**Exome sequencing analysis**

Genomic DNA was extracted from peripheral venous blood or saliva using the salting out method1. 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 annotation3. *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 reported4. Variants were categorised in line with the American College of Medical Genetics (ACGM) guidance to remove ‘benign’ variants and identify ‘pathogenic’ and ‘likely pathogenic’ variants5.

**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.

X1= X- μcontrols/σcontrols

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.

Reference:

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.