagnosis was established using the Porto criteria[18]; all children had compatible history, examination and laboratory investigation results, and infectious causes excluded. All were investigated with upper gastrointestinal endoscopy and ileo-colonoscopy. Written informed consent was obtained from the attending parent of all children, and the child where appropriate. In the initial recruitment interview, clinical data and venous blood samples (10 ml for DNA extraction and 8 ml for plasma extraction) were collected. Additional comprehensive clinical data were extracted from patient records. For each patient we gathered information on gender, dates of birth and initial diagnosis, disease extent currently and at diagnosis using the Paris classification,[19] disease activity score at diagnosis (using the paediatric CD activity index (PCDAI) and the paediatric ulcerative colitis activity index (PUCAI)), height and weight currently and at first diagnosis, time to and date of first relapse, treatment history (use of steroids, immunomodulators, biological therapies, surgery), history of potential aetiological and modifying conditions such as smoking, gastrointestinal infection and other autoimmune disease, and family history.

Ethics statement

This study was approved by the Southampton and South West Hampshire Research Ethics Committee (REC) (09/H0504/125) and University Hospital Southampton Foundation Trust Research & Development (RHM CHI0497).

Selection of samples

Eight patient samples from our PIBD cohort as previously described were selected for exome sequencing for this study. These eight patients were selected based on age of diagnosis, disease severity or positive family history in a first degree relative. Selection criteria and patient phenotypic characteristics are summarised in table 1.

DNA and plasma extraction

Genomic DNA was extracted from EDTA anticoagulated peripheral venous blood samples using the salting out method. Plasma was isolated from lithium–heparin anticoagulated peripheral venous blood samples using standard methods.

Exome sequencing

Targeted exome capture was performed using the SureSelect Human All Exon 50Mb kit (Agilent). The Illumina HiSeq system was used to generate sequence data. These steps were conducted at the Wellcome Trust Centre for Human Genetics at Oxford University. The resultant paired end sequencing data were aligned against the human genome reference sequence 18 (hg18) using the Novoalign software (2.06.09MT, Novocraft Technologies, Selangor, Malaysia). Duplicate reads, resulting from PCR clonality or optical duplicates, and reads mapping to multiple locations were excluded from downstream analysis. Depth and breadth of sequence coverage was calculated with custom scripts and the BedTools package.[20] Single nucleotide substitutions and small insertion deletions were identified and quality filtered within the SamTools software package[21] and in-house software tools. Variants were annotated with respect to genes and transcripts with the Annovar tool.[22] Summary statistics for exome sequencing, mapping and coverage are shown in supplementary table 1 (available online only). Data from the 1000 Genomes Project (1KG) phase I (2010 November release) were utilised using LiftOver (University of California Santa Cruz Genome Browser, http://genome.ucsc.edu/cgi-bin/hgLiftOver) for the conversion of 2010 November coordinates to hg18. Variants were characterised as novel if they were previously unreported in the dbSNP129, dbSNP132, 1KG data and our 22 in-house reference exomes (supplementary table 2). Southampton reference exomes for evaluating the burden of mutation comprised independent DNA samples from unrelated individuals who were exome sequenced on the same platform at the same time as part of other local projects. Each reference exome was from a patient with a distinct clinical diagnosis but no history of gastrointestinal or autoimmune disease. The clinical phenotypes of the 22 reference exomes included 10 with leukaemia, 5 with lymphoma, 4 with Beckwith–Wiedemann syndrome and 3 with macrocephaly malformation syndrome.

The National Heart Lung and Blood Institute Exome Sequencing Project Exome Variant Server (http://evs.gs.washington.edu/EVS/) (Feb 2012) was used as a reference dataset for rare variant allele frequency in a European American population (table 2). This project contains exome data from approximately 3500 European American individuals taken from 12 disease cohorts with a range of heart, lung or blood disorders.

Selection of a panel of known IBD genes

We constructed a panel of high priority genes previously shown to be strongly associated with IBD. Our aim was to include all

Table 1 Summary of patient phenotypes and characteristics (specific selection criteria are in bold)

Sample ID	Age at diagnosis (years)	Sex	Disease	Phenotype description and selection criteria	Ethnicity	Family history
Proband 1	11	Male	CD	**Severe disease requiring surgery**/ Stricturing ileo-colonic disease requiring right hemicolectomy within 6 months of diagnosis.	White British	–
Proband 2	7	Female	CD	**Early age of onset**/non-stricturing, non-penetrating mild to moderate pancolitis, disease resistant to treatment.	White British	–
Proband 3	6	Male	CD	**Early age of onset**/non-stricturing, non-penetrating granulomatous colitis and duodenitis. **Mother diagnosed CD aged 21 years**.	White British	+
Proband 4	6	Female	CD	**Early age of onset**/non-penetrating pancolitis with possible ileo-caecal stricture.	White British	–
Proband 5	13	Male	CD	Non-stricturing, non-penetrating, colitis. **Family history** including maternal CD and maternal grandparental UC.	White British	+
Proband 6	9	Male	UC	**Severe** left sided colitis, also with oral pemphigus.	White British	–
Proband 7	2	Male	UC	**Early age of onset**/mild to moderate pancolitis.	White British	–
Proband 8	3.5	Male	IBDU	**Early age of onset**/left sided colitis.	Iraqi	–

CD, Crohn's disease; IBDU, inflammatory bowel disease unclassified; UC, ulcerative colitis.

Table 2 Characterisation of non-synonymous, stopgain and indel variants with an alternative allele frequency of <0.05 or not reported in 1000 genomes across 39 known IBD genes

Gene	Chromosome	Exon	Variant	Functionally implicated in pathway	Base pair location in hg18	rs ID number in dbSNP 132	Nucleotide change	Protein change	Frequency in 1000 genome	Frequency in NHLBI ESP	Sift score	Grantham Score	Grantham	Polyphen2	Observed n/8	No. homozygote	No. heterozygote	proband 1	proband 2	proband 3	proband 4	proband 5	proband 6	proband 7	proband 8
BACH2	6	7	ns	B-cell regulation	90,717,215	NR	C1331T	S444L	NR	NR	0	145	MR	B	1	0	1						◊		
BACH2	6	7	ns	B-cell regulation	90,717,675	rs61754114	C871G	L291V	0.010	0.030	0	32	C	PrD	1	0	1							◊	
BSN	3	5	ns	Presynaptic cytoskeletal support	49,665,631	rs35762866	G3638A	G1213D	0.035	0.108	0.13	94	MC	PrD	2	0	2			.	.				
BSN	3	5	ns	Presynaptic cytoskeletal support	49,667,511	NR	G5518A	E1840K	NR	NR	0.07	56	MC	PrD	1	0	1							.	
BTNL2	6	6	ns	T-cell negative regulation	32,470,681	rs41521946	C1178A	P393Q	NR	0.003	0.75	76	MC	B	3	0	3		.	.		.			
BTNL2	6	5	ns	T-cell negative regulation	32,471,871	rs28362679	C1001T	S334L	0.014	0.020	0	145	MR	PrD	1	0	1								◊
C1orf93	1	5	ns	Prostaglandin processing	2,509,900	NR	G526A	G176R	NR	NR	0	125	MR	PrD	1	0	1						◊		
CD19	16	13	ns	B-cell receptor signalling	28,857,552	rs34763945	G1544A	R515H	0.025	0.066	0.56	29	C	PrD	2	0	2	.		.					
CDKAL1	6	8	ns	Methylthiotransferase family	20,889,397	NR	G560A	R187K	NR	NR	0.40	26	C	B	1	0	1					.			
CXCR1	2	2	ns	Chemokine receptor	218,737,177	rs16858808	C1003T	R335C	0.018	0.030	0.09	180	R	PrD	2	0	2					.	.		
CXCR1	2	2	ns	Chemokine receptor	218,738,088	rs16858811	T92G	M31R	0.040	0.032	0.60	91	MC	B	2	0	2					.	.		
ERAP2	5	6	ns	Antigen presentation	96,253,828	rs75263594	C1040T	T347M	0.013	0.033	0.01	81	MC	PrD	1	0	1			◊					
ERRFI1	1	4	ns	Epithelial barrier function	7,996,921	rs34781518	G325A	D109N	0.005	0.015	0.14	23	C	B	1	0	1								
FUT2	19	2	ns	Blood group antigen synthesis	53,898,797	rs602662	G772A	G258S	NR	0.515	0.06	56	MC	PrD	5	2	3	.			.	.	.	.	
GMPPB	3	5	ns	Catalyses mannose processing	49,735,146	NR	G448C	E150Q	NR	NR	0.01	29	C	B	1	0	1					◊			
GSDMB	17	7	ns	Unknown	35,316,029	rs35104165	A710G	D237G	0.012	0.036	0	94	MC	B	1	1	0	◊							
HORMAD2	22	2	ns	Unknown	28,819,945	rs34150968	G4A	A2T	0.004	0.012	0	58	MC	PoD	1	0	1					◊			
ICAM1	19	5	ns	Leukocyte adhesion ligand	10,256,141	-	G988A	V330M	0.001	0.003	0	21	C	PrD	1	0	1								◊
ICAM1	19	5	ns	Leukocyte adhesion ligand	10,256,252	-	C1099T	R367C	0.004	0.000	0	180	R	PrD	1	0	1								◊
IL10	1	2	ns	Innate immune recognition	205,011,338	-	C211A	L71M	NR	NR	0.07	15	C	PrD	1	0	1						.		
IL10	1	1	ns	Innate immune recognition	205,012,361	-	G43A	G15R	NR	0.002	0.04	125	MR	PoD	1	0	1							◊	
IL18RAP	2	11	ns	Enhances IL18 binding	102,433,811	-	C1282A	L428M †	NR	0.001	0.15	15	C	PrD	1	0	1	.							
IL18RAP	2	11	ns	Enhances IL18 binding	102,433,812	-	T1283A	L428Q †	NR	0.001	0.15	113	MR	PrD	1	0	1	.							
IL1RL1	2	11	ns	T-helper cell function	102,334,643	rs10192036	C1501A	Q501K †	NR	0.019	1.00	53	MC	B	4	2	2			.	.	.		.	
IL1RL1	2	11	ns	T-helper cell function	102,334,644	rs10204137	A1502G	Q501R †	0.018	0.032	0.61	43	C	B	4	2	2			.	.	.		.	
IL1RL2	2	11	ns	Interleukin receptor	102,217,903	-	C1412T	A471V	NR	NR	0	64	MC	PrD	1	0	1							◊	
JAK2	9	9	ns	Th17-cell differentiation	5,055,003	rs2230723	C1177G	L393V	0.016	0.006	0.38	32	C	B	1	0	1		.						
KIF21B	1	33	ns	Microtubule-binding protein	199,210,518	-	C4722A	D1574E	NR	NR	0.1	45	C	PrD	1	0	1				.				
LRRK2	12	18	ns	Autophagy	38,958,256	rs10878307	A2167G	I723V	0.046	0.070	0.52	29	C	B	2	1	1	.					.		
LTA	6	3	ns	Cytokine receptor interaction	31,648,736	rs2229092	A152C	H51P	0.039	0.072	0.3	77	MC	B	2	0	2					.		.	
MST1	3	17	sg	Apoptosis	49,696,816	-	C1951T	R651X	0.004	0.013	0.14	-	-	-	1	0	1				.				
MST1	3	13	ns	Apoptosis	49,697,765	rs62262682	G1478T	R493L	0.015	0.058	0.09	102	MR	B	1	0	1					.			
MTMR3	22	17	ns	Lipid phosphatase	28,745,983	rs61737780	C2335T	L779F	0.005	0.012	0.21	22	C	PoD	1	0	1					.			
MTMR3	22	17	ns	Lipid phosphatase	28,746,527	rs41278853	A2879G	N960S	0.041	0.086	0.08	46	C	B	1	0	1				.				
NOD2	16	4	ns	Autophagy	49,303,427	rs2066844	C2104T	R702W	0.029	0.047	0	101	MR	PrD	2	0	2	◊		◊					
NOD2	16	9	ns	Autophagy	49,314,777	rs5743291	G2863A	V955I	0.044	0.095	0.46	29	C	B	1	0	1				.				
NOD2	16	11	fi	Autophagy	49,321,282	-	3019_3020insC	L1007fs	NR‡	NR	-	-	-	-	1	0	1			.					
PARK7	1	5	ns	Autophagy	7,953,581	rs71653619	G293A	R98Q	0.003	0.012	0.50	43	C	B	2	0	2				.		.		
PNMT	17	3	sg	Adrenaline processing	35,080,063	-	C744A	Y248X	NR	0.088	0.18	-	-	-	1	0	1						.		
PTGER4	5	3	ns	Epithelial barrier function	40,727,650	rs111866313	G880A	V294I	0.009	0.027	0.51	29	C	B	1	0	1		.						
RTEL1	20	24	ns	DNA repair	61,791,572	rs35640778	G2051A	R684Q	0.003	0.018	0.58	43	C	B	1	0	1							.	
SEC16A	9	23	ns	Endoplasmic reticulum traffic	138,465,668	rs45519739	C6173T	T2058M	NR	0.015	0.01	81	MC	-	1	0	1			◊					
SEC16A	9	3	ns	Endoplasmic reticulum traffic	138,490,409	-	G1480C	G494R	NR	NR	0	125	MR	-	1	0	1						◊		
SEC16A	9	3	ns	Endoplasmic reticulum traffic	138,490,852	-	G1037A	R346H	NR	0.001	0.13	29	C	-	1	0	1		.						
SEC16A	9	3	ns	Endoplasmic reticulum traffic	138,490,870	-	G1019A	G340E	NR	0.002	0.07	98	MC	-	1	0	1	.							
SH2B1	16	1	ns	Adaptor for TYK receptors	28,785,470	-	T554A	L185Q	NR	NR	0.04	113	MR	PrD	1	0	1								◊
SH2B1	16	5	ns	Adaptor for TYK receptors	28,790,787	-	A1495G	T499A	NR	NR	0.19	58	MC	PoD	1	0	1		.						
SMAD3	15	3	ns	TGF-β signalling pathway	65,244,752	rs35874463	A376G	I126V	0.018	0.053	0.65	29	C	B	2	0	2	.	.						
SNAPC4	9	17	ns	RNA polymerase transcription	138,396,228	rs34569521	G2186A	R729Q	0.046	0.088	0	43	C	PrD	1	0	1							◊	
SNAPC4	9	10	ns	RNA polymerase transcription	138,402,740	-	G1100C	G367A	NR‡	0.003	0	60	MC	PrD	1	0	1							◊	
SP140	2	24	ns	Nuclear body protein	230,883,793	rs62192163	T2266C	C756R	0.048	0.245	0.44	180	R	B	1	1	0	.							
SULT1A2	16	2	ns	Sulphate conjugation	28,514,733	rs1136703	T20C	I7T	0.014	0.059	0.12	89	MC	B	1	0	1	.							
TAGAP	6	6	ns	T cell regulation	159,382,412	rs41267765	G439A	E147K	0.014	0.020	0.64	56	MC	B	1	0	1	.							
TAGAP	6	5	ns	T cell regulation	159,383,130	-	G283A	G95S	NR	NR	0.58	56	MC	B	1	0	1	.							
THADA	2	28	nd	Apoptosis	43,508,785 *	-	4014_4016del	1338_1339del	NR	NR	-	-	-	-	1	0	1	.							
TYK2	19	20	ns	Th17-cell differentiation	10,325,843	rs35018800	C2783T	A928V	0.003	0.008	0	64	MC	PrD	1	0	1				◊				
TYK2	19	8	ns	Th17-cell differentiation	10,336,649	rs2304255	G1087A	G363S	0.035	0.080	0.54	56	MC	B	1	0	1								.
ZNF365	10	3	ns	Zinc finger	64,084,667	-	C97A	L33I	0.006	0.000	0	5	C	U	1	0	1	◊							

Novel variants are shown in grey.
Where a specific variant is present in a proband, this is indicated by a dot (.).
Where a specific variant is present in a proband and has a SIFT score of <0.05 this is indicated by a ◊.
*Indicates the first bp location of a 3-bp deletion.
†Indicates a dinucleotide variant (that for *IL18RAP* results in a codon change from CTG>AAG, resulting in p.L428K amino acid change).
‡NR indicates variants that despite not being reported in dbSNP132 or 1000 genomes, are reported in dbSNP129 or seen in our in-house reference exomes and are therefore not characterised as novel.
B, benign; C, conservative; fi, frameshift insertion; MC, moderately conservative; MR, moderately radical; nd, non-frameshift deletion; NR, not reported; ns, non-synonymous; PoD, possibly damaging; PrD, probably damaging; R, radical; sg, stopgain; U, unknown.

genes with convincing evidence for disease causality in previous studies. Selection was based on the findings of two genome-wide meta-analyses of IBD,[5 6] one genome-wide association study of early-onset IBD,[15] and one linkage study in consanguineous families with early-onset IBD.[16] Gene names were cross referenced with the Human Genome Nomenclature Committee to ensure that the most up-to-date versions of gene names were applied (http://www.genenames.org/). Our consolidated panel represented 169 genes (supplementary table 3).

Evaluation of spectrum of mutation and predicted functional impact

Exome data from our eight patients were cross-referenced against our gene panel described above. Synonymous variants

were excluded from analysis due to their decreased likelihood of functional effect on protein. SIFT ('sorting intolerant from tolerant') scores[23] were annotated using Annovar, or where scores were missing, were derived indirectly using the database of non-synonymous functional prediction.[24] A small number of additional missing scores were obtained from the SIFT server at http://sift.jcvi.org. SIFT is a sequence homology-based tool that predicts whether an amino acid substitution is likely to affect protein function. Variants with SIFT scores of <0.05 are considered 'deleterious', and SIFT therefore allows prioritisation of amino acid changes by ranking according to score.

We examined in silico predictions from the Polyphen2 (Polymorphism Phenotyping v2) server at http://genetics.bwh.harvard.edu/pph2/bgi.shtml.[25] Polyphen2 uses a probability model to generate thresholds and classify polymorphisms as benign, possibly damaging or probably damaging, based on 11 predictive features relating to sequence, phylogenetic and structural information which characterise the substitution. Additional functional predictions of the result of each amino acid change were derived from Grantham scores,[26] which predict the effect of amino acid substitutions according to chemical properties including polarity and molecular volume. The Grantham distance, d, between two amino acids is classified as conservative ($0<d\leq50$), moderately conservative ($50<d\leq100$), moderately radical ($100<d\leq150$) or radical ($d>150$).[27] Radical changes predicted by these scores are linked to clinical phenotypes.[28]

Burden of mutation

Using only novel variants or variants with an alternative allele frequency of <0.05 in the 1000 genomes data, a χ^2 contingency test was performed to test for an excess of rare potentially deleterious variants (non-synonymous and frameshift indels) compared to neutral synonymous variants, within the panel of known IBD genes in our eight cases compared to 22 reference exome samples from non-IBD patients.

RESULTS

Exome sequencing

On average, each PIBD exome had 78% of mappable bases of the Gencode defined exome represented by coverage of at least 20 reads (supplementary table 1). For each patient approximately 23 000 variants were found. After exclusion of synonymous variants, approaching 13 000 variants were found per patient, of which approximately 300 were novel (supplementary table 2).

Characterisation of mutations in genes known to be associated with IBD

Across all eight exomes, we found 332 variants (excluding synonymous) among 104 of our panel of 169 genes (supplementary table 4). Of these, approximately 40% (122) were found in HLA class genes. Seventeen were novel variants not previously reported in public databases or our own in-house database of non-IBD patient reference exomes.

Table 2 describes the set of variants remaining after removal of splicing, common (where the alternative allele frequency in 1000 genomes is reported as >0.05) and *HLA* variants. Fifty-eight variants within 39 genes remain, of which 17 are novel.

The χ^2 analysis to test for an excess of deleterious rare variants in known and candidate IBD genes in IBD cases listed in table 2 compared to 22 reference exomes did not reach statistical significance (supplementary table 5).

Crohn's disease patient profiles

Only two patients with early onset CD exhibit rare potentially deleterious variations within *NOD2*.

Proband 1 was diagnosed with CD aged 11 years and required a right hemicolectomy for extensive ileo-caecal stricturing. He is a heterozygote carrier of the *NOD2* R702W variant that is associated with a twofold increase in odds ratio of CD.[29] In addition he harbours potentially damaging mutations in *GSDMB* and *ZNF365* and a dinucleotide variant of undetermined functionality on one chromosomal copy of the *IL18RAP* gene. The presence of ileal disease and a stenotic phenotype in this patient is also consistent with his *NOD2* variant profile.[29]

Proband 2 carries a novel variant in each of the *SEC16A* and *SH2B1* genes. This patient also has a rare variant in *JAK2*; however, SIFT scoring suggests none of these mutations are likely to be particularly deleterious.

Proband 3 is the second patient with *NOD2* variation and carries both the R702W variant and the L1007 frameshift insertion. Carriage of two or more high risk alleles in *NOD2* confers a 17-fold increased risk of IBD.[29] Exome analysis cannot determine if both variants have been co-inherited on the same chromosome. Proband 3 additionally possesses potentially deleterious variants in *ERAP2* and *SEC16A*.

Proband 4 presented with severe disease aged 6 years. She carries the *NOD2* V955I variant, but this is predicted to be innocuous as is her private variant in *KIF21B*. She is a heterozygote for a number of previously seen variants with borderline (~0.05) SIFT scores (*FUT2*, *MTMR3*). The most distinct rare (frequency of 0.003) and potentially deleterious variant observed in this patient is the A928V variant in the *TYK2* gene.

Proband 5 possesses one variant in the *GMPBB* gene and another in *HORMAD2*, both estimated by SIFT to be harmful. The former is ascertained as novel to this individual, whereas the latter occurs in <0.5% of chromosomes studied in the thousand genomes project, but in just over 1% of the 3500 exomes tested in Exome Variant Server.

UC and IBDU patient profiles

Proband 6 has a histological diagnosis of UC and carries novel deleterious mutations in the *BACH2*, *C1orf93* and *SEC16A* genes. A fourth novel variant in the *IL10* gene also has a low SIFT score.

Proband 7 is a boy, diagnosed aged 2, and similar to our other UC patient, exhibits a potentially functionally detrimental mutation in *BACH2* and a second very rare and possibly damaging mutation in *IL10*. The *IL1RL2* and *SNAPC4* genes are also apparently compromised in this individual.

Proband 8 was diagnosed at a young age with IBDU, and possesses two possibly harmful variants in *ICAM1*, one in *BTNL2* and a novel deleterious variant in *SH2B1*.

Predicted functional impact

Figure 1 illustrates relationships between SIFT, Grantham and Polyphen2 scores for all non-synonymous variants in table 2. There is particularly close agreement between SIFT and Polyphen2 scores as noted previously.[30] Agreement with Grantham scores is less clear, but there is striking concordance between the vast majority of variants with a SIFT score >0.2 (benign) being independently designated benign by Polyphen2 and conservative by Grantham. Notably, two variants are classified as radical by Grantham and probably damaging by SIFT and/or Polyphen2—*CXCR1* (R335C) and *ICAM1* (R367C)—with the latter being classified as radical/damaging by all three criteria.

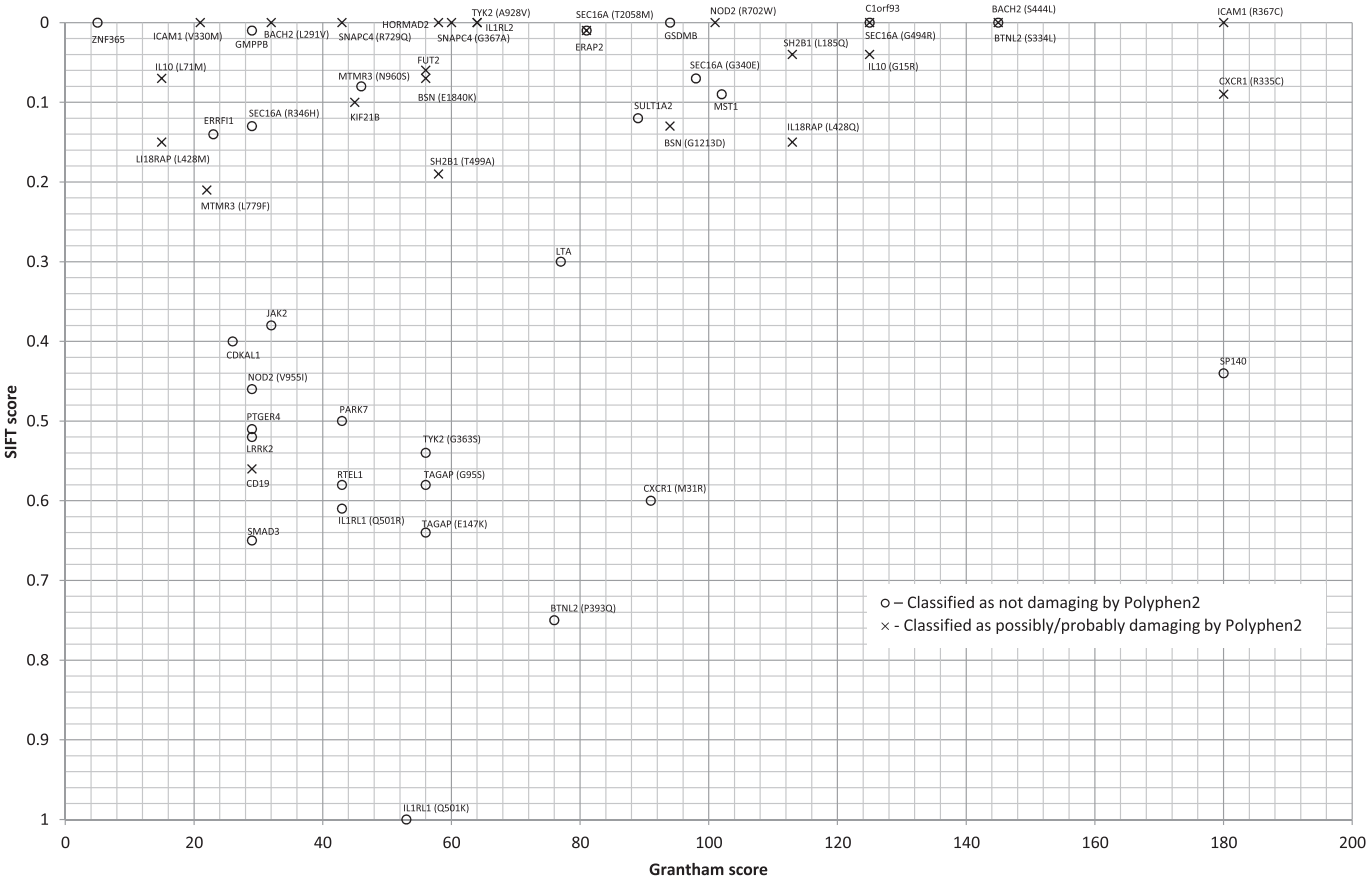

Figure 1 In silico functional predictions.

DISCUSSION

In this study we have applied exome sequencing, which allows the screening of the complete spectrum of variation within protein coding genes. There is abundant evidence that such regions are likely to be highly enriched for disease causing variation.[31] We have focused on the identification of rare and novel variation within genes known to contain causal variants or identified as candidate genes for IBD. Excluding *HLA* variants and considering only rare non-synonymous, stop-gain mutations and indels, we uncovered 58 variants across 39 genes, of which 17 were not previously reported. Of these, 35% (20 variants) have SIFT scores under 0.05, 12 of these are also classified as probably damaging by Polyphen2 and five of these (*BTNL2*: S334L; *C1orf93*: G176R; *ICAM1*: R367C; *NOD2*: R702W and *SH2B1*: L185Q) are also classified as moderately radical or radical by Grantham score. One variant, *CXCR1* (R335C), has a borderline SIFT score of 0.09 and is classified as probably damaging by Polyphen2 and radical by Grantham score. These variants may compromise protein function and contribute to the PIBD phenotype in these patients.

Our study included five patients with childhood onset CD. The variant profiles show that four of these patients carry potentially deleterious mutations in one or more IBD candidate genes. One child had a 17-fold increased risk of IBD on the basis of his *NOD2* profile alone. Others in this group bear variants with likely impact on antigen presentation (*ERAP2*), endoplasmic reticulum trafficking (*SEC16A*) and T-helper cell differentiation. A variant in the *IL18RAP* gene was recently reported by Rivas *et al*[7] to carry a threefold OR for CD, and variants in the same gene have also been implicated in coeliac disease.[32] We identify a rare, non-synonymous, two-base pair mutation in this gene in one of our severely affected early onset CD cases. Our study examined only two patients with a clear diagnosis of UC and intriguingly we observe unique, potentially deleterious variation in both the B-cell regulatory gene *BACH2* and *IL10* genes in both patients. Interestingly, defective *IL10* functioning is already recognised in UC pathogenesis,[33 34] whereas although other components of B-cell signalling (*IL7R* and *IRF5*) have shown previous association with UC,[6] variation in *BACH2* has shown previous association with CD only. Our patient with undetermined IBD is the only patient with rare *ICAM1* variants. This gene, in which our IBDU patient carries two functionally damaging variants, plays a role in cell-mediated inflammation and has been identified as a therapeutic target in IBD.[35]

Assessing our results obtained for each individual in our cohort with IBD, we can see clearly that it is possible to generate an individualised variant profile for each patient. Individualised profiles are already being usefully applied to refine disease diagnosis. For example, Franke *et al*[36] reported recently on a whole genome sequencing undertaken on a 47-year-old patient diagnosed with CD in her 20s. Her case was particularly severe, as she had failed standard treatments including anti-TNF, had undergone multiple bowel resections, and required intermittent parenteral nutrition. Sequencing in this patient revealed multiple 'hits' in the autophagy pathways. This prompted in-depth mycobacterial diagnostics and ultimately resulted in a diagnosis of chronic active *Mycobacterium avium* infection.

Although suggestive and interesting mutation profiles have emerged from our small panel, it is clear that our picture is far from complete. Proband 3 displays rare variation across many genes, but not one of these appears to have potential functional consequence. Furthermore, in 65 genes previously linked to IBD, we identified no variants in our eight probands. It is possible

that these genes do not contribute to disease in this small group, consistent with a high degree of genetic heterogeneity in this complex disease. It is also possible that limitations of sequencing technology or the analytical pipeline could have resulted in failure to call true variants. By focusing our analysis on exomes, we rely on the fact that many of the non-coding SNP variants previously implicated by GWAS simply flag coding variants in the genomic vicinity. Protein-coding genes harbour about 85% of the mutations with large effects for disease-related traits,[37] but it is entirely possible that restriction of the exome capture to coding regions might have overlooked non-coding variants with significant impact on protein expression. By tabulating rare and novel variants, we are focusing attention on those variants hypothesised to have larger effect sizes on the assumption that such variants confer significant genetic contribution to childhood severe and/familial disease.[38] However, for any complex disease, multiple common susceptibility variants, each contributing very modest effect sizes, should not be ignored.

SIFT, Polyphen2 and Grantham scores provide an indication of potential causality but they must be interpreted with caution, particularly for complex traits. Kumar *et al*[39] describe in silico prediction such as SIFT as effective for monogenic disease, but consider such tools to be less effective for lower penetrance variants associated with complex diseases. Furthermore, one study compiled in silico prediction scores and found pairwise agreement between all methods to be in the range 60–70%, implying fairly substantial disagreement.[24] These and other studies underpin the difficulty in ascribing functional evidence and translational importance of genetic variants, and the particular difficulty in heterogeneous complex disease. However, it is notable that published evidence demonstrates a clear functional impact for two of the six variants listed above as having an overall deleterious score by two or more of the in silico measures. The *CXCR1* gene R335C variant has been previously implicated in chronic obstructive pulmonary disease and asthma.[40] The two *CXCR1* mutations listed in table 2 (R335C and M31R) are in tight linkage disequilibrium and both are known to alter the structure and charge of the protein at the respective positions. The N-terminus of CXCR1 protein has been identified as potentially important for receptor–ligand binding, leading to the suggestion that the M31R variant may affect this interaction. This led to the hypothesis that both polymorphisms could impact receptor function through alterations in structure.[41] The upper right quadrant of figure 1 indicates those variants where all three *in silico* prediction tools are concordant in ascribing detrimental effects of the variant. Mutations such as the rare R376C *ICAM1* variant may modify the function of the encoded glycoprotein expressed on immune and endothelial cells and should be prioritised for functional assessment. Another non-synonymous variant highlighted by the in silico scores is the *NOD2* R702W variant which, together with the *NOD2* L1007fs variant, has been found to impair the activation of the NF-κB pathway in response to muramyl dipeptide (MDP), a bacterial wall component, with the L1007fs mutant unable to respond.[42] NOD2 is localised to the cell membrane but the L1007fs polymorphism disrupts this association and thus the protein has cytoplasmic distribution. Forcing the L1007fs mutant protein to associate with the plasma membrane does not lead to activation of the NF-κB pathway in response to MDP; thus it is not the localisation of the NOD2 mutant, but rather an inability to respond to MDP, that affects induction of the NF-κB pathway. The L1007fs mutation has been shown to produce a truncated protein with impaired function.[43] The *NOD2* R702W variant occurred in four of the 22 non-IBD reference exomes, representing a higher than expected frequency. Although the reference exomes were composed of germline DNA from patients with diverse diagnoses (various lymphomas, leukaemias and congenital growth disorders), all four of these IBD negative controls had a diagnosis of chronic lymphocytic leukaemia. Interestingly, a population based cohort study of 47 679 Swedish patients with CD or UC, reported a 20% increased risk of haematopoietic cancers in these patients.[44] However, the role of *NOD2* polymorphisms has been further investigated in a variety of cancers, with most finding no association.[45] Recently, however, Sivakumaran *et al*[46] found abundant evidence for pleiotropy in complex disease, defined as one gene having an effect on multiple phenotypes. The authors identified many genes harbouring variants associated with CD and other immune-mediated phenotypes. These associations include a CD association with chronic lymphocytic leukaemia, through the *SP140* gene (within which a rare variant is listed in table 2). Other gene/disease associations linked with CD include *BACH2* with type 1 diabetes and coeliac disease, *IL18RAP* in coeliac disease, *IL1RL1* with eosinophil count and coeliac disease, *MST1* with UC and primary sclerosing cholangitis, *ZNF365* with breast cancer, and *NOD2* with leprosy, among many others.[47] All of these genes contain rare variants listed in table 2 within the eight patients we have exome sequenced.

The abundance of potentially damaging variants arising from next generation sequencing renders interpretation of the potential impact of disease challenging. However, focusing on early onset and other forms of 'severe' phenotype, including familial cases, coupled with our ability to filter variants identified with increasingly large and reliable databases of apparently neutral variants, offers the prospect of identifying important rare variants involved in complex traits such as IBD. This is the first study whereby a cohort of patients have been exome sequenced with the specific aim of generating a unique and personalised profile of rare variants across known disease genes for each patient. The rare variant profiles presented here provide a relatively small number of potential causal variants and include many mutations classed as deleterious by in silico prediction, a number of potential compound heterozygotes and a number of variants for which there is established functional evidence of roles in disease. These data, assessed from the perspective of individual patients, provide one of the first glimpses of personal mutation profiles and establish a foundation to elucidate the disease significance of these variants in future next-generation sequencing analyses of PIBD patients.

Author affiliations
[1]Genetic Epidemiology and Genomic Informatics Group, Human Genetics & Genomic Medicine, Faculty of Medicine, University of Southampton, Duthie Building (Mailpoint 808), University Hospital Southampton NHS Foundation Trust, Southampton, UK
[2]NIHR Biomedical Research Unit (Nutrition, Diet & Lifestyle), University Hospital Southampton NHS Foundation Trust, Mailpoint 218, Southampton General Hospital, Tremona Road, Southampton, UK
[3]Paediatric Medical Unit, University Hospital Southampton NHS Foundation Trust, Southampton General Hospital, Tremona Road, Southampton, UK
[4]Human Genetics & Genomic Medicine, Human Genetics, Faculty of Medicine, University of Southampton Duthie Building (Mailpoint 808), University Hospital Southampton NHS Foundation Trust, Southampton, SO16 6YD, UK
[5]Division of Genetics and Molecular Medicine, King's College London School of Medicine, Guy's Hospital, London, UK

Acknowledgements The authors would like to thank Nikki J Graham from the DNA laboratory in Human Genetics & Genomic Medicine, University of Southampton; and David Buck and Lorna Gregory from the Wellcome Trust Centre for Human Genetics, Oxford University.

Contributors KC was responsible for analysis, and with AEW, interpretation of data, drafting of the manuscript, critical revision of article and final approval. RMB, NA and

CW were responsible for acquisition of data, critical revision and final approval of article. AC, JG, RU-G, WT, JWH and MS were responsible for interpretation of data, critical revisions and final approval. SE was responsible for conception, design, acquisition of data, analysis and interpretation of data, drafting, revision and approval of the final manuscript.

Funding This project was supported by: NIHR Biomedical Research Unit (Nutrition, Diet & Lifestyle), University Hospital Southampton NHS Foundation Trust with specific thanks to Liz Blake, Senior Paediatric Research Sister, and Rachel Haggarty, Senior Children's Research Nurse; University Hospital Southampton Foundation Trust R&D; and the Crohn's in Childhood Research Association (CICRA).

Competing interests None.

Patient consent Obtained.

Ethics approval This study was approved by the Southampton & South West Hampshire Research Ethics Committee (REC) (09/H0504/125).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Sawczenko A,** Sandhu BK, Logan RF, *et al*. Prospective survey of childhood inflammatory bowel disease in the British Isles. *Lancet* 2001;**357**:1093—4.
2. **Khor B,** Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011;**474**:307—17.
3. **Bengtson MB,** Solberg C, Aamodt G, *et al*. Familial aggregation in Crohn's disease and ulcerative colitis in a Norwegian population-based cohort followed for ten years. *J Crohns Colitis* 2009;**3**:92—9.
4. **Spehlmann ME,** Begun AZ, Burghardt J, *et al*. Epidemiology of inflammatory bowel disease in a German twin cohort: results of a nationwide study. *Inflamm Bowel Dis* 2008;**14**:968—76.
5. **Franke A,** McGovern DP, Barrett JC, *et al*. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;**42**:1118—25.
6. **Anderson CA,** Boucher G, Lees CW, *et al*. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 2011;**43**:246—52.
7. **Rivas MA,** Beaudoin M, Gardet A, *et al*. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011;**43**:1066—73.
8. **Bodmer W,** Tomlinson I. Rare genetic variants and the risk of cancer. *Curr Opin Genet Dev* 2010;**20**:262—7.
9. **Gilissen C,** Hoischen A, Brunner HG, *et al*. Unlocking Mendelian disease using exome sequencing. *Genome Biol* 2011;**12**:228.
10. **Worthey EA,** Mayer AN, Syverson GD, *et al*. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med* 2011;**13**:255—62.
11. **Day-Williams AG,** Zeggini E. The effect of next-generation sequencing technology on complex trait research. *Eur J Clin Invest* 2011;**41**:561—7.
12. **de Ridder L,** Weersma RK, Dijkstra G, *et al*. Genetic susceptibility has a more important role in pediatric-onset Crohn's disease than in adult-onset Crohn's disease. *Inflamm Bowel Dis* 2007;**13**:1083—92.
13. **Biank V,** Broeckel U, Kugathasan S. Pediatric inflammatory bowel disease: clinical and molecular genetics. *Inflamm Bowel Dis* 2007;**13**:1430—8.
14. **Lacher M,** Kappler R, Berkholz S, *et al*. Association of a CXCL9 polymorphism with pediatric Crohn's disease. *Biochem Biophys Res Commun* 2007;**363**:701—7.
15. **Imielinski M,** Baldassano RN, Griffiths A, *et al*. Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat Genet* 2009;**41**:1335—40.
16. **Glocker EO,** Kotlarz D, Boztug K, *et al*. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med* 2009;**361**:2033—45.
17. **Cooper GM,** Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* 2011;**12**:628—40.
18. **IBD Working Group of the European Society for Paediatric Gastroenterology, Hepatology and Nutrition.** Inflammatory bowel disease in children and adolescents: recommendations for diagnosis—the Porto criteria. *J Pediatr Gastroenterol Nutr* 2005;**41**:1—7.
19. **Levine A,** Griffiths A, Markowitz J, *et al*. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflamm Bowel Dis* 2011;**17**:1314—21.
20. **Quinlan AR,** Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841—2.
21. **Li H,** Handsaker B, Wysoker A, *et al*. The sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078—9.
22. **Wang K,** Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;**38**:e164.
23. **Ng PC,** Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;**31**:3812—14.
24. **Liu X,** Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* 2011;**32**:894—9.
25. **Adzhubei IA,** Schmidt S, Peshkin L, *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248—9.
26. **Grantham R.** Amino acid difference formula to help explain protein evolution. *Science* 1974;**185**:862—4.
27. **Li WH,** Wu CI, Luo CC. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J Mol Evol* 1984;**21**:58—71.
28. **Botstein D,** Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003;(33 Suppl):228—37.
29. **Economou M,** Trikalinos TA, Loizou KT, *et al*. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am J Gastroenterol* 2004;**99**:2393—404.
30. **Rudd MF,** Williams RD, Webb EL, *et al*. The predicted impact of coding single nucleotide polymorphisms database. *Cancer Epidemiol Biomark Prev* 2005;**14**:2598—604.
31. **Lehne B,** Lewis CM, Schlitt T. Exome localization of complex disease association signals. *BMC Genomics* 2011;**12**:92.
32. **Dubois PC,** Trynka G, Franke L, *et al*. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;**42**:295—302.
33. **Franke A,** Balschun T, Karlsen TH, *et al*. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet* 2008;**40**:1319—23.
34. **Festen EA,** Stokkers PC, van Diemen CC, *et al*. Genetic analysis in a Dutch study sample identifies more ulcerative colitis susceptibility loci and shows their additive role in disease risk. *Am J Gastroenterol* 2010;**105**:395—402.
35. **Philpott JR,** Miner PB Jr. Antisense inhibition of ICAM-1 expression as therapy provides insight into basic inflammatory pathways through early experiences in IBD. *Expert Opin Biol Ther* 2008;**8**:1627—32.
36. **Franke A,** Kuehbacher T, Nikolaus S, *et al*. The complete individual genome of a Female Crohn's disease patient—What can you Learn? *Gastroenterol* 2011;**140** (5 Suppl 1):S-90.
37. **Majewski J,** Schwartzentruber J, Lalonde E, *et al*. What can exome sequencing do for you? *J Med Genet* 2011;**48**:580—9.
38. **Bodmer W,** Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008;**40**:695—701.
39. **Kumar S,** Dudley JT, Filipski A, *et al*. Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends Genet* 2011;**27**:377—86.
40. **Stemmler S,** Arinir U, Klein W, *et al*. Association of interleukin-8 receptor alpha polymorphisms with chronic obstructive pulmonary disease and asthma. *Genes Immun* 2005;**6**:225—30.
41. **Vasilescu A,** Terashima Y, Enomoto M, *et al*. A haplotype of the human CXCR1 gene protective against rapid disease progression in HIV-1+ patients. *Proc Natl Acad Sci U S A* 2007;**104**:3354—9.
42. **Lecine P,** Esmiol S, Metais JY, *et al*. The NOD2-RICK complex signals from the plasma membrane. *J Biol Chem* 2007;**282**:15197—207.
43. **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.
44. **Askling J,** Brandt L, Lapidus A, *et al*. Risk of haematopoietic cancer in patients with inflammatory bowel disease. *Gut* 2005;**54**:617—22.
45. **Yazdanyar S,** Nordestgaard BG. NOD2/CARD15 genotype, cardiovascular disease and cancer in 43,600 individuals from the general population. *J Intern Med* 2010;**268**:162—70.
46. **Sivakumaran S,** Agakov F, Theodoratou E, *et al*. Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet* 2011;**89**:607—18.
47. **Lees CW,** Barrett JC, Parkes M, *et al*. New IBD genetics: common pathways with other diseases. *Gut* 2011;**60**:1739—53.

Supplementary material to:

Next generation exome sequencing of paediatric inflammatory bowel disease patients identifies rare and novel variants in candidate genes.

Katja Christodoulou, Anthony E Wiskin, Jane Gibson, William Tapper, Claire Willis , Nadeem A Afzal, Rosanna Upstill-Goddard, John W Holloway, Michael A Simpson, R Mark Beattie, Andrew Collins & Sarah Ennis

Contents:

Patient Vignettes

Proband 1
Crohn's disease diagnosed aged 11 years presenting with a one year history of intermittent abdominal pain, decreased appetite, loose stools (including nocturnal stooling) and poor height and weight gain. Investigation showed ileo-colonic disease with stricturing disease in the proximal ileum. Histology demonstrated chronic inflammation with colonic granulomata and relative preservation of crypt architecture. He received an initial treatment course of exclusive enteral nutrition and was started on azathioprine. A **right hemi-colectomy** was performed within six months of diagnosis for persistent stricturing disease with pre stenostic dilatation. His disease has subsequently been well controlled with azathioprine.

Proband 2
Crohn's disease **diagnosed aged 7 years** presenting with a 6 month history of intermittent abdominal pain, loose bloody stools and static weight. Initial investigation demonstrated mild patchy pancolitis. Recurrent histology has shown preservation of glandular architecture but moderately active colitis. Her disease was resistant to treatment with exclusive enteral nutrition and several courses of corticosteroids in combination with azathioprine. Induction with Infliximab improved her symptoms but after one year of maintenance therapy her symptoms returned. She has responded to higher dosing.

Proband 3
Crohn's disease **diagnosed aged 6 years** presenting with a two month history of weight loss, abdominal pain and vomiting. **Positive family history** (maternal Crohn's disease diagnosed age 21). Histology demonstrated granulomatous inflammation in the stomach, ileum and colon. Tuberculosis and immunodeficiency were excluded. He responded well to treatment with exclusive enteral nutrition and has been well to date. He has multiple IgE mediated food allergies.

Proband 4
Crohn's disease **diagnosed aged 6 years** presenting with an eight month history of diarrhoea and acute cryptosporidium infection. She was severely malnourished at presentation with acute weight loss, abdominal pain and worsening diarrhoea and was dependent on parenteral nutrition for several weeks. Endoscopy and histology demonstrated patchy colitis with preservation of crypt architecture which has been confirmed on repeat endoscopy. Treatment with corticosteroids was successful and she has been well subsequently.

Proband 5
Crohn's disease diagnosed aged 13 years presenting with a six month history of diarrhoea (including nocturnal stooling) abdominal pain and mouth ulcers plus a **positive family history** of IBD (maternal Crohn's disease and grandmaternal Ulcerative Colitis). Endoscopy and histology showed patchy colonic inflammation with relative preservation of crypt architecture. His disease has been successfully managed with amino-salicylates.

Proband 6
Ulcerative Colitis (left sided) diagnosed aged 9 years presenting with a two month history of bloody diarrhoea. Histology demonstrated crypt abscesses and cryptitis with diffuse inflammatory cell

infiltrate. He also has **oral pemphigus which presented at age 11 years with severe oral ulceration.** He responded to corticosteroids and longer term aminosalicylate and azathioprine as maintenance.

Proband 7
Ulcerative Colitis (pancolitis) **diagnosed aged 2 years.** Histology demonstrated widespread crypt distortion with cryptitis and increased inflammatory cells more pronounced distally. He required prolonged treatment with corticosteroids and azathioprine to achieve remission but remains well on azathioprine. He also has primary hypothyroidism although auto-antibody screen is repeatedly negative.

Proband 8
Colitis (left sided) classified as IBDU **diagnosed aged 3 years** presenting with a four month history of bloody diarrhoea. Histology showed active colitis with occasional crypt abscesses and no granulomata. He responded to initial treatment with corticosteroids and has been maintained in remission on amino-salicylates.

Supplementary Table 1: Summary statistics for exome sequencing - mapping and coverage

Sequenced exomes	Proband 1	Proband 2	Proband 3	Proband 4	Proband 5	Proband 6	Proband 7	Proband 8
Total no. read seqs	50,583,874	54,278,594	51,698,058	61,686,454	46,971,966	108,712,050	73,873,640	72,246,856
Total no. aligned reads	49,651,350	53,141,656	50,904,320	60,644,577	45,894,857	106,505,676	72,596,746	70,939,697
Total no. unique alignments	45,762,617	49,013,160	47,027,836	56,068,870	42,335,161	98,532,568	67,085,167	65,387,188
Mapped to target reads +/-150bp (%)	72.23	72.45	73.89	73.43	66.38	67.89	70.12	69.80
Mapped to target reads (%)	65.51	65.34	67.04	66.17	59.71	61.48	63.73	63.36
Target bases with coverage >1 (%)	98.57	97.80	97.93	98.20	98.22	98.92	98.54	98.60
Target bases with coverage >5 (%)	92.38	91.36	91.78	92.63	91.77	94.89	93.48	93.53
Target bases with coverage >10 (%)	86.74	85.10	86.06	87.43	85.45	91.02	88.81	88.77
Target bases with coverage >20 (%)	75.81	72.00	75.43	77.94	72.73	85.12	81.04	80.76
Mean read depth across exome	46.87	40.20	51.07	54.63	41.72	99.67	70.76	68.03

Supplementary Table 2: Summary statistics for exome sequencing - number of variants of different classes identified by exome sequencing in eight PIBD cases

Variant type	Proband 1			Proband 2			Proband 3			Proband 4			Proband 5			Proband 6			Proband 7			Proband 8		
	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel
Synonymous	10,100	9,993	107	10,096	9,974	122	10,231	10,145	86	10,255	10,149	106	10,038	9,926	112	10,588	10,477	111	10,339	10,226	113	10,734	10,506	228
Heterozygous	6,130	6,030	100	6,046	5,925	121	6,223	6,139	84	6,317	6,212	105	6,040	5,935	105	6,362	6,256	106	6,218	6,111	107	6,614	6,400	214
Homozygous	3,970	3,963	7	4,050	4,049	1	4,008	4,006	2	3,938	3,937	1	3,998	3,991	7	4,226	4,221	5	4,121	4,115	6	4,120	4,106	14
Non-synonymous	9,420	9,204	216	9,452	9,246	206	9,589	9,405	184	9,542	9,329	213	9,366	9,151	215	9,678	9,457	221	9,784	9,591	193	10,037	9,679	358
Heterozygous	5,940	5,733	207	5,956	5,753	203	5,937	5,757	180	5,986	5,776	210	5,836	5,642	194	5,843	5,629	214	6,097	5,912	185	6,372	6,034	338
Homozygous	3,480	3,471	9	3,496	3,493	3	3,652	3,648	4	3,556	3,553	3	3,530	3,509	21	3,835	3,828	7	3,687	3,679	8	3,665	3,645	20
Frameshift indel	184	172	12	189	176	13	188	175	13	184	179	5	171	162	9	189	178	11	193	182	11	197	191	6
Heterozygous	56	44	12	74	61	13	59	48	11	50	45	5	56	48	8	51	40	11	62	52	10	62	56	6
Homozygous	128	128	0	115	115	0	129	127	2	134	134	0	115	114	1	138	138	0	131	130	1	135	135	0
Non-frameshift Indel	183	174	9	173	166	7	179	167	12	187	173	14	164	154	10	203	191	12	198	181	17	184	163	21
Heterozygous	116	108	8	110	103	7	95	85	10	119	106	13	98	90	8	120	112	8	121	104	17	108	90	18
Homozygous	67	66	1	63	63	0	84	82	2	68	67	1	66	64	2	83	79	4	77	77	0	76	73	3
Splicing	2,518	2,471	47	2,481	2,431	50	2,559	2,513	46	2,623	2,571	52	2,418	2,379	39	2,662	2,615	47	2,615	2,573	42	2,720	2,630	90
Heterozygous	1,494	1,449	45	1,459	1,414	45	1,519	1,475	44	1,596	1,546	50	1,412	1,376	36	1,551	1,507	44	1,549	1,509	40	1,650	1,566	84
Homozygous	1,024	1,022	2	1,022	1,017	5	1,040	1,038	2	1,027	1,025	2	1,006	1,003	3	1,111	1,108	3	1,066	1,064	2	1,070	1,064	6
StopLoss / gain	117	110	7	123	114	9	110	104	6	115	110	5	113	103	10	122	112	10	111	103	8	127	117	10
Heterozygous	85	78	7	92	83	9	79	73	6	82	77	5	79	69	10	87	77	10	84	76	8	92	82	10
Homozygous	32	32	0	31	31	0	31	31	0	33	33	0	34	34	0	35	35	0	27	27	0	35	35	0
TOTAL	22,522	22,124	398	22,514	22,107	407	22,856	22,509	347	22,906	22,511	395	22,270	21,875	395	23,442	23,030	412	23,240	22,856	384	23,999	23,286	713

Supplementary Table 3: Panel of 169 selected genes associated with IBD

AAMP
ADAD1
AMIGO3
APEH
ARPC2
ATG16L1
BACH2
BSN
BTNL2
C11orf30
C1orf106
C1orf93
C2orf74
CAPN10
CARD9
CCL11
CCL2
CCL7
CCNY
CCR6
CD19
CD244
CDKAL1
CPEB4
CREM
CXCR1
CXCR2
DAP
DENND1B
DNMT3A
EIF3C
ERAP2
ERRFI1
ESRRA
EXOC3
FADS1
FASLG
FCGR2A
FCGR2B
FIGNL1
FUT2
GALC
GCKR
GMPPB
GNA12
GPR35
GPR65

GPX1
GPX4
GSDMB
HLA-DQA1
HLA-DQA2
HLA-DRA
HLA-DRB1
HLA-DRB5
HORMAD2
HSPA6
ICAM1
ICAM3
ICOSLG
IFNG
IKZF1
IKZF3
IL10
IL10RA
IL10RB
IL12B
IL17REL
IL18R1
IL18RAP
IL19
IL1R2
IL1RL1
IL1RL2
IL2
IL20
IL21
IL23R
IL26
IL27
IL2RA
IL3
IL7R
INPP5E
IRF1
IRF5
IRGM
ITLN1
JAK2
KIF1A
KIF21B
LACC1
LAT
LIF
LRRK2
LSP1

LST1
LTA
LTB
MLX
MMEL1
MST1
MTMR3
MUC1
MUC19
NDFIP1
NKX2-3
NOD2
ORMDL3
PARK7
PIM3
PLCH2
PLCL1
PNMT
PRDM1
PRDX5
PSMG1
PTGER4
PTPN2
PTPN22
PUS10
RASIP1
REL
RNPEPL1
RTEL1
SBNO2
SCAMP3
SDCCAG3
SEC16A
SERINC3
SH2B1
SLC11A1
SLC22A4
SLC22A5
SLC2A4RG
SMAD3
SNAPC4
SP140
STAT3
STMN3
SULT1A1
SULT1A2
TAB1
TAGAP
THADA

TNF
TNFRSF14
TNFRSF6B
TNFRSF9
TNFSF11
TNFSF15
TNFSF18
TNFSF4
TNFSF8
TNPO3
TYK2
UBA7
UBE2D1
UTS2
VAMP3
YDJC
ZBTB46
ZFP36L1
ZFP90
ZGPAT
ZMIZ1
ZNF365
ZPBP
ZPBP2

Supplementary Table 4: Characterisation of variants in eight PIBD probands across 104 known IBD genes

Gene	Chromosome	Exon	Variant	Functionally implicated in pathway	Base pair location in hg18	rs ID number in dbSNP 132	Nucleotide change	Protein change	Frequency in 1000 genome	Frequency in NHLBI ESP	Sift score	Grantham Score	Grantham	Polyphen2	Observed n/8	No. homozygote	No. heterozygote	proband 1	proband 2	proband 3	proband 4	proband 5	proband 6	proband 7	proband 8
AAMP	2	N/A	sp	Cell migration	218,838,149	rs115942379	G>A	-	0.019	0.028	-	-	-	-	2	0	2					.	.		
ARPC2	2	N/A	sp	Cell migration	218,811,825	rs10169718	A>G	-	0.428	0.493	-	-	-	-	6	1	5	.	.	.	.			.	.
ATG16L1	2	9	ns	Autophagy	233,848,107	rs2241880	A646G	T216A	0.397	0.527	0.46	58	MC	B	5	3	2	.	.	.			.		.
BACH2	6	7	ns	B-cell regulation	90,717,215	NR	C1331T	S444L	NR	NR	0	145	MR	B	1	0	1						◊		
BACH2	6	7	ns	B-cell regulation	90,717,675	rs61754114	C871G	L291V	0.010	0.030	0	32	C	PrD	1	0	1							◊	
BSN	3	5	ns	Presynaptic cytoskeletal support	49,664,214	rs34762726	G2221A	A741T	0.230	0.299	0.23	58	MC	PoD	4	0	4	.	.	.					.
BSN	3	5	ns	Presynaptic cytoskeletal support	49,665,631	rs35762866	G3638A	G1213D	0.035	0.108	0.13	94	MC	PrD	2	0	2			.	.				
BSN	3	5	ns	Presynaptic cytoskeletal support	49,667,511	NR	G5518A	E1840K	NR	NR	0.07	56	MC	PrD	1	0	1							.	
BSN	3	8	ns	Presynaptic cytoskeletal support	49,676,302	rs2005557	G11587A	A3863T	0.598	0.517	0.27	58	MC	PoD	6	1	5			.	.	.	.	.	.
BTNL2	6	6	ns	T-cell negative regulation	32,470,681	rs41521946	C1178A	P393Q	NR	0.003	0.75	76	MC	B	3	0	3		.	.		.			
BTNL2	6	6	ns	T-cell negative regulation	32,470,719	rs28362677	G1140A	M380I	0.154	0.142	0.45	10	C	B	3	0	3		.	.		.			
BTNL2	6	6	ns	T-cell negative regulation	32,470,723	rs28362678	C1136T	P379L	0.158	0.142	1.00	98	MC	B	3	0	3		.	.		.			
BTNL2	6	5	ns	T-cell negative regulation	32,471,871	rs28362679	C1001T	S334L	0.014	0.020	0	145	MR	PrD	1	0	1								◊
BTNL2	6	3	ns	T-cell negative regulation	32,478,794	rs28362680	C605T	A202V	0.163	0.084	1.00	64	MC	B	1	0	1			.					
BTNL2	6	3	ns	T-cell negative regulation	32,478,813	rs2076523	A586G	K196E	0.360	0.367	1.00	56	MC	B	3	0	3		.	.	.				
BTNL2	6	3	ns	T-cell negative regulation	32,478,857	rs28362681	G542A	R181Q	0.116	0.076	0.45	43	C	B	1	0	1			.					
BTNL2	6	2	ns	T-cell negative regulation	32,480,841	rs28362682	T280A	W94R	0.132	0.075	1.00	101	MR	B	1	0	1			.					
C11orf30	11	N/A	sp	Transcriptional repression	75,904,830	rs2508740	G>A	-	0.616	0.609	-	-	-	-	8	2	6	.	.	.	.	.	.	.	.
C1orf106	1	9	ns	Unknown	199,147,492	rs45547233	A1248C	R416S	0.053	0.115	0.28	110	MR	B	1	0	1		.						
C1orf106	1	9	ns	Unknown	199,147,601	rs296520	C1357T	R453C	0.672	0.705	0.03	180	R	B	3	3	0	◊						◊	◊
C1orf106	1	N/A	sp	Unknown	199,144,473	rs41299637	T>G	-	0.158	0.272	-	-	-	-	5	1	4		.	.	.	.	.		
C1orf93	1	5	ns	Prostaglandin processing	2,509,900	NR	G526A	G176R	NR	NR	0	125	MR	PrD	1	0	1						◊		
CARD9	9	2	ns	Innate immune recognition	138,386,317	rs4077515	G35A	S12N	0.312	0.420	0.42	46	C	B	6	2	4	.	.	.			.	.	.
CD19	16	3	ns	B-cell receptor signalling	28,851,897	rs2904880	C520G	L174V	0.843	0.685	0.17	32	C	B	8	8	0	.	.	.	.	.	.	.	.
CD19	16	13	ns	B-cell receptor signalling	28,857,552	rs34763945	G1544A	R515H	0.025	0.066	0.56	29	C	PrD	2	0	2	.		.					
CDKAL1	6	8	ns	Methylthiotransferase family	20,889,397	NR	G560A	R187K	NR	NR	0.40	26	C	B	1	0	1					.			
CDKAL1	6	12	ns	Methylthiotransferase family	21,173,428	rs77152992	C1226T	P409L	0.074	0.044	0	98	MC	B	1	0	1							◊	
CPEB4	5	N/A	sp	Endoplasmic reticulum stress	173,309,112	rs12652907	A>G	-	0.186	0.071	-	-	-	-	2	0	2				.	.			
CXCR1	2	2	ns	Chemokine receptor	218,737,177	rs16858808	C1003T	R335C	0.018	0.030	0.09	180	R	PrD	2	0	2					.	.		
CXCR1	2	2	ns	Chemokine receptor	218,737,353	rs2234671	G827C	S276T	0.106	0.054	1.00	58	MC	B	1	0	1			.					
CXCR1	2	2	ns	Chemokine receptor	218,738,088	rs16858811	T92G	M31R	0.040	0.032	0.60	91	MC	B	2	0	2					.	.		
DENND1B	1	N/A	sp	Antigen presentation	195,842,931	-	->T	-	NR‡	NR	-	-	-	-	7	7	0	.	.	.	.	.	.		.
DNMT3A	2	N/A	sp	DNA methylation	25,323,417	rs2276599	C>T	-	0.628	0.747	-	-	-	-	7	4	3	.	.		.	.	.	.	.
EIF3C	16	N/A	sp	RNA transport	28,641,978	rs444518	G>T	-	NR	0.133	-	-	-	-	3	0	3	.		.	.				
ERAP2	5	6	ns	Antigen presentation	96,253,828	rs75263594	C1040T	T347M	0.013	0.033	0.01	81	MC	PrD	1	0	1			◊					
ERAP2	5	7	ns	Antigen presentation	96,256,756	rs2549782	G1176T	K392N	0.536	0.522	1.00	94	MC	B	8	5	3	.	.	.	.	.	.	.	.
ERAP2	5	N/A	sp	Antigen presentation	96,261,652	rs2248374	A>G	-	0.523	0.522	-	-	-	-	8	5	3	.	.	.	.	.	.	.	.

ERRFI1	1	4	ns	Epithelial barrier function	7,996,921	rs34781518	G325A	D109N	0.005	0.015	0.14	23	C	B	1	0	1								
FCGR2A	1	3	sg	Phagocytosis	159,742,828	rs9427397	C184T	Q62X	0.059	0.131	1.00	-	-	-	2	0	2	.					.		
FCGR2A	1	3	ns	Phagocytosis	159,742,829	rs9427398	A185G	Q62R	0.073	0.131	0.08	43	C	B	2	0	2	.					.		
FCGR2A	1	4	ns	Phagocytosis	159,746,369	rs1801274	A497G	H166R	0.430	0.512	0.04	29	C	B	7	2	5		◊	◊	◊	◊	◊	◊	◊
FCGR2B	1	4	ns	Phagocytosis	159,910,422	rs1050501	T674C	I225T	0.090	0.097	0.67	89	MC	B	1	0	1						.		
FUT2	19	2	sg	Blood group antigen synthesis	53,898,486	rs601338	G461A	W154X	0.331	0.490	0.10	-	-	-	5	2	3	.			.	.	.	.	
FUT2	19	2	ns	Blood group antigen synthesis	53,898,797	rs602662	G772A	G258S	NR	0.515	0.06	56	MC	PrD	5	2	3	.			.	.	.	.	
GALC	14	14	ns	Lysosome formation	87,477,641	rs398607	T1616C	I539T	0.450	0.485	0.28	89	MC	B	4	1	3	.	.			.			.
GALC	14	6	ns	Lysosome formation	87,512,465	rs34362748	G673A	D225N	0.072	0.149	0.10	23	C	PrD	5	0	5		.	.	.		.	.	
GALC	14	N/A	sp	Lysosome formation	87,477,487	rs448805	G>C	-	0.929	0.999	-	-	-	-	1	1	0								.
GALC	14	10	ns	Lysosome formation	87,499,443	rs17687109	T1199C	L400P	0.074	NR	-	98	MC	B	5	0	5		.	.	.		.	.	
GALC	14	N/A	sp	Lysosome formation	87,477,482	-	->A	-	NR‡	NR	-	-	-	-	2	0	2	.	.						
GALC	14	N/A	sp	Lysosome formation	87,477,482	-	G>-	-	NR‡	NR	-	-	-	-	1	0	1					.			
GMPPB	3	5	ns	Catalyses mannose processing	49,735,043	rs1466685	A551G	Q184R	0.988	1.000	0.01	43	C	B	8	8	0	◊	◊	◊	◊	◊	◊	◊	◊
GMPPB	3	5	ns	Catalyses mannose processing	49,735,146	NR	G448C	E150Q	NR	NR	0.01	29	C	B	1	0	1					◊			
GPR35	2	2	ns	Receptor for kynurenic acid	241,218,365	rs3749171	C323T	T108M	0.149	0.170	0.01	81	MC	PoD	2	1	1			◊		◊			
GPR35	2	2	ns	Receptor for kynurenic acid	241,218,922	rs3749172	A880C	S294R	0.464	0.432	0.23	110	MR	B	6	3	3	.	.		.		.	.	.
GPX1	3	2	ns	Oxidative stress	49,369,838	rs1050450	C599T	P200L	0.217	0.295	0.14	98	MC	B	4	1	3	.	.	.					.
GSDMB	17	8	ns	Unknown	35,315,722	rs2305480	C892T	P298S	0.304	0.444	0.2	74	MC	B	7	2	5		.	.	.	.	.	.	.
GSDMB	17	8	ns	Unknown	35,315,743	rs2305479	G871A	G291R	0.324	0.481	0	125	MR	PrD	8	3	5	◊	◊	◊	◊	◊	◊	◊	◊
GSDMB	17	7	ns	Unknown	35,316,029	rs35104165	A710G	D237G	0.012	0.036	0	94	MC	B	1	1	0	◊							
GSDMB	17	N/A	sp	Unknown	35,317,995	rs11078928	T>C	-	0.308	0.435	-	-	-	-	7	2	5		.	.	.	.	.	.	.
HLA-DQA1	6	1	ns	Major histocompatibility complex	32,713,235	rs1047989	C22A	L8M	0.427	0.549	0.48	15	C	B	5	5	0			.	.	.	.	.	
HLA-DQA1	6	1	ns	Major histocompatibility complex	32,713,244	rs1047992	G31A	A11T	0.054	0.065	0.46	58	MC	B	2	2	0					.		.	
HLA-DQA1	6	1	ns	Major histocompatibility complex	32,713,262	rs12722039	G49A	V17M	0.048	0.039	0.58	21	C	B	1	0	1								.
HLA-DQA1	6	2	ns	Major histocompatibility complex	32,717,083	rs1129740	G101A	C34Y	0.480	0.460	1.00	194	R	B	5	5	0			.	.	.	.	.	
HLA-DQA1	6	2	ns	Major histocompatibility complex	32,717,104	rs1071630	T122C	F41S	0.480	0.436	1.00	155	R	B	5	5	0			.	.	.	.	.	
HLA-DQA1	6	2	ns	Major histocompatibility complex	32,717,125	rs12722051	A143T	Y48F	0.170	0.166	0.89	22	C	B	1	0	1				.				
HLA-DQA1	6	2	ns	Major histocompatibility complex	32,717,151	rs10093	C169G	Q57E	0.195	0.287	0.21	29	C	B	3	0	3		.		.				.
HLA-DQA1	6	2	ns	Major histocompatibility complex	32,717,170	rs1142323	A188G	E63G	0.283	0.245	0.33	98	MC	B	5	4	1			.	.	.	.	.	
HLA-DQA1	6	2	ns	Major histocompatibility complex	32,717,185	rs1142324	C203T	A68V	0.310	0.369	1.00	64	MC	B	5	5	0			.	.	.	.	.	
HLA-DQA1	6	2	ns	Major histocompatibility complex	32,717,190	rs1142326	C208T	R70W	0.022	NR	0	101	MR	PrD	2	2	0			◊				◊	
HLA-DQA1	6	2	ns	Major histocompatibility complex	32,717,194	rs1142328	G212T	W71L	0.036	0.161	1.00	61	MC	B	2	2	0			.				.	
HLA-DQA1	6	3	ns	Major histocompatibility complex	32,717,784	rs707952	C389T	T130I	0.206	0.251	0.60	89	MC	B	5	4	1			.	.	.	.	.	
HLA-DQA1	6	3	ns	Major histocompatibility complex	32,717,851	rs707950	G456C	Q152H	0.450	0.453	1.00	24	C	B	5	5	0			.	.	.	.	.	
HLA-DQA1	6	3	ns	Major histocompatibility complex	32,717,930	rs707949	T535C	F179L	0.096	0.142	0.03	22	C	B	5	4	1			◊	◊	◊	◊	◊	
HLA-DQA1	6	3	ns	Major histocompatibility complex	32,717,947	rs707963	T552G	D184E	0.084	0.159	0.87	45	C	B	5	4	1			.	.	.	.	.	
HLA-DQA1	6	3	ns	Major histocompatibility complex	32,717,952	rs707962	T557G	I186S	0.084	0.143	0.08	142	MR	B	5	4	1			.	.	.	.	.	
HLA-DQA1	6	3	ns	Major histocompatibility complex	32,717,987	rs1129957	C592A	Q198K	NR	NR	0.53	53	MC	B	4	4	0			.		.	.	.	
HLA-DQA1	6	4	ns	Major histocompatibility complex	32,718,415	rs35087390	G664A	A222T	0.012	0.015	0.02	58	MC	B	1	0	1								◊
HLA-DQA1	6	4	ns	Major histocompatibility complex	32,718,439	rs9260	A688G	M230V	0.682	0.743	1.00	21	C	B	7	5	2		.	.	.	.	.	.	.
HLA-DQA1	6	4	ns	Major histocompatibility complex	32,718,465	rs1048430	C714G	F238L	0.210	0.197	1.00	22	C	B	1	0	1				.				
HLA-DQA1	6	4	ns	Major histocompatibility complex	32,718,473	rs9272793	A722G	Q241R	0.357	0.443	0.60	43	C	B	5	5	0			.	.	.	.	.	
HLA-DQA1	6	N/A	sp	Major histocompatibility complex	32,713,167	rs9272426	A>G	-	0.256	0.351	-	-	-	-	4	3	1			.	.		.	.	
HLA-DQA1	6	N/A	sp	Major histocompatibility complex	32,713,303	rs9272434	C>T	-	0.145	0.232	-	-	-	-	5	5	0			.	.	.	.	.	
HLA-DQA1	6	N/A	sp	Major histocompatibility complex	32,717,724	rs9272744	C>T	-	0.276	0.312	-	-	-	-	1	0	1				.				
HLA-DQA1	6	N/A	sp	Major histocompatibility complex	32,718,534	rs1130117	G>T	-	0.180	0.240	-	-	-	-	4	4	0			.		.	.	.	
HLA-DQA1	6	N/A	sp	Major histocompatibility complex	32,718,356	rs9272783	C>T	-	NR	0.035	-	-	-	-	5	4	1			.	.	.	.	.	
HLA-DQA1	6	N/A	sp	Major histocompatibility complex	32,718,357	rs9272784	T>A	-	NR	0.043	-	-	-	-	5	4	1			.	.	.	.	.	

HLA-DQA2	6	3	ns	Major histocompatibility complex	32,821,576	-	T362C	F121S	0.034	0.001	1.00	155	R	B	1	0	1						.		
HLA-DQA2	6	3	ns	Major histocompatibility complex	32,821,603	rs116163401	C389T	T130I	0.060	0.002	0.60	89	MC	B	3	0	3			.		.	.		
HLA-DQA2	6	3	ns	Major histocompatibility complex	32,821,749	rs34847266	T535C	F179L	0.023	NR	0.03	22	C	PoD	2	0	2					◊		◊	
HLA-DQA2	6	4	ns	Major histocompatibility complex	32,822,061	rs9276436	T680C	V227A	0.089	0.078	0.01	64	MC	B	1	0	1				◊				
HLA-DQA2	6	4	ns	Major histocompatibility complex	32,822,121	rs2071800	G740A	G247D	0.054	0.072	0.04	94	MC	B	1	0	1								◊
HLA-DQA2	6	N/A	sp	Major histocompatibility complex	32,817,287	rs2051600	A>G	-	0.803	0.768	-	-	-	-	7	5	2		.	.	.	.	.	.	.
HLA-DQA2	6	N/A	sp	Major histocompatibility complex	32,821,166	rs2213565	C>T	-	0.811	0.769	-	-	-	-	7	5	2		.	.	.	.	.	.	.
HLA-DQA2	6	N/A	sp	Major histocompatibility complex	32,821,832	rs4398729	G>A	-	0.030	0.062	-	-	-	-	1	0	1						.		
HLA-DQA2	6	N/A	sp	Major histocompatibility complex	32,822,163	rs74201397	A>-	-	NR	NR	-	-	-	-	5	1	4		.	.		.		.	.
HLA-DRA	6	4	ns	Major histocompatibility complex	32,519,624	rs7192	T724G	L242V	0.613	0.614	0.22	32	C	B	5	2	3		.		.	.		.	.
HLA-DRA	6	N/A	sp	Major histocompatibility complex	32,518,193	rs3129885	T>C	-	0.151	0.174	-	-	-	-	3	1	2			.	.		.		
HLA-DRA	6	N/A	sp	Major histocompatibility complex	32,519,501	rs2239804	T>C	-	0.510	0.471	-	-	-	-	4	1	3				.	.		.	.
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,058	rs3175105	A176G	Y59C	0.010	0.007	0.03	194	R	B	2	0	2		◊			◊			
HLA-DRB1	6	1	ns	Major histocompatibility complex	32,665,464	rs116331390	A34T	M12L	0.011	0.008	0.01	15	C	B	6	0	6			◊	◊	◊	◊	◊	◊
HLA-DRB1	6	4	ns	Major histocompatibility complex	32,656,534	rs3830125	G730A	A244T	0.013	NR	0.03	58	MC	PrD	2	0	2			◊				◊	
HLA-DRB1	6	N/A	sp	Major histocompatibility complex	32,656,492	rs3830128	G>A	-	0.246	0.193	-	-	-	-	3	0	3					.		.	.
HLA-DRB1	6	N/A	sp	Major histocompatibility complex	32,656,498	rs3830127	C>T	-	0.207	0.184	-	-	-	-	3	0	3					.		.	.
HLA-DRB1	6	3	ns	Major histocompatibility complex	32,657,430	rs77637983	G534C	Q178H	0.247	0.303	0.25	24	C	B	5	0	5			.		.	.	.	.
HLA-DRB1	6	3	ns	Major histocompatibility complex	32,657,453	rs2308768	A511G	M171V	0.770	0.777	1.00	21	C	B	7	5	2		.	.	.	.	.	.	.
HLA-DRB1	6	3	ns	Major histocompatibility complex	32,657,459	rs112408735	G505A	A169T	0.290	0.231	0.24	58	MC	PoD	6	0	6			.	.	.	.	.	.
HLA-DRB1	6	3	ns	Major histocompatibility complex	32,657,479	rs78466762	T485G	L162R	0.814	0.749	1.00	102	MR	B	6	4	2		.	.		.	.	.	.
HLA-DRB1	6	3	ns	Major histocompatibility complex	32,657,567	rs1136795	T397G	S133A	0.123	0.151	1.00	99	MC	B	6	0	6			.	.	.	.	.	.
HLA-DRB1	6	N/A	sp	Major histocompatibility complex	32,659,856	rs9269939	A>G	-	0.223	0.329	-	-	-	-	2	2	0						.	.	
HLA-DRB1	6	N/A	sp	Major histocompatibility complex	32,659,857	rs9269940	T>C	-	0.189	0.134	-	-	-	-	2	2	0						.	.	
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,906	rs29029549	C328T	H110Y	0.261	0.176	0.08	83	MC	PoD	1	0	1				.				
HLA-DRB1	6	4	ns	Major histocompatibility complex	32,656,512	rs71547382	G752A	R251K	0.019	0.000	0.13	26	C	B	1	0	1			.					
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,917	rs9269941	C317A	T106N	0.121	0.195	0.18	65	MC	B	3	2	1			.	.		.		
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,995	rs1059582	C239G	T80R	0.020	0.002	0.01	71	MC	PrD	1	0	1							◊	
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,034	rs17878951	T200C	V67A	0.020	0.052	0.13	64	MC	B	1	0	1							.	
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,056	rs1059346	T178A	F60I	0.020	0.074	0.51	21	C	B	2	0	2		.			.			
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,069	rs1059572	C165G	F55L	0.020	0.029	0.42	22	C	B	2	0	2		.			.			
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,935	rs9269942	C299A	A100E	0.248	0.501	0.13	107	MR	B	3	3	0			.		.	.		
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,936	rs1064592	G298A	A100T	0.182	0.365	0.1	58	MC	B	4	3	1		.	.	.		.		
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,948	rs17886918	A286C	I96L	0.317	0.385	1.00	5	C	B	5	2	3		.	.	.		.	.	
HLA-DRB1	6	1	ns	Major histocompatibility complex	32,665,467	-	T31A	C11S	0.023	0.012	0.13	112	MR	B	5	0	5			.	.		.	.	.
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,007	rs17884945	T227A	F76Y	0.276	0.376	1.00	22	C	B	7	1	6		.	.	.	.	.	.	.
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,026	rs56158521	G208A	D70N	0.070	0.184	0.02	23	C	PrD	1	0	1				◊				
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,037	rs17883134	C197A	S66Y	0.622	0.687	1.00	144	MR	B	3	1	2					.		.	.
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,038	rs16822820	T196A	S66T	0.160	0.207	0.08	58	MC	B	3	0	3			.	.		.		
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,053	rs1064664	T181C	Y61H	0.340	0.278	0.21	83	MC	B	4	1	3			.			.	.	.
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,059	rs113465897	T175C	Y59H	0.102	0.012	0.27	83	MC	B	1	0	1							.	
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,977	rs17885129	A257T	D86V	0.032	0.030	0.17	152	R	B	1	0	1				.				
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,070	rs16822516	T164A	F55Y	0.153	0.242	0.37	22	C	B	3	2	1			.	.		.		
HLA-DRB1	6	N/A	sp	Major histocompatibility complex	32,665,392	-	A>T	-	0.127	NR	-	-	-	-	4	0	4			.	.		.	.	
HLA-DRB1	6	1	ns	Major histocompatibility complex	32,665,413	rs9270299	T85G	S29A	0.731	0.819	1.00	99	MC	B	7	6	1		.	.	.	.	.	.	.
HLA-DRB1	6	1	ns	Major histocompatibility complex	32,665,457	-	C41T	A14V	0.134	0.086	0.29	64	MC	B	5	0	5			.		.	.	.	.
HLA-DRB1	6	1	ns	Major histocompatibility complex	32,665,461	rs9270303	A37G	T13A	0.789	0.733	1.00	58	MC	B	6	4	2			.	.	.	.	.	.
HLA-DRB1	6	1	ns	Major histocompatibility complex	32,665,484	rs707953	A14G	K5R	0.376	0.383	0.84	60	MC	B	6	0	6			.	.	.	.	.	.
HLA-DRB1	6	1	ns	Major histocompatibility complex	32,665,424	-	C74G	P25R	0.043	0.019	0.02	103	MR	B	4	0	4			◊	◊		◊	◊	

HLA-DRB1	6	2	sg	Major histocompatibility complex	32,659,913	-	C321A	Y107X	NR‡	0.129	-	-	-	-	1	0	1				.				
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,926	rs17886882	C308G	A103G	NR	NR	0.09	60	MC	B	2	2	0			.			.		
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,927	-	G307C	A103P	NR‡	NR	0.06	27	C	B	2	2	0			.			.		
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,929	rs1059596	C305G	A102G	NR	NR	0.1	60	MC	B	2	2	0			.			.		
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,935	rs9269942	C299G	A100G	NR	NR	0.13	60	MC	B	1	0	1		.						
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,937	rs1064591	G297C	Q99H	NR	NR	0	24	C	B	1	1	0					◊			
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,938	rs17884070	A296G	Q99R	NR	NR	0.34	43	C	B	1	0	1				.				
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,939	rs17881965	C295G	Q99E	NR	NR	0.09	29	C	B	1	1	0					.			
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,659,974	rs1059584	C260A	A87D	NR	0.002	0	126	MR	PrD	1	0	1							◊	
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,035	rs17878614	G199C	V67L	NR	NR	0.22	32	C	B	3	0	3			.	.		.		
HLA-DRB1	6	2	ns	Major histocompatibility complex	32,660,063	rs1059575	C171A	D57E	NR	NR	0.71	45	C	B	2	0	2		.			.			
HLA-DRB1	6	N/A	sp	Major histocompatibility complex	32,665,397	-	C>G	-	NR‡	0.217	-	-	-	-	1	0	1				.				
HLA-DRB1	6	1	ns	Major histocompatibility complex	32,665,414	-	G84C	L28F	NR‡	NR	0.04	22	C	B	4	0	4			◊	◊		◊	◊	
HLA-DRB5	6	4	ns	Major histocompatibility complex	32,594,380	rs41553512	G694A	V232I	0.013	0.050	0.24	29	C	B	1	1	0								.
HLA-DRB5	6	3	ns	Major histocompatibility complex	32,595,143	-	C634G	P212A	0.056	0.094	0.01	27	C	B	3	2	1			◊	◊		◊		
HLA-DRB5	6	3	ns	Major histocompatibility complex	32,595,148	rs1136633	C629T	T210M	0.127	0.143	0.15	81	MC	B	1	0	1				.				
HLA-DRB5	6	3	ns	Major histocompatibility complex	32,595,152	-	G625A	V209M	0.120	NR	0.08	21	C	B	1	0	1				.				
HLA-DRB5	6	3	ns	Major histocompatibility complex	32,595,220	rs1059662	C557T	T186I	0.008	0.027	0.19	89	MC	B	1	0	1								.
HLA-DRB5	6	3	ns	Major histocompatibility complex	32,595,287	rs3200405	A490G	S164G	0.530	0.701	1.00	56	MC	B	7	5	2		.	.	.	.	.	.	.
HLA-DRB5	6	3	ns	Major histocompatibility complex	32,595,331	rs114293611	A446G	N149S	0.199	0.242	0.93	46	C	B	3	3	0			.		.	.		
HLA-DRB5	6	2	ns	Major histocompatibility complex	32,597,686	rs41556512	G344T	G115V	NR	NR	0.95	109	MR	B	1	0	1							.	
HLA-DRB5	6	2	ns	Major histocompatibility complex	32,597,731	-	G299A	R100K	0.016	0.026	0.78	26	C	B	2	1	1			.		.			
HLA-DRB5	6	2	ns	Major histocompatibility complex	32,597,732	rs41551116	A298G	R100G	NR	NR	0.01	125	MR	B	1	0	1					◊			
HLA-DRB5	6	2	ns	Major histocompatibility complex	32,597,744	rs41562819	T286A	F96I	NR	NR	0.71	21	C	B	1	0	1					.			
HLA-DRB5	6	2	ns	Major histocompatibility complex	32,597,770	rs41562816	C260A	A87D	NR	0.004	0	126	MR	PoD	2	1	1					◊		◊	
HLA-DRB5	6	2	ns	Major histocompatibility complex	32,597,803	-	A227T	Y76F	0.017	0.109	0.57	22	C	B	2	1	1					.		.	
HLA-DRB5	6	2	ns	Major histocompatibility complex	32,597,831	rs78961241	T199G	L67V	0.601	0.655	1.00	32	C	B	1	1	0								.
HLA-DRB5	6	2	ns	Major histocompatibility complex	32,597,854	rs41546317	A176G	D59G	0.270	0.095	0.03	94	MC	B	1	1	0								◊
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,892	rs76748970	G88A	G30R	0.120	0.207	1.00	125	MR	B	1	0	1				.				
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,895	rs71549220	G85T	A29S	0.031	0.081	0.01	99	MC	B	2	0	2		◊			◊			
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,896	rs72508462	G84C	L28F	0.128	0.102	0.04	22	C	B	2	0	2						◊	◊	
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,921	rs17211043	T59C	M20T	0.251	0.336	0.16	81	MC	B	3	0	3		.		.	.			
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,938	-	G42T	K14N	NR‡	0.000	0	94	MC	B	3	0	3			◊			◊	◊	
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,939	rs78935256	A41C	K14T	0.614	0.670	0.04	78	MC	B	2	1	1		◊		◊				
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,939	-	A41T	K14M	NR‡	NR	0	95	MC	B	1	1	0								◊
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,940	rs77365746	A40G	K14E	0.635	0.738	0	56	MC	B	7	6	1		◊	◊	◊	◊	◊	◊	◊
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,948	-	A32G	Y11C	0.296	0.255	1.00	194	R	B	4	3	1			.	.		.	.	
HLA-DRB5	6	1	ns	Major histocompatibility complex	32,605,966	-	A14G	K5R	0.130	0.087	0.83	26	C	B	3	0	3			.			.	.	
HLA-DRB5	6	N/A	sp	Major histocompatibility complex	32,597,650	rs115833135	C>A	-	0.010	0.002	-	-	-	-	1	1	0						.		
HLA-DRB5	6	N/A	sp	Major histocompatibility complex	32,594,302	-	G>A	-	0.028	0.067	-	-	-	-	1	1	0				.				
HLA-DRB5	6	N/A	sp	Major histocompatibility complex	32,597,653	-	T>C	-	0.050	0.022	-	-	-	-	3	0	3			.		.		.	
HLA-DRB5	6	N/A	sp	Major histocompatibility complex	32,605,879	rs79192142	C>T	-	0.331	0.344	-	-	-	-	1	0	1				.				
HLA-DRB5	6	N/A	sp	Major histocompatibility complex	32,594,310	-	C>G	-	NR‡	0.063	-	-	-	-	1	1	0				.				
HLA-DRB5	6	N/A	sp	Major histocompatibility complex	32,595,122	rs113524741	C>A	-	NR	NR	-	-	-	-	1	0	1				.				
HLA-DRB5	6	N/A	sp	Major histocompatibility complex	32,597,652	-	A>G	-	NR‡	0.015	-	-	-	-	3	0	3			.		.		.	
HLA-DRB5	6	N/A	sp	Major histocompatibility complex	32,605,874	rs76562035	A>T	-	NR	NR	-	-	-	-	2	0	2						.	.	
HORMAD2	22	2	ns	Unknown	28,819,945	rs34150968	G4A	A2T	0.004	0.012	0	58	MC	PoD	1	0	1					◊			
HSPA6	1	1	ns	Oxidative stress	159,761,664	rs1079109	C592T	L198F	0.120	0.117	0.01	22	C	PrD	1	0	1	◊							
ICAM1	19	5	ns	Leukocyte adhesion ligand	10,256,141	-	G988A	V330M	0.001	0.003	0	21	C	PrD	1	0	1								◊

ICAM1	19	5	ns	Leukocyte adhesion ligand	10,256,252	-	C1099T	R367C	0.004	0.000	0	180	R	PrD	1	0	1								◊
ICAM1	19	6	ns	Leukocyte adhesion ligand	10,256,683	rs5498	A1405G	K469E	0.322	0.432	1.00	56	MC	B	5	1	4	.	.	.			.		.
ICAM3	19	7	ns	Leukocyte adhesion ligand	10,305,603	rs2230399	G1574C	S525T	0.094	0.083	0.17	58	MC	PoD	1	0	1								.
ICAM3	19	3	ns	Leukocyte adhesion ligand	10,307,568	rs2304237	A428G	D143G	0.148	0.221	1.00	94	MC	B	2	0	2	.		.					
ICAM3	19	N/A	sp	Leukocyte adhesion ligand	10,305,826	rs2278442	G>A	-	0.648	0.641	-	-	-	-	7	2	5	.	.	.	.		.	.	.
ICOSLG	21	3	ns	T-cell regulation	44,481,202	rs11558819	G382A	V128I	0.222	0.277	0.23	29	C	PoD	3	0	3	.				.			.
IL10	1	2	ns	Innate immune recognition	205,011,338	-	C211A	L71M	NR	NR	0.07	15	C	PrD	1	0	1						.		
IL10	1	1	ns	Innate immune recognition	205,012,361	-	G43A	G15R	NR	0.002	0.04	125	MR	PoD	1	0	1							◊	
IL10RA	11	4	ns	Immune cell recruitment	117,369,273	rs3135932	A475G	S159G	0.072	0.168	0.38	56	MC	B	2	0	2		.	.					
IL10RA	11	5	ns	Immune cell recruitment	117,370,056	rs2228055	A670G	I224V	0.104	0.054	1.00	29	C	B	1	0	1		.						
IL10RA	11	7	ns	Immune cell recruitment	117,374,880	rs2229113	A1051G	R351G	0.814	0.685	0.2	125	MR	B	7	3	4	.	.	.	.	.		.	.
IL10RB	21	2	ns	Immune cell recruitment	33,562,658	rs2834167	A139G	K47E	0.276	0.259	0.2	56	MC	B	4	0	4	.		.			.		.
IL17REL	22	14	ns	Unknown	48,777,607	rs5771069	T998C	L333P	0.558	0.500	0	98	MC	B	2	1	1			◊			◊		
IL17REL	22	5	ns	Unknown	48,781,321	rs9617090	G208A	G70R	0.276	0.412	0	125	MR	B	2	0	2					◊		◊	
IL18R1	2	N/A	sp	IL18 mediated signal transduction	102,350,711	rs1420098	T>C	-	0.337	0.382	-	-	-	-	4	2	2	.	.				.		.
IL18RAP	2	11	ns	Enhances IL18 binding	102,433,811	-	C1282A †	L428M	NR	0.001	0.15	15	C	PrD	1	0	1	.							
IL18RAP	2	11	ns	Enhances IL18 binding	102,433,812	-	T1283A †	L428Q	NR	0.001	0.15	113	MR	PrD	1	0	1	.							
IL18RAP	2	N/A	sp	Enhances IL18 binding	102,429,952	rs11465723	G>A	-	0.074	0.159	-	-	-	-	2	0	2			.		.			
IL19	1	7	ns	Uncertain	205,082,580	rs2243191	T524C	F175S	0.638	0.780	0.31	155	R	B	4	4	0			.	.	.		.	
IL1R2	2	N/A	sp	Immune tolerance	102,008,997	-	->T	-	NR‡	NR	-	-	-	-	3	0	3		.		.		.		
IL1RL1	2	3	ns	T-helper cell function	102,321,900	rs1041973	C233A	A78E	0.327	0.257	0.81	107	MR	B	5	0	5		.	.		.	.		.
IL1RL1	2	11	ns	T-helper cell function	102,334,439	rs4988956	G1297A	A433T	0.415	0.388	0.13	58	MC	PoD	4	2	2			.	.	.		.	
IL1RL1	2	11	ns	T-helper cell function	102,334,643	rs10192036	C1501A †	Q501K	NR	0.019	1.00	53	MC	B	4	2	2			.	.	.		.	
IL1RL1	2	11	ns	T-helper cell function	102,334,644	rs10204137	A1502G†	Q501R	0.018	0.032	0.61	43	C	B	4	2	2			.	.	.		.	
IL1RL1	2	11	ns	T-helper cell function	102,334,788	rs10192157	C1646T	T549I	0.416	0.386	0.04	89	MC	B	4	2	2			◊	◊	◊		◊	
IL1RL1	2	11	ns	T-helper cell function	102,334,794	rs10206753	T1652C	L551S	0.416	0.385	0.60	145	MR	PrD	4	2	2			.	.	.		.	
IL1RL1	2	N/A	sp	T-helper cell function	102,323,723	rs13029918	A>G	-	0.016	0.026	-	-	-	-	1	0	1						.		
IL1RL1	2	N/A	sp	T-helper cell function	102,326,078	rs62152661	A>G	-	0.050	0.094	-	-	-	-	1	0	1					.			
IL1RL2	2	11	ns	Interleukin receptor	102,217,903	-	C1412T	A471V	NR	NR	0	64	MC	PrD	1	0	1							◊	
IL1RL2	2	11	ns	Interleukin receptor	102,218,140	rs2302612	T1649C	L550P	0.292	0.190	0	98	MC	B	2	0	2		◊				◊		
IL23R	1	2	ns	Th17-cell differentiation	67,406,400	rs1884444	G9T	Q3H	0.499	0.530	0.01	24	C	B	7	2	5	◊	◊	◊	◊	◊	◊		◊
IL23R	1	7	ns	Th17-cell differentiation	67,457,975	rs7530511	T929C	L310P	0.854	0.871	0.65	98	MC	B	8	8	0	.	.	.	.	.	.	.	.
IL26	12	N/A	sp	Mucosal immunity	66,881,986	rs10748100	T>C	-	0.220	0.085	-	-	-	-	2	0	2		.		.				
IL27	16	4	ns	IL-10 signalling	28,420,904	rs181206	T356C	L119P	0.149	0.317	0.09	98	MC	PoD	1	1	0	.							
IL2RA	10	N/A	sp	T cell regulation	6,106,206	rs11256369	C>G	-	0.172	0.222	-	-	-	-	3	1	2				.	.	.		
IL3	5	1	ns	Haematopoietic growth factor	131,424,377	rs40401	C79T	P27S	0.444	0.229	0.83	74	MC	B	1	0	1					.			
IL7R	5	2	ns	B and T-cell regulation	35,896,825	rs1494558	T197C	I66T	0.586	0.672	1.00	89	MC	PoD	8	5	3	.	.	.	.	.	.	.	.
IL7R	5	4	ns	B and T-cell regulation	35,906,947	rs1494555	G412A	V138I	0.707	0.679	0.55	29	C	B	8	5	3	.	.	.	.	.	.	.	.
IL7R	5	6	ns	B and T-cell regulation	35,910,332	rs6897932	C731T	T244I	0.162	0.263	0.37	89	MC	B	5	0	5	.	.	.		.	.		
IL7R	5	8	ns	B and T-cell regulation	35,912,031	rs3194051	A1066G	I356V	0.230	0.266	0.79	29	C	B	4	0	4		.			.	.	.	
INPP5E	9	N/A	sp	Converts phosphatidylinositol	138,447,420	rs73566945	G>A	-	0.134	0.247	-	-	-	-	1	0	1						.		
IRF1	5	N/A	sp	MHC class I gene regulator	131,849,971	rs2070724	A>G	-	0.341	0.315	-	-	-	-	1	1	0								.
IRF5	7	6	ns	B-cell regulation	128,374,610	rs113806178	G572A	R191Q	0.172	0.206	0.58	43	C	U	1	0	1				.				
ITLN1	1	4	ns	Epithelial barrier function	159,118,450	rs2274907	T326A	V109D	0.583	0.685	0.74	152	R	B	8	3	5	.	.	.	.	.	.	.	.
ITLN1	1	N/A	sp	Epithelial barrier function	159,117,560	rs2236515	T>C	-	0.656	0.690	-	-	-	-	8	3	5	.	.	.	.	.	.	.	.
JAK2	9	9	ns	Th17-cell differentiation	5,055,003	rs2230723	C1177G	L393V	0.016	0.006	0.38	32	C	B	1	0	1		.						
KIF21B	1	33	ns	Microtubule-binding protein	199,210,518	-	C4722A	D1574E	NR	NR	0.1	45	C	PrD	1	0	1				.				
LAT	16	N/A	sp	T-cell, NK cell, mast cell signalling	28,905,612	rs4788115	T>A	-	0.132	0.169	-	-	-	-	4	0	4		.			.		.	.
LRRK2	12	18	ns	Autophagy	38,958,256	rs10878307	A2167G	I723V	0.046	0.070	0.52	29	C	B	2	1	1	.					.		

LRRK2	12	34	ns	Autophagy	39,000,168	rs11564148	T4939A	S1647T	0.253	0.299	0.80	58	MC	B	3	1	2				.		.	.	
LRRK2	12	N/A	sp	Autophagy	38,931,524	rs7955902	C>A	-	0.286	0.378	-	-	-	-	3	1	2				.		.	.	
LRRK2	12	N/A	sp	Autophagy	38,967,411	-	->T	-	NR‡	NR	-	-	-	-	6	3	3			.	.	.	.	.	.
LRRK2	12	N/A	sp	Autophagy	39,003,220	rs41286460	A>G	-	NR	0.004	-	-	-	-	1	0	1		.						
LRRK2	12	N/A	sp	Autophagy	39,039,321	-	->T	-	NR	NR	-	-	-	-	5	2	3	.	.	.	.				.
LRRK2	12	N/A	sp	Autophagy	39,043,461	-	T>-	-	NR	NR	-	-	-	-	4	4	0			.	.	.		.	
LSP1	11	N/A	sp	Cell migration	1,861,732	-	T>-	-	NR	NR	-	-	-	-	1	1	0							.	
LTA	6	2	ns	Cytokine receptor interaction	31,648,535	rs2229094	T37C	C13R	0.249	0.272	0.47	180	R	B	4	1	3	.		.		.		.	
LTA	6	3	ns	Cytokine receptor interaction	31,648,736	rs2229092	A152C	H51P	0.039	0.072	0.3	77	MC	B	2	0	2					.		.	
LTA	6	3	ns	Cytokine receptor interaction	31,648,763	rs1041981	C179A	T60N	0.377	0.328	0.49	65	MC	PoD	5	1	4		.	.	.	.	.		
MLX	17	7	ns	Transcription regulator	37,975,555	rs665268	A506G	Q169R	0.276	0.275	0.14	43	C	PoD	6	1	5	.	.	.	.	.		.	
MMEL1	1	16	ns	Metalloprotease	2,516,606	rs3748816	T1553C	M518T	0.470	0.332	0.99	81	MC	B	4	1	3	.	.	.		.			
MMEL1	1	N/A	sp	Metalloprotease	2,526,932	rs4074787	C>T	-	0.022	0.045	-	-	-	-	1	0	1				.				
MMEL1	1	N/A	sp	Metalloprotease	2,517,993	rs2843401	T>C	-	0.560	0.683	-	-	-	-	7	3	4	.	.		.	.	.	.	.
MST1	3	18	ns	Apoptosis	49,696,536	rs3197999	C2107T	R703C	0.203	0.294	0.30	180	R	PoD	4	1	3	.	.	.					.
MST1	3	17	sg	Apoptosis	49,696,816	-	C1951T	R651X	0.004	0.013	0.14	-	-	-	1	0	1				.				
MST1	3	13	ns	Apoptosis	49,697,765	rs62262682	G1478T	R493L	0.015	0.058	0.09	102	MR	B	1	0	1					.			
MST1	3	1	ns	Apoptosis	49,701,074	rs62262686	C55T	P19S	0.111	0.138	0.45	74	MC	PrD	4	0	4			.	.	.			.
MST1	3	N/A	sp	Apoptosis	49,701,032	rs62262685	T>C	-	0.150	0.231	-	-	-	-	6	0	6	.	.	.	.	.			.
MTMR3	22	17	ns	Lipid phosphatase	28,745,983	rs61737780	C2335T	L779F	0.005	0.012	0.21	22	C	PoD	1	0	1					.			
MTMR3	22	17	ns	Lipid phosphatase	28,746,527	rs41278853	A2879G	N960S	0.041	0.086	0.08	46	C	B	1	0	1				.				
MTMR3	22	N/A	sp	Lipid phosphatase	28,704,920	rs737907	C>T	-	0.094	0.092	-	-	-	-	1	0	1							.	
NOD2	16	4	ns	Autophagy	49,302,125	rs2066842	C802T	P268S	0.122	0.271	0.26	74	MC	B	5	1	4	.		.		.	.	.	
NOD2	16	4	ns	Autophagy	49,303,427	rs2066844	C2104T	R702W	0.029	0.047	0	101	MR	PrD	2	0	2	◊		◊					
NOD2	16	9	ns	Autophagy	49,314,777	rs5743291	G2863A	V955I	0.044	0.095	0.46	29	C	B	1	0	1				.				
NOD2	16	11	fi	Autophagy	49,321,282	-	3019_3020insC	L1007fs	NR‡	NR	-	-	-	-	1	0	1			.					
PARK7	1	5	ns	Autophagy	7,953,581	rs71653619	G293A	R98Q	0.003	0.012	0.50	43	C	B	2	0	2				.		.		
PIM3	22	6	ns	T-cell regulation	48,742,697	rs4077129	T899C	V300A	0.751	0.753	0.51	64	MC	B	5	4	1	.	.		.		.		.
PLCL1	2	2	ns	Intracellular signalling	198,658,485	rs1064213	G1999A	V667I	0.344	0.491	0	29	C	PrD	5	4	1		◊	◊	◊	◊		◊	
PNMT	17	3	sg	Adrenaline processing	35,080,063	-	C744A	Y248X	NR	0.088	0.18	-	-	-	1	0	1						.		
PRDM1	6	4	ns	B-cell activation/T-cell regulation	106,654,065	rs811925	C609G	D203E	0.130	0.167	0.95	45	C	PrD	2	0	2			.	.				
PRDX5	11	1	ns	Oxidative stress	63,842,361	rs7938623	A98G	Y33C	0.949	0.999	0	194	R	B	6	6	0	◊	◊	◊	◊		◊	◊	
PSMG1	21	N/A	sp	Proteasome assembly chaperone	39,469,257	rs9305670	A>G	-	0.861	NR	-	-	-	-	7	7	0	.	.	.	.	.	.	.	
PTGER4	5	3	ns	Epithelial barrier function	40,727,650	rs111866313	G880A	V294I	0.009	0.027	0.51	29	C	B	1	0	1		.						
PTPN22	1	12	ns	B cell activation	114,179,091	rs2476601	T1693C	W565R	0.963	0.909	1.00	101	MR	B	8	6	2	.	.	.	.	.	.	.	.
RTEL1	20	24	ns	DNA repair	61,791,572	rs35640778	G2051A	R684Q	0.003	0.018	0.58	43	C	B	1	0	1							.	
RTEL1	20	32	ns	DNA repair	61,796,554	rs3208008	A3126C	Q1042H	0.743	0.773	0.25	24	C	B	7	4	3	.	.		.	.	.	.	.
SBNO2	19	N/A	sp	Immune tolerance	1,060,213	rs2159133	A>G	-	0.473	0.349	-	-	-	-	1	1	0					.			
SBNO2	19	N/A	sp	Immune tolerance	1,067,935	rs7251039	G>A	-	0.557	0.505	-	-	-	-	2	2	0		.				.		
SBNO2	19	N/A	sp	Immune tolerance	1,075,031	rs2024092	G>A	-	0.224	0.205	-	-	-	-	1	0	1						.		
SDCCAG3	9	9	ns	Modulation of TNF response	138,418,401	rs1131992	G1135A	V379M	0.084	0.133	0.15	21	C	B	2	0	2		.				.		
SDCCAG3	9	7	ns	Modulation of TNF response	138,419,458	rs3812577	G911A	R304Q	0.082	0.124	0.18	43	C	PrD	2	0	2		.				.		
SDCCAG3	9	N/A	sp	Modulation of TNF response	138,418,474	rs12235378	G>A	-	0.040	0.029	-	-	-	-	2	0	2		.	.					
SEC16A	9	N/A	sp	Endoplasmic reticulum traffic	138,477,880	rs11145753	G>A	-	0.108	0.137	-	-	-	-	1	0	1		.						
SEC16A	9	3	ns	Endoplasmic reticulum traffic	138,488,774	rs3812594	C3115T	R1039C	0.141	0.260	0.06	180	R		4	0	4	.		.			.	.	
SEC16A	9	23	ns	Endoplasmic reticulum traffic	138,465,668	rs45519739	C6173T	T2058M	NR	0.015	0.01	81	MC		1	0	1			◊					
SEC16A	9	3	ns	Endoplasmic reticulum traffic	138,490,409	-	G1480C	G494R	NR	NR	0	125	MR		1	0	1						◊		
SEC16A	9	3	ns	Endoplasmic reticulum traffic	138,490,852	-	G1037A	R346H	NR	0.001	0.13	29	C		1	0	1		.						

SEC16A	9	3	ns	Endoplasmic reticulum traffic	138,490,870	-	G1019A	G340E	NR	0.002	0.07	98	MC		1	0	1	.								
SH2B1	16	1	ns	Adaptor for TYK receptors	28,785,470	-	T554A	L185Q	NR	NR	0.04	113	MR	PrD	1	0	1									◊
SH2B1	16	5	ns	Adaptor for TYK receptors	28,790,742	rs7498665	A1450G	T484A	0.230	0.383	0.17	58	MC	B	3	1	2	.		.	.					
SH2B1	16	5	ns	Adaptor for TYK receptors	28,790,787	-	A1495G	T499A	NR	NR	0.19	58	MC	PoD	1	0	1		.							
SLC11A1	2	15	ns	Macrophage membrane transport	218,967,976	rs17235409	G1627A	D543N	0.063	0.019	0.02	23	C	B	1	0	1									◊
SLC22A4	5	5	ns	Oxidative stress	131,690,961	rs272893	T917C	I306T	0.568	0.616	0.09	89	MC	B	8	4	4	.	.	.	.	.	.	.	.	.
SLC22A4	5	9	ns	Oxidative stress	131,704,219	rs1050152	C1507T	L503F	0.161	NR	0.1	22	C	B	7	3	4	.	.	.	.	.	.	.		
SLC22A5	5	N/A	sp	Solute transport	131,747,898	rs4551059	A>G	-	1.000	NR	-	-	-	-	8	8	0	.	.	.	.	.	.	.	.	.
SLC2A4RG	20	6	ns	CHO metabolism	61,844,151	rs8957	G699T	E233D	0.727	0.723	0.63	45	C	B	7	5	2	.	.	.		.	.	.	.	.
SMAD3	15	3	ns	TGF-β signalling pathway	65,244,752	rs35874463	A376G	I126V	0.018	0.053	0.65	29	C	B	2	0	2	.	.							
SNAPC4	9	22	ns	RNA polymerase transcription	138,390,697	rs3812561	C4342T	P1448S	0.055	0.031	0	74	MC	B	2	0	2		◊	◊						
SNAPC4	9	17	ns	RNA polymerase transcription	138,396,228	rs34569521	G2186A	R729Q	0.046	0.088	0	43	C	PrD	1	0	1							◊		
SNAPC4	9	10	ns	RNA polymerase transcription	138,402,740	-	G1100C	G367A	NR‡	0.003	0	60	MC	PrD	1	0	1							◊		
SP140	2	16	ns	Nuclear body protein	230,857,352	rs4972946	G1546A	E516K	0.954	0.919	1.00	56	MC	B	8	7	1	.	.	.	.	.	.	.	.	.
SP140	2	24	ns	Nuclear body protein	230,883,793	rs62192163	T2266C	C756R	0.048	0.245	0.44	180	R	B	1	1	0	.								
SP140	2	N/A	sp	Nuclear body protein	230,814,454	rs75177776	T>C	-	0.002	0.005	-	-	-	-	1	0	1	.								
SP140	2	N/A	sp	Nuclear body protein	230,865,756	rs6710297	A>G	-	0.129	0.266	-	-	-	-	4	1	3	.	.		.					.
STAT3	17	N/A	sp	IL-10 signalling	37,729,179	-	->A	-	NR‡	NR	-	-	-	-	2	0	2							.		.
SULT1A1	16	7	ns	Sulphate conjugation	28,524,986	rs1801030	G667A	V223M	0.915	0.998	0.07	21	C	B	8	8	0	.	.	.	.	.	.	.	.	.
SULT1A1	16	7	ns	Sulphate conjugation	28,525,015	rs1042028	G638A	R213H	0.129	0.256	0.01	29	C	B	2	0	2	◊		◊						
SULT1A2	16	8	ns	Sulphate conjugation	28,510,894	rs27742	A844G	K282E	0.977	0.999	0.57	56	MC	B	8	8	0	.	.	.	.	.	.	.	.	.
SULT1A2	16	7	ns	Sulphate conjugation	28,511,156	rs1059491	A704C	N235T	0.204	0.363	0.02	65	MC	PrD	2	0	2	◊		◊						
SULT1A2	16	2	ns	Sulphate conjugation	28,514,697	rs10797300	C56T	P19L	0.062	0.144	0.39	98	MC	PrD	5	0	5		.		.	.	.			.
SULT1A2	16	2	ns	Sulphate conjugation	28,514,733	rs1136703	T20C	I7T	0.014	0.059	0.12	89	MC	B	1	0	1	.								
TAGAP	6	6	ns	T cell regulation	159,382,412	rs41267765	G439A	E147K	0.014	0.020	0.64	56	MC	B	1	0	1	.								
TAGAP	6	5	ns	T cell regulation	159,383,130	-	G283A	G95S	NR	NR	0.58	56	MC	B	1	0	1	.								
THADA	2	N/A	sp	Apoptosis	43,579,469	rs35419456	C>A	-	0.026	0.032	-	-	-	-	1	0	1		.							
THADA	2	29	ns	Apoptosis	43,478,688	rs33979934	A4153T	T1385S	0.194	0.245	0.24	58	MC	B	4	0	4				.		.	.	.
THADA	2	24	ns	Apoptosis	43,586,327	rs7578597	A3559G	T1187A	0.137	0.107	0.35	58	MC	PoD	1	0	1			.						
THADA	2	14	ns	Apoptosis	43,651,123	rs17031056	G2095A	V699I	0.201	0.201	-	29	C	B	3	1	2		.			.	.			
THADA	2	28	nd	Apoptosis	43,508,785 *	-	c.4014_4016del	p.1338_1339del	NR	NR	-	-	-	-	1	0	1	.								
THADA	2	N/A	sp	Apoptosis	43,671,460	-	T>-	-	NR‡	NR	-	-	-	-	2	1	1		.			.				
TNFRSF14	1	1	ns	Cytokine receptor interaction	2,486,265	-	A50G	K17R	0.540	0.470	0.17	26	C	PrD	1	1	0			.						
TYK2	19	20	ns	Th17-cell differentiation	10,325,843	rs35018800	C2783T	A928V	0.003	0.008	0	64	MC	PrD	1	0	1				◊					
TYK2	19	8	ns	Th17-cell differentiation	10,336,649	rs2304255	G1087A	G363S	0.035	0.080	0.54	56	MC	B	1	0	1									.
TYK2	19	8	ns	Th17-cell differentiation	10,336,652	rs2304256	G1084T	V362F	0.273	0.282	0.07	50	C	B	4	1	3				.	.	.	.		
TYK2	19	N/A	sp	Th17-cell differentiation	10,333,933	rs280519	A>G	-	0.509	0.508	-	-	-	-	5	3	2				.	.	.	.	.	
TYK2	19	N/A	sp	Th17-cell differentiation	10,334,138	rs280520	A>G	-	0.305	0.220	-	-	-	-	1	0	1									.
UBA7	3	N/A	sp	Activates ubiquitin	49,822,627	rs1799845	A>G	-	0.209	0.299	-	-	-	-	3	0	3					.	.	.		
UBA7	3	N/A	sp	Activates ubiquitin	49,823,418	rs28535523	C>T	-	0.140	0.179	-	-	-	-	4	0	4	.				.	.	.		
UTS2	1	1	ns	Oxidative stress	7,835,616	rs34305100	T35C	I12T	0.078	0.175	0	89	MC	PoD	1	0	1				◊					
UTS2	1	1	ns	Oxidative stress	7,836,017	rs228648	C62T	T21M	0.491	0.570	0	81	MC	PoD	6	2	4	◊	◊		◊	◊		◊	◊	
UTS2	1	1	ns	Oxidative stress	7,836,032	rs13306061	G47A	R16Q	0.088	0.175	0	43	C	B	1	0	1				◊					
YDJC	22	5	ns	Unknown	20,312,892	rs2298428	G787A	A263T	0.160	0.185	0.35	58	MC	B	1	0	1		.							
ZBTB46	20	2	ns	Zinc finger	61,892,524	rs2281929	A31G	T11A	0.210	0.118	0.18	58	MC	B	2	0	2				.		.			
ZGPAT	20	2	ns	Zinc finger	61,810,559	rs1291212	C183G	S61R	0.963	0.923	0.15	110	MR	B	8	8	0	.	.	.	.	.	.	.	.	.
ZNF365	10	5	ns	Zinc finger	63,829,339	rs3758490	G1009T	A337S	0.455	0.612	0.23	99	MC	B	7	2	5	.	.	.	.	.	.			.
ZNF365	10	3	ns	Zinc finger	64,084,667	-	C97A	L33I	0.006	0.000	0	5	C	U	1	0	1	◊								

ZNF365	10	4	ns	Zinc finger	64,085,190	rs7076156	A184G	T62A	0.865	0.732	0.80	58	MC	U	8	5	3	.	.	.	.	.	.	.	.
ZPBP	7	N/A	sp	Zona pellucida binding protein	49,993,668	rs988392	C>T	-	0.829	0.797	-	-	-	-	7	6	1	.	.	.		.	.	.	.
ZPBP2	17	5	ns	Zona pellucida binding protein	35,282,160	rs11557467	G518T	S173I	0.399	0.488	0.12	142	MR	B	8	3	5	.	.	.	.	.	.	.	.

Novel variants are shown in grey.

N/A = not applicable, NR = not reported, NR‡ indicates variants that despite not being reported in dbSNP132 or 1000 genomes, are reported in dbSNP129 or seen in our in-house control exomes and are therefore not characterised as novel.

* Indicates the first bp location of a 3-bp deletion.

Where a specific variant is present in a proband, this is indicated by a dot (**.**)

Where a specific variant is present in a proband and has a SIFT score of < 0.05, this is indicated by **◊**

†indicates a dinucleotide variant (that for IL18RAP results in a codon change from CTG > AAG, resulting in p.L428K amino acid change).

ns=nonsynonymous; sg=stopgain; sp=splicing; fi=frameshift insertion; nd=nonframeshift deletion.

C=conservative; MC=moderately conservative; MR=moderately radical; R=radical.

B=benign; PoD=possibly damaging; PrD=probably damaging; U=unknown

HLA gene variants should be considered with caution due to known challenges of accurate alignment of short read data and consequent difficulty in robust identification of variants from highly divergent HLA haplotypes.

Supplementary Table 5: Chi-squared contingency testing for excess of rare variants in IBD candidate genes in cases compared to controls

	Synonymous	Non-synonymous and non-frameshift
Cases n=8	61	77
Controls n=22	149	208

Pearson χ^2 (1 degree freedom) = 0.25, p = 0.62