Supplemental Digital Content

Supplementary Methods

Exome sequencing analysis

Genomic DNA was extracted from peripheral venous blood or saliva using the salting out method[1]. Fragmented DNA was then subjected to adaptor ligation, exome library enrichment performed using Agilent SureSelect Human All Exon capture kit (V5 & V6) and sequenced on Illumina HiSeq platforms. Alignment against the human genome (hg38) was performed using Burrows-Wheeler Aligner (BWA)[2], variants called using Genome Analysis Toolkit (GATK v3.6) and ANNOVAR for variant annotation[3]. *NOD2* variants were reported in line with previously published data. Briefly, variants with a CADD score of >15 and a minor allele frequency of <0.01/novel, or variants reported as pathogenic in the CLINVAR database or human genetic mutation database were reported[4]. Variants were categorised in line with the American College of Medical Genetics (ACGM) guidance to remove 'benign' variants and identify 'pathogenic' and 'likely pathogenic' variants[5].

Statistical analysis and data visualisation

Statistical analysis was performed using the GraphPad Prism software, version 7. Cytokine induction between the patient cohort and controls were compared using unpaired t-tests in a two-tailed manner. For application of hierarchical clustering and generation of radar plots, raw cytokine data were normalised using RobustScaler and StandardScaler respectively, embedded in the python scikit-learn package (version 0.19.01). Normalisation of raw data is a common requirement for machine learning applications as these programmes are designed on the assumption that the data values vary on comparable scales. Presence of frequent

outliers can affect the objective and predictive performance of many machine learning algorithms. Statistical graphics including hierarchical dendogram and radar plots were generated using python scikit-learn and R software package.

For generation of radar plots, raw cytokine data for every patient were normalised using the standardScaler function from python scikit-learn package, which removes the mean and scales the data to unit variance. Each value was subtracted from the mean (μ) and divided by the standard deviation (σ). Both μ and σ were calculated on paediatric controls.

$X^1 = X - \mu_{controls}/\sigma_{controls}$

Normalised values per patient per stimulus were plotted along the spokes or vectors of the radar plot, each spoke representing a specific cytokine response. Cytokine data from paediatric controls were used to define the boundaries of normality in the patient cohort. The upper and lower range of normality of cytokine responses were assigned as +/- 2 SD across the mean of controls. Values within 2 SD represented normal responses and, those above and below 2 SD indicated hyper-inflammatory and hypo-inflammatory responses respectively.

For hierarchical clustering, the raw cytokine data were normalised with RobustScaler and the R software package used to generate a dendogram. Data transformation and scaling statistics in RobustScaler are based on percentiles and therefore not influenced by a few number of large marginal outliers. RobustScaler centres the data using the median calculated on controls and then scales the values according to inter quartile ranges.

References:

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

2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25(14):1754-60.

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

4. Ashton JJ, Andreoletti G, Coelho T, et al. Identification of Variants in Genes Associated with Single-gene Inflammatory Bowel Disease by Whole-exome Sequencing. *Inflamm Bowel Dis* 2016;22(10):2317-27.

5. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17(5):405-24.

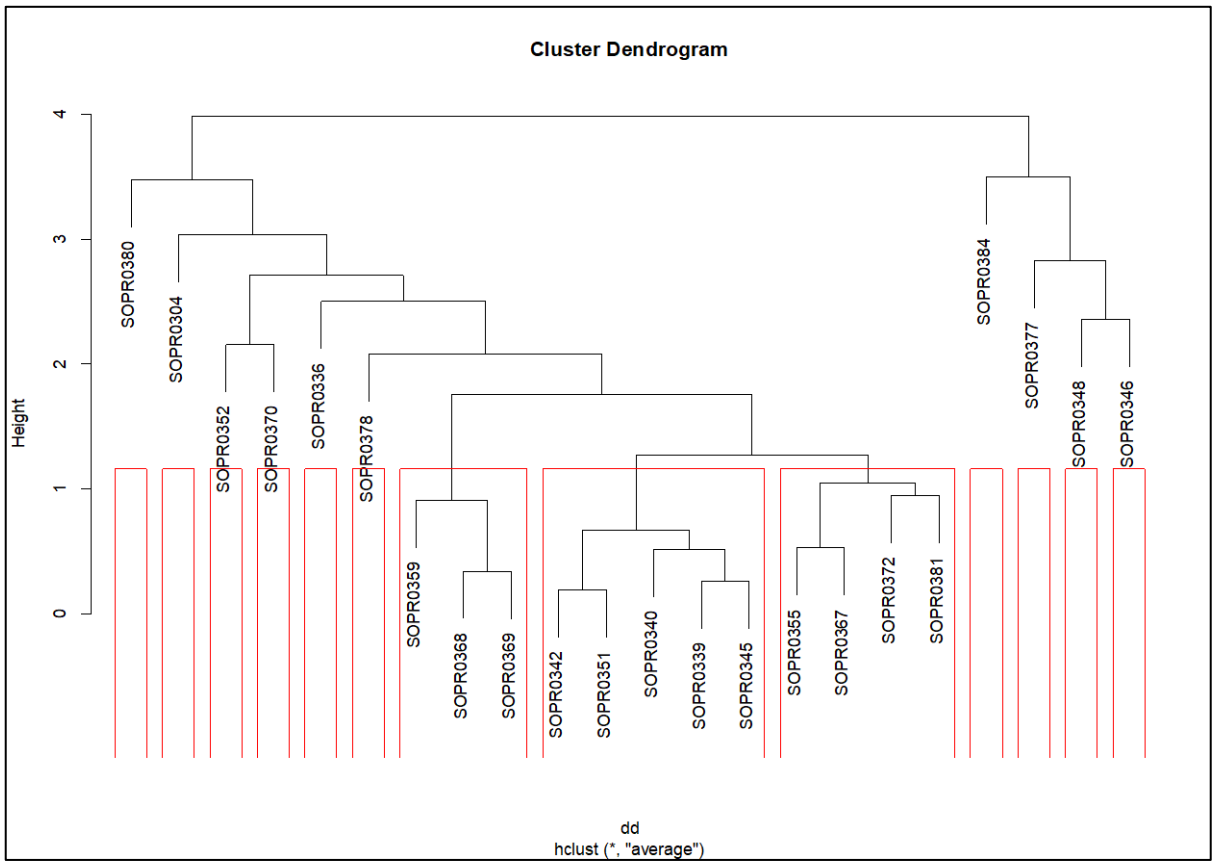


Figure S1. Static height cut-off for cluster identification

The figure demonstrates three distinct clusters generated using a static height cut-off at 1.25. Clusters 1, 2 and 3 had five, four and three patients respectively. Ten individuals fell in the 'unclustered' group at the applied height cut-off.

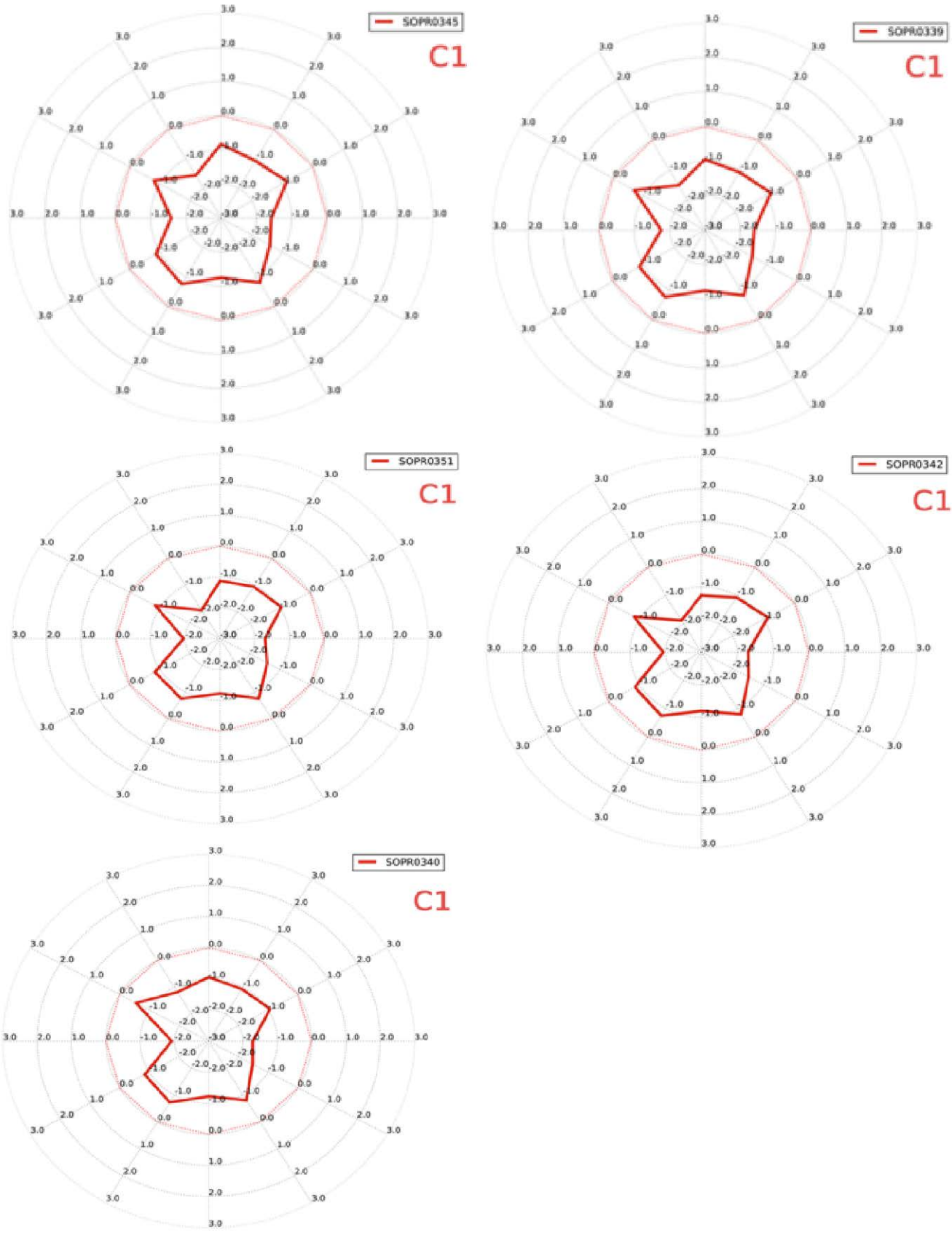


Figure S1-A. Radar plots in cluster 1 (5 individuals)

Figure S1-B. Radar plots in cluster 2

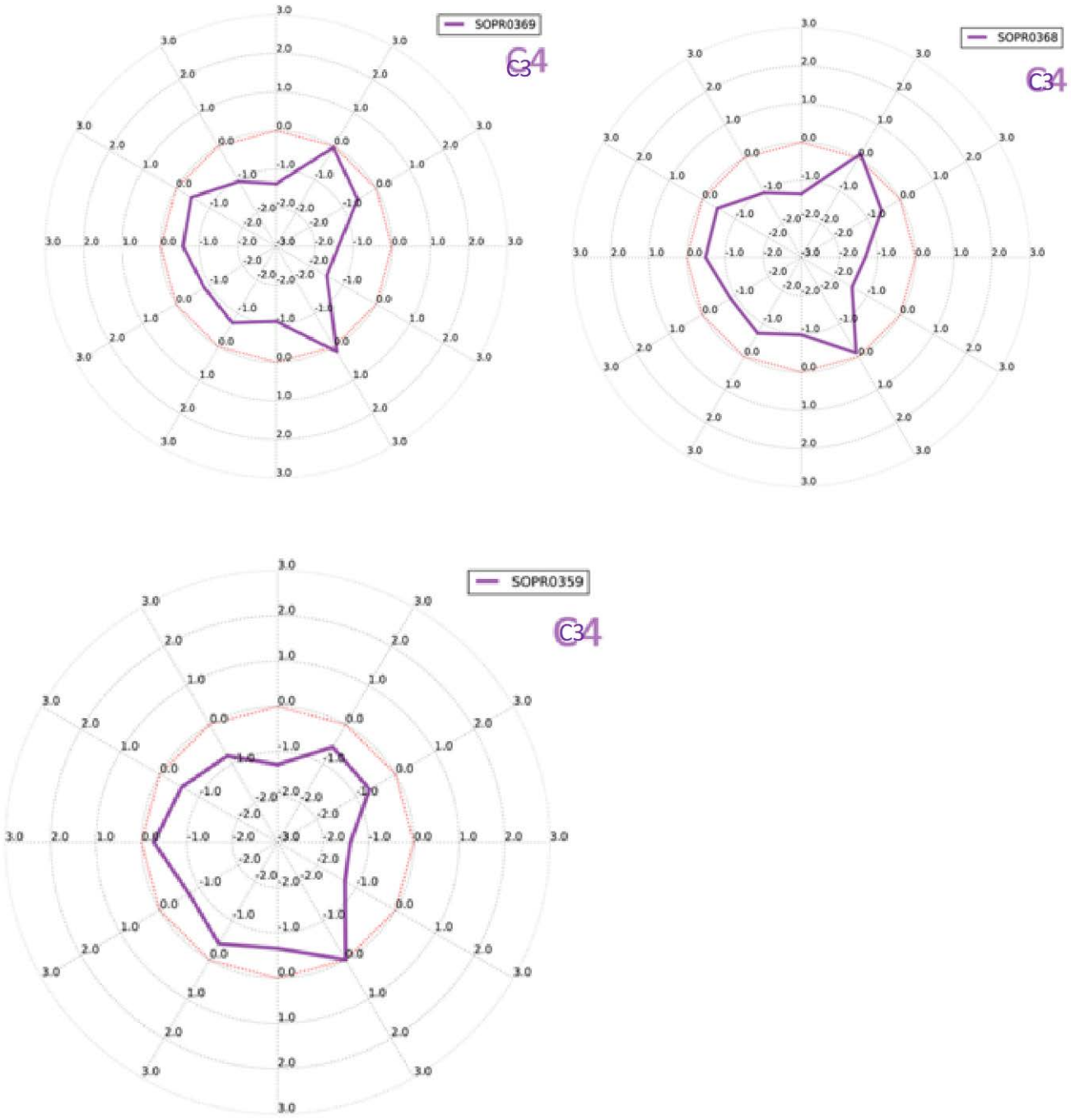


Figure S1-C. Radar plots in cluster 3

Figure S2 (A-C). Individual Radar plots

This figure shows immune response profiles for the patients in clusters 1-3 on individual radar plots. Patients within a cluster are put together and indicated by a unique colour. Each radar plot has 3 sectors representing a stimulant with the 4 cytokines analysed per sector. The red-dotted line is the mean obtained from cytokine responses observed in the control group.

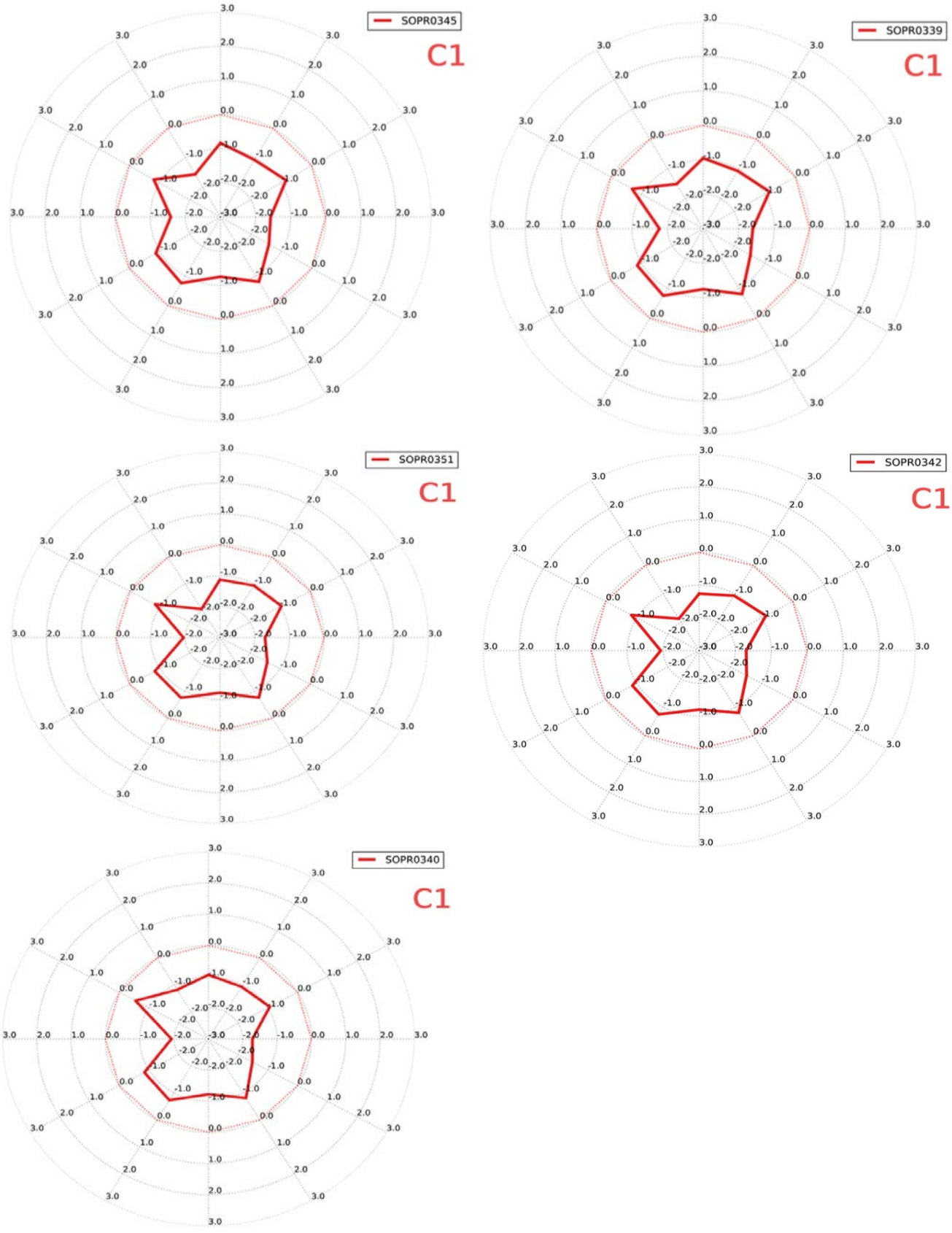
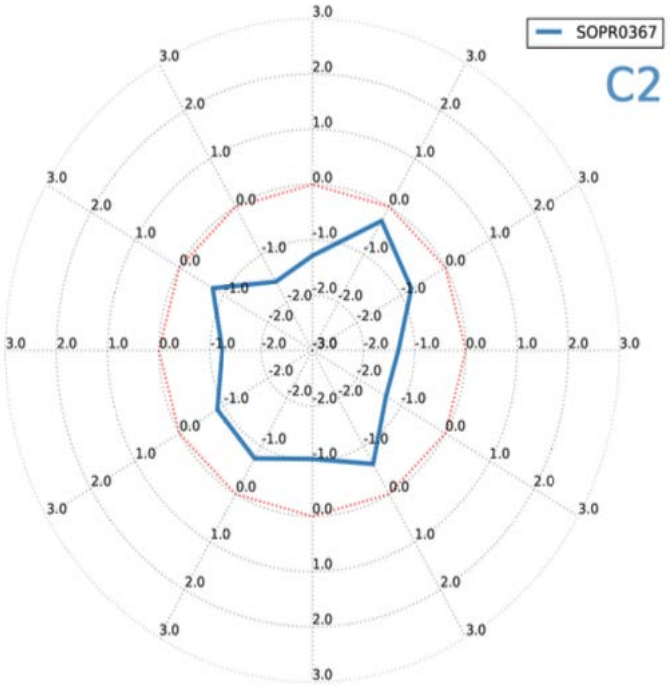
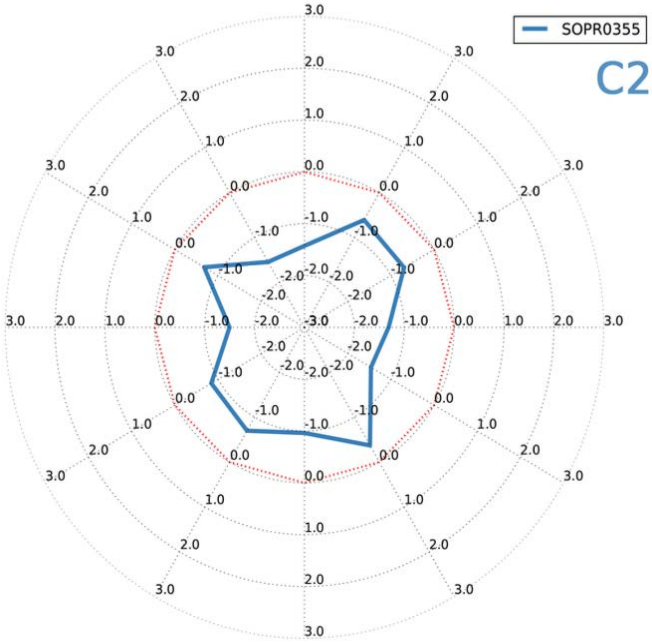
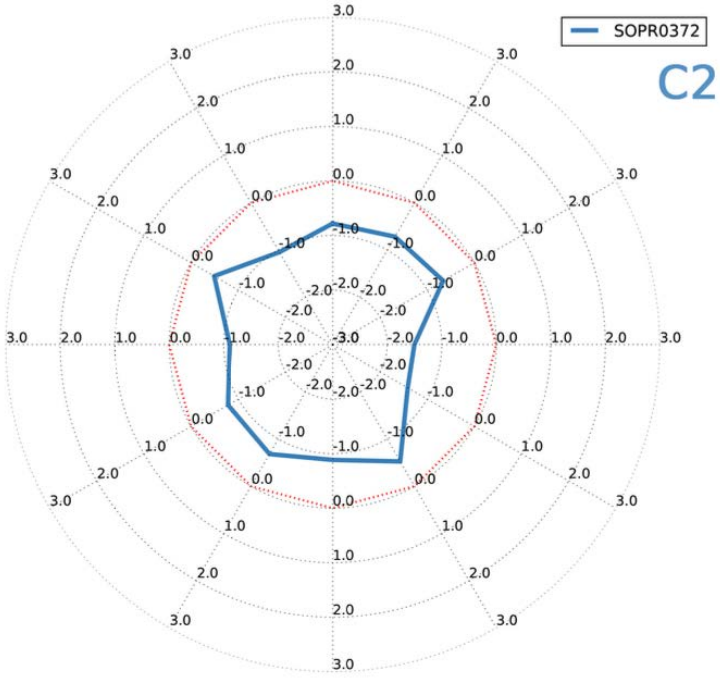


Figure S1-A. Radar plots in cluster 1 (5 individuals)







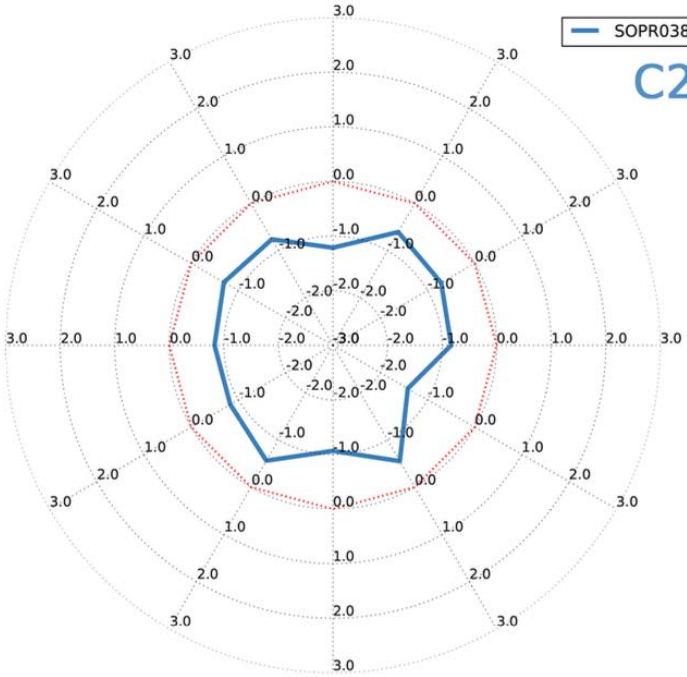
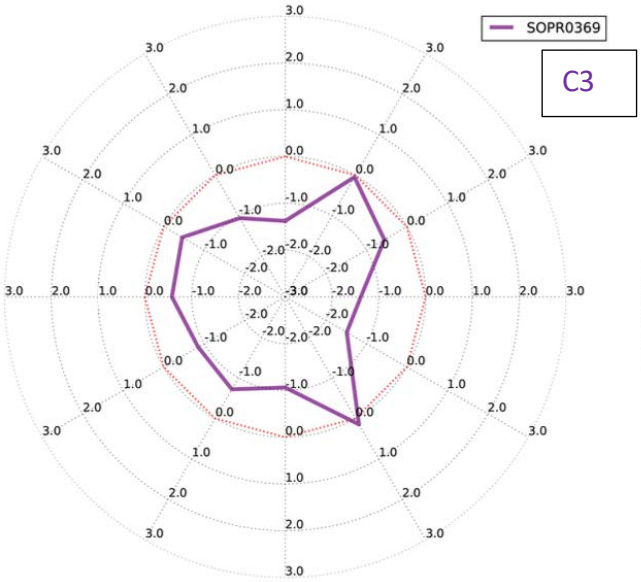
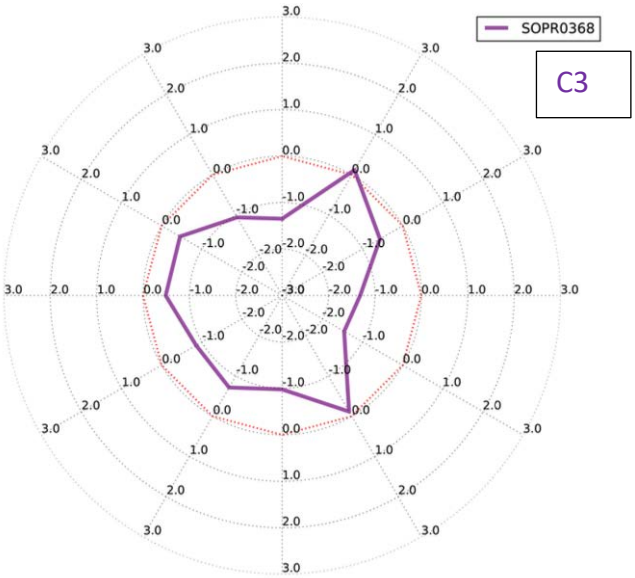


Figure S1-B. Radar plots in cluster 2





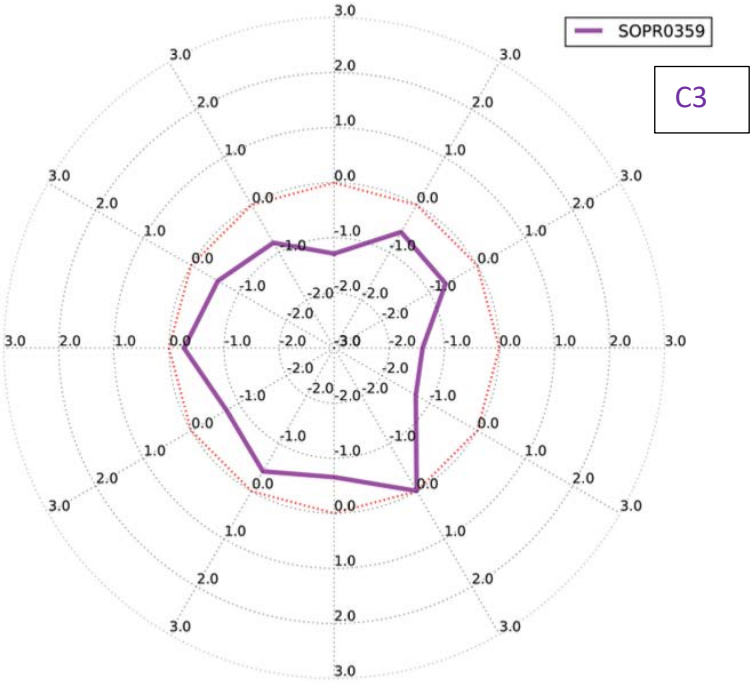


Figure S1-C. Radar plots in cluster 3

Table S1. Patient cohort characteristics

	Proband	Gender	Age at diagnosis	Diagnosis	PCDAI/ PUCAI	Severity at diagnosis	Paris Classification (disease location)	Abdo pain	Diarrhoea	Lethargy	Nocturnal stools	Reduced appetite	Rectal Bleeding	Weight loss	CRP	Faecal calprotectin(mg/g)
Crohn's Disease																
1	SOPR0304	M	10	CD	30	Mod-sev	L2L4a	✓	✓	✓	✓	✓	✓	×	4	>1800
2	SOPR0336	F	12	CD	45	Mod-sev	L3	✓	✓	✓	×	✓	×	×	67	NA
3	SOPR0339	M	15	CD	32.5	Mod-sev	L2L4a	✓	✓	×	✓	✓	×	✓	11	NA
4	SOPR0342	M	5	CD	55	Mod-sev	L2	✓	✓	×	✓	×	✓	×	78	NA
5	SOPR0345	M	14	CD	45	Mod-sev	L3	✓	✓	×	×	×	×	✓	62	NA
6	SOPR0348	M	8	CD	50	Mod-sev	L3	✓	✓	✓	×	✓	×	✓	65	NA
7	SOPR0351	F	7	CD	12.5	Mild	L2L4a	✓	×	×	×	×	×	×	NA	673
8	SOPR0352	F	16	CD	32.5	Mod-sev	L3	✓	×	✓	×	×	×	✓	25	NA
9	SOPR0359	F	9	CD	22.5	Mild	L2	✓	✓	✓	×	×	×	×	1	2664
10	SOPR0368	M	15	CD	47.5	Mod-sev	L4a	✓	✓	✓	×	×	×	✓	51	3024
11	SOPR0370	M	11	CD	35	Mod-sev	L3L4a	×	✓	×	×	✓	✓	×	47	NA
12	SOPR0377	F	13	CD	32.5	Mod-sev	L3	✓	✓	✓	×	✓	×	✓	12	360
13	SOPR0378	M	9	CD	30	Mod-sev	L2	✓	×	✓	✓	✓	✓	✓	2	NA
14	SOPR0380	F	16	CD	52.5	Mod-sev	L3L4a	✓	✓	×	×	✓	×	✓	87	NA
Ulcerative Colitis																
1	SOPR0340	F	10	UC	45	Mod-sev	E1	✓	✓	×	×	✓	✓	✓	NA	>3000
2	SOPR0346	M	16	UC	45	Mod-sev	E2	✓	✓	×	✓	×	✓	×	NA	NA
3	SOPR0355	F	13	UC	75	Severe	E4	✓	✓	✓	×	×	✓	✓	2	1690
4	SOPR0367	F	13	UC	60	Moderate	E4	✓	✓	×	✓	×	✓	×	7	NA
5	SOPR0369	M	15	UC	40	Moderate	E1	✓	✓	×	×	×	✓	×	1	166
6	SOPR0372	M	14	UC	55	Moderate	E2	✓	✓	✓	×	×	✓	×	2	NA
7	SOPR0381	F	13	UC	45	Moderate	E4	✓	✓	×	×	×	✓	✓	1	NA
8	SOPR0384	M	14	UC	70	Severe	E3	×	✓	×	✓	×	✓	×	2	NA

Patient characteristics are presented along with the PCDAI (mild 10-29, moderate-severe >30) and PUCAI (mild 10-34, moderate 35-64, severe >65) scores for CD and UC respectively. Disease location/s as per Paris classification for Crohn's disease include L1 (distal 1/3 of ileum), L2 (colonic), L3 (ileo-colonic) & L4 (upper GI disease; L4a- proximal to the ligament of

Treitz, L4b- distal to the ligament of Treitz); UC E1(ulcerative proctitis), E2 (left-sided UC), E3 (extensive, up to hepatic flexure) & E4 (pancolitis). NA- not available.

Table S2. Regression analysis between immune profiles and disease activity scores

Multivariable linear regression between immune responses and normalised disease scores

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.026	1.320		.020	.985
	IL-10_LPS	1.154	1.724	.661	.670	.520
	IL-1B_LPS	-2.778	2.690	-1.722	-1.033	.329
	IL-6_LPS	2.031	3.975	.796	.511	.622
	TNF-a_LPS	1.500	1.439	1.169	1.042	.324
	IL-10_MDP	-.682	1.113	-.396	-.613	.555
	IL-1B_MDP	-.238	1.006	-.100	-.237	.818
	IL-6_MDP	-.023	.941	-.018	-.025	.981
	TNF-a_MDP	-.650	2.026	-.239	-.321	.756
	IL-10_PAM	-.527	1.312	-.371	-.402	.697
	IL-1B_PAM	1.587	1.659	1.182	.956	.364
	IL-6_PAM	-1.254	2.874	-.732	-.436	.673
	TNF-a_PAM	-.409	1.496	-.199	-.274	.791

Dependent Variable: Normalised disease activity

Multi-variable linear regression was performed between the twelve cytokine responses and the normalised disease activity scores. The disease activity scores were normalised as the scoring tools are different in CD and UC. There was no evidence of a correlation between the immune profiles and the disease activity scores.

Table S3. Cytokine data before normalisation (raw data)

		LPS				MDP				Pam3CSK4			
Cytokine Read-out		IL-10	IL-1β	IL-6	TNF-α	IL-10	IL-1β	IL-6	TNF-α	IL-10	IL-1β	IL-6	TNF-α
Normal range (pg/ml)	Subjects	3433-15289	3030-5133	1797-4737	1829-5885	420-3084	195-734	1044-2627	62-278	2119-7835	482-1351	652-4440	194-537
SOPR0304	**Patient**	1189.5	1441.5	4640	858.5	606.5	381	4470.5	328	1439.5	654	4711	345.5
SOPR0336	**Patient**	2415.5	177.5	1599	263	1	1786	1786	96	222	233.5	1283	126.5
SOPR0339	**Patient**	3907	1115	1274	576.5	31.5	9.5	15	2	276	109	613.5	7.5
SOPR0340	**Patient**	3953.5	1666	1993.5	274	58.5	33	52	1.5	85	41	293	5
SOPR0342	**Patient**	1823.5	487	575	262.5	30	13.5	96	3	218	54.5	553	8
SOPR0345	**Patient**	4738.5	1022.5	679.5	931	4	3	5	2	479	128	629	5.5
SOPR0346	**Patient**	9352	3111	691.5	5127	1574.5	237.5	2339	469.5	5089.5	737	2437	753.5
SOPR0348	**Patient**	14620.5	1897	2522	2498	1789	90	2183	169	8011.5	523	2520	327.5
SOPR0351	**Patient**	2741	382	499	225.5	95	7	53	18.5	406	50	220	11
SOPR0352	**Patient**	10332	1587.5	691.5	1564.5	3815.5	182	868	100.5	2579	102	740.5	36
SOPR0355	**Patient**	688.5	1047.5	1331	1097	59	60	514	151.5	292	283.5	1173	221.5
SOPR0359	**Patient**	1644.5	2323.5	1929.5	3470.5	292.5	318	1062	277	886.5	234.5	1286	239.5
SOPR0367	**Patient**	1628	1005.5	953.5	1595.5	32.5	18.5	381	63.5	708	279.5	898.5	367.5
SOPR0368	**Patient**	1235.5	1874.5	2470	3053.5	29	46	484.5	239	299	279.5	1632	563.5
SOPR0369	**Patient**	1033	1875.5	2477.5	2882	43	42.5	337	330	209.5	239.5	1672	486.5
SOPR0370	**Patient**	10522.5	3252	2966.5	2988	3998	67	1056.5	58	215	904.5	3224	232
SOPR0372	**Patient**	5067	1899	2310	1847	198.5	75.5	653.5	98.5	455.5	170	1288	170.5
SOPR0377	**Patient**	9962	2829	5287.5	4699.5	4020	558.5	5237.5	289	7857	1047	5708	829
SOPR0378	**Patient**	4167	2127.5	2213.5	1112	372	155	676	126	2881.5	1079	2990	368.5
SOPR0380	**Patient**	8317.5	1594	9733.5	2594.5	203.5	21.5	537	61	5046	298	9546	306
SOPR0381	**Patient**	2137	2380.5	1259	2393.5	97	186.5	330.5	90.5	443.5	570.5	1109	221.5
SOPR0384	**Patient**	7247	4965.5	1763	4954	1421	634	1464.5	118.5	3408	1902	1672	183
SOPR0330	Control	5672	3619.5	7062.5	3646	1716.5	727	6629	380	8649	1586	6676	703
SOPR0364	Control	12134	5284.5	2104.5	5688.5	7912.5	3143	558.5	594	6515.5	2240	1	308
SOPR0365	Control	18471.5	2978	3009	2453	4412	215	2116	59	8423.5	574	3090	102
SOPR0371	Control	15704	2587	1794.5	3276.5	1233	136	1310	19.5	8975	1153	1761	102
SOPR0374	Control	842	686.5	2463.5	391.5	86.5	16.5	182	18.5	150	26	488.5	59
SOPR0375	Control	7150.5	2285	3095.5	3847	1092	226	2489	290	4167.5	1307.5	3160	1476
SOPR0376	Control	16661	2997	6052.5	4382.5	1973.5	246.5	4131.5	90	8780	768.5	5972	267.5
SOPR0379	Control	20174.5	6225.5	19389.5	8160	272	12	1805.5	67	6291.5	504.5	12646	768.5
SOPR0382	Control	2226.5	3786.5	1560	4828.5	514.5	1297.5	1579.5	1158	918	1089.5	1498	1143
SOPR0383	Control	4427.5	6301	1833	3330.5	758	943.5	1549	109.5	471.5	1431.5	1574	140

The table shows raw cytokine data prior to normalisation. Effector cytokine responses were assessed following stimulation of the PBMCs with the receptor-specific ligands including LPS (TLR4 stimulant), MDP (NOD2 stimulant) and Pam3CSK4 (TLR1-2 stimulant). Induction of four cytokines including IL-10, IL-1□, IL-6 and TNF-□ were assessed per stimulant, thereby generating 12 assay conditions in total.

Table S4. *NOD2* variants and MDP-mediated immune responses

Subjects	Diagnosis	Cluster	NOD2 Variants	Location (hg19)	Zygosity	Variant type	ACMG individual variant classification	HGMD	NOD2 induction (with MDP)				Combined MDP-induced immune responses
									IL-10	IL-1B	IL-6	TNF-α	
SOPR0339	CD	1	G908R; R708H	rs2066845; rs35285618	Compound het	NS; NS	Strong; Supporting	DFP; Not reported	-0.856	-0.753	-1.242	-0.807	-3.659
SOPR0372	UC	2	V955I	rs5743291	het	NS	Strong	DM?	-0.784	-0.681	-0.885	-0.525	-2.875
SOPR0359	CD	3	V955I	rs5743292	het	NS	Strong	DM?	-0.743	-0.415	-0.656	-0.004	-1.818
SOPR0336	CD	4	R702W	rs2066844	het	NS	Strong	DFP	-0.870	1.195	-0.251	-0.533	-0.458
SOPR0352	CD	4	R702W	rs2066845	het	NS	Strong	DFP	0.792	-0.564	-0.765	-0.519	-1.056
SOPR0370	CD	4	R702W	rs2066846	het	NS	Strong	DFP	0.872	-0.690	-0.659	-0.643	-1.121
SOPR0380	CD	4	R702W	rs2066847	het	NS	Strong	DFP	-0.782	-0.740	-0.950	-0.635	-3.106
SOPR0377	CD	4	R702W	rs2066848	het	NS	Strong	DFP	0.881	-0.151	1.680	0.031	2.441
SOPR0340	UC	1	.	.	.	.	.	.	-0.845	-0.728	-1.221	-0.808	-3.602
SOPR0342	CD	1	.	.	.	.	.	.	-0.857	-0.749	-1.197	-0.804	-3.607
SOPR0345	CD	1	.	.	.	.	.	.	-0.868	-0.761	-1.248	-0.807	-3.684
SOPR0351	CD	1	.	.	.	.	.	.	-0.829	-0.756	-1.221	-0.759	-3.564
SOPR0355	UC	2	.	.	.	.	.	.	-0.845	-0.698	-0.963	-0.371	-2.876
SOPR0367	UC	2	.	.	.	.	.	.	-0.856	-0.744	-1.037	-0.627	-3.264
SOPR0381	UC	2	.	.	.	.	.	.	-0.828	-0.559	-1.066	-0.549	-3.001
SOPR0378	CD	4	.	.	.	.	.	.	-0.708	-0.594	-0.872	-0.445	-2.619
SOPR0368	CD	3	.	.	.	.	.	.	-0.858	-0.713	-0.979	-0.115	-2.666
SOPR0369	UC	3	.	.	.	.	.	.	-0.851	-0.717	-1.062	0.150	-2.480
SOPR0348	CD	4	.	.	.	.	.	.	-0.091	-0.665	-0.029	-0.320	-1.104
SOPR0346	UC	4	.	.	.	.	.	.	-0.184	-0.503	0.058	0.557	-0.072
SOPR0384	UC	4	.	.	.	.	.	.	-0.251	-0.068	-0.431	-0.467	-1.217
SOPR0304	CD	4	.	.	.	.	.	.	-0.606	-0.346	1.251	0.144	0.443
t-tests between patients with *NOD2* variants and those without									0.060	0.170	0.482	0.637	0.203

The table shows NOD2 variants identified in eight patients alongside the MDP-induced immune responses. The p values for t-tests comparing MDP-mediated immune responses between individuals with NOD2 variants and those without are indicated in the bottom row. *[Abbreviations: ACMG- American College of Medical Genetics (individual variant classification for evidence of pathogenicity include: VS- very strong, S- strong, M- moderate, Su- supporting); het- heterozygote; hg19- human genome version 19; HGMD- The Human Gene Mutation Database (DFP- disease associated mutation with functional evidence; ? DM- disease causing mutation, the question mark indicates a degree of uncertainty of the pathogenic potential of the mutation based on updated reports on HGMD)]*

Table S5. Unpaired t-tests comparing cytokine data between each cluster and controls

Receptors	TLR4-induction (with LPS)			TLR1-2 induction (with Pam3CSK4)	
Cytokines	IL-10	IL-1B	TNF-a	IL-10	IL-1B
Cluster 1	0.054	0.006	0.002	0.010	0.004
Cluster 2	0.052	0.046	0.053	0.023	0.042
Cluster 3	0.057	0.152	0.493	0.047	0.052
Cluster 4	.	.	.	.	.

The 5 dysfunctional TLR-mediated responses were compared between each cluster and the control group using unpaired t-tests. The table shows p values obtained after comparing the respective clusters against the control samples. T-tests were not performed in the 'unclustered' group (cluster 4).

Table S6. Assessing differences in immune profiles between CD and UC

		TLR4 induction				NOD2 induction				TLR1-2 induction				Sum
Subjects	Phenotype	IL-10	IL-1B	IL-6	TNF-a	IL-10	IL-1B	IL-6	TNF-a	IL-10	IL-1B	IL-6	TNF-a	
SOPR0304	CD	-1.353	-1.312	-0.038	-1.622	-0.606	-0.346	1.251	0.144	-1.127	-0.692	0.282	-0.343	-5.762
SOPR0336	CD	-1.172	-2.054	-0.627	-1.929	-0.870	1.195	-0.251	-0.533	-1.480	-1.396	-0.662	-0.809	-10.587
SOPR0339	CD	-0.952	-1.503	-0.690	-1.768	-0.856	-0.753	-1.242	-0.807	-1.464	-1.604	-0.846	-1.062	-13.548
SOPR0342	CD	-1.260	-1.872	-0.825	-1.930	-0.857	-0.749	-1.197	-0.804	-1.481	-1.695	-0.863	-1.061	-14.593
SOPR0345	CD	-0.829	-1.558	-0.805	-1.585	-0.868	-0.761	-1.248	-0.807	-1.405	-1.572	-0.842	-1.066	-13.345
SOPR0348	CD	0.632	-1.044	-0.448	-0.776	-0.091	-0.665	-0.029	-0.320	0.775	-0.911	-0.321	-0.382	-3.580
SOPR0351	CD	-1.124	-1.934	-0.840	-1.949	-0.829	-0.756	-1.221	-0.759	-1.426	-1.702	-0.955	-1.055	-14.549
SOPR0352	CD	-0.002	-1.226	-0.803	-1.257	0.792	-0.564	-0.765	-0.519	-0.797	-1.615	-0.811	-1.002	-8.570
SOPR0359	CD	-1.286	-0.794	-0.563	-0.274	-0.743	-0.415	-0.656	-0.004	-1.287	-1.394	-0.661	-0.569	-8.646
SOPR0368	CD	-1.347	-1.057	-0.458	-0.489	-0.858	-0.713	-0.979	-0.115	-1.457	-1.319	-0.566	0.120	-9.238
SOPR0370	CD	0.026	-0.248	-0.362	-0.523	0.872	-0.690	-0.659	-0.643	-1.482	-0.273	-0.127	-0.585	-4.696
SOPR0377	CD	-0.057	-0.497	0.087	0.361	0.881	-0.151	1.680	0.031	0.730	-0.035	0.557	0.685	4.272
SOPR0378	CD	-0.913	-0.909	-0.508	-1.491	-0.708	-0.594	-0.872	-0.445	-0.710	0.018	-0.192	-0.294	-7.618
SOPR0380	CD	-0.300	-1.222	0.948	-0.726	-0.782	-0.740	-0.950	-0.635	-0.083	-1.288	1.614	-0.427	-4.590
SOPR0340	UC	-0.945	-1.180	-0.551	-1.924	-0.845	-0.728	-1.221	-0.808	-1.519	-1.717	-0.935	-1.067	-13.440
SOPR0346	UC	-0.147	-0.331	-0.803	0.582	-0.184	-0.503	0.058	0.557	-0.071	-0.554	-0.344	0.524	-1.215
SOPR0355	UC	-1.427	-1.543	-0.679	-1.499	-0.845	-0.698	-0.963	-0.371	-1.459	-1.312	-0.692	-0.607	-12.095
SOPR0367	UC	-1.289	-1.568	-0.752	-1.241	-0.856	-0.744	-1.037	-0.627	-1.339	-1.319	-0.768	-0.296	-11.836
SOPR0369	UC	-1.377	-1.057	-0.457	-0.577	-0.851	-0.717	-1.062	0.150	-1.483	-1.386	-0.555	-0.043	-9.415
SOPR0372	UC	-0.780	-1.043	-0.489	-1.112	-0.784	-0.681	-0.885	-0.525	-1.412	-1.502	-0.661	-0.715	-10.589
SOPR0381	UC	-1.213	-0.760	-0.693	-0.830	-0.828	-0.559	-1.066	-0.549	-1.416	-0.832	-0.710	-0.607	-10.062
SOPR0384	UC	-0.458	0.758	-0.595	0.492	-0.251	-0.068	-0.431	-0.467	-0.558	1.395	-0.555	-0.689	-1.427
t-tests		0.3584	0.1708	0.2618	0.2983	0.2885	0.5814	0.3718	0.5079	0.4464	0.5623	0.2086	0.5893	0.8112

Normalised cytokine data following stimulation with the receptor-specific ligands including LPS (TLR4 stimulant), MDP (NOD2 stimulant) and Pam3CSK4 (TLR1-2 stimulant). Induction of 4 cytokines including IL-10, IL-1□, IL-6 and TNF-□ were assessed per stimulant, thereby generating 12 assay conditions in total. The column at the extreme right represents summation of the values across all the cells in each row. The bottom row includes p values obtained by applying a 2-tailed student t-test across the values between patients with CD and UC. No significant differences were observed.