

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

Supplementary Tables S1-5

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 stenotic 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 aminosaliclylate 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

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