# Deleterious genetic variation across the NOD-signaling pathway is associated with reduced *NFKB-*signaling transcription and upregulation of alternative inflammatory transcripts in paediatric inflammatory bowel disease

Short title- Genomic variation impacts NOD-signaling

James J Ashton PhD1,2, Konstantinos Boukas PhD 3, Imogen S Stafford MSci1,4, Guo Cheng PhD 1,5, Rachel Haggarty RN5, Tracy AF Coelho PhD 2, Akshay Batra MD2, Nadeem A Afzal MD2, Anthony P. Williams PhD 3,4, Marta E Polak PhD 6,7, R Mark Beattie FRCPCH2, and Sarah Ennis PhD 1

1. Department of Human Genetics and Genomic Medicine, University of Southampton, Southampton, UK
2. Department of Paediatric Gastroenterology, Southampton Children’s Hospital, Southampton, UK
3. Wessex Investigational Sciences Hub, University of Southampton Faculty of Medicine, Southampton General Hospital, Southampton, UK
4. Institute for Life Sciences, University of Southampton, Southampton, UK
5. NIHR Southampton Biomedical Research Centre, University Hospital Southampton, Southampton, UK
6. Clinical and Experimental Sciences, Sir Henry Wellcome Laboratories, Faculty of Medicine, University of Southampton, Southampton, UK
7. Institute for Life Sciences, University of Southampton, Southampton, UK

Correspondence to-

Professor Sarah Ennis

Human Genetics & Genomic Medicine

University of Southampton

Duthie Building (Mailpoint 808)

Southampton General Hospital

Southampton

SO16 6YD

UK

Tel: +44 (0)23 8079 8614

s.ennis@soton.ac.uk

**Conflicts of interest**

The authors declare no conflicts of interest

**Funding**

JJA was funded during this work by an Action Medical Research training fellowship. JJA is currently funded by an NIHR clinical lectureship and an ESPR post-doctoral grant.

**Data availability**

The RNA sequencing data underlying this article are available in GEO repository and can be accessed with at the accession number GSE153974. The exome sequencing data underlying this article cannot be shared publicly due to ethical considerations. The data will be shared on reasonable request to the corresponding author.

**Author contributions**

JJA, RMB and SE conceived the study. Patients were recruited by JJA and RH. Patient samples were acquired by JJA, RH, TAC, AB, NAA and RMB. Samples were processed and sequenced by JJA, KB, ISS and GC. Analyses were performed by JJA, KB, JD, ISS and GC, under the guidance of MP, AW and SE. BV performed histological analysis for all patients. JJA wrote the manuscript with help from all authors. All authors approved the final manuscript prior to submission.

Abstract word count- 250

Word count- 4905

**Summary**

NOD-signaling pathway transcriptomics, alongside whole exome sequencing demonstrates the

impact of variation across multiple genes/complexes on downstream transcription. Overall, summed

variation across the pathway implicates multiple genetic hits in reduced NFKB transcription, with

upregulation of alternative inflammatory transcripts.

**Abstract**

**Background-**Inflammatory bowel disease (IBD) may arise with inadequate immune response to intestinal bacteria. *NOD2* is an established gene in Crohn’s disease pathogenesis, with deleterious variation associated with reduced *NFKB*-signaling. We hypothesised that deleterious variation across the *NOD2*-signaling pathway impacts on transcription.

**Methods-**Treatment-naïve paediatric IBD patients had ileal biopsies for targeted-autoimmune RNA-sequencing and blood for whole-exome-sequencing collected at diagnostic endoscopy. Utilising GenePy, *a per individual, per gene* score, genes within the NOD-signaling pathway were assigned a quantitative score representing total variant burden. Where multiple genes formed complexes, GenePy scores were summed to create a ‘complex’ score. Normalised transcript expression of 95-genes within this pathway were retrieved. Regression analysis was performed to determine the impact of genomic variation on gene transcription.

**Results-**Thirty-nine patients were included. Limited clustering of patients based on NOD-signaling transcripts was related to underlying genomic variation. Patients harbouring deleterious variation in *NOD2* had reduced *NOD2* (β=-0.702, p=4.3x10-5) and increased *NFKBIA* (β=0.486, p=0.001), reflecting reduced *NFKB*-signal activation. Deleterious variation in the *NOD2-RIPK2* complex was associated with increased *NLRP3* (β=0.8, p=3.1475x10-8) and *TXN* (β=-0.417, p=8.4x10-5) transcription, components of the *NLRP3*-inflammasome. Deleterious variation in the *TAK1-TAB* complex resulted in reduced *MAPK14* transcription (β=-0.677, p=1.7x10-5), a key signal transduction protein in the *NOD2*-signaling cascade and increased *IFNA1* (β=0.479, p=0.001), indicating reduced transcription of *NFKB* activators and alternative interferon transcription in these patients.

**Conclusions**-Data integration identified perturbation of *NOD2*-signaling transcription correlated with genomic variation. A hypoimmune *NFKB*-signaling transcription response was observed. Alternative inflammatory pathways were activated and may represent therapeutic targets in specific patients.

## Key words

IBD; Crohn’s disease; NOD2; WES; transcriptome

## Introduction

Paediatric inflammatory bowel disease has a complex pathogenesis, reflecting genetic and environmental influences on disease aetiology. Contemporary genomics has pointed to the molecular cause of disease being specific to an individual, or family. Despite this, the resulting immune dysfunction frequently converges on common inflammatory pathways such as TNF-α production and Il17-signaling1. Identification of the precise cause of disease within an individual patient is challenging and integration of multi-omic data to assess the impact of genetic variation on gene expression, immune function, and the relationship with intestinal microbiota remains difficult2. Typically, functional assessment of deleterious genomic variation in an individual is through quantification of downstream cytokine levels3.. Stimulation allows assessment of the impact of a mutation on downstream immune function4. Analysis may focus on transfection of cell lines with the mutation of interest, although accurate reflection of *in vivo* response is difficult to achieve when multiple variants are contributing to a complex disease phenotype. Whilst these techniques are effective in Mendelian disease, in polygenic IBD, where multiple genetic variants appear to result in the same phenotype, establishing the effect of multiple mutations is challenging.

Expression quantitative trait loci (eQTLs) have been identified that impact on *NOD2* expression5. Consolidation of genetic variation across a protein complex or pathway, would allow grouping of patients with deleterious variation in several interacting genes. Consolidation can occur through a mathematical model based on a whole gene pathogenicity score, that sums the variation within genes, taking into account the key variation metrics6. This score reflects a patient’s burden of variant pathogenicity within genes of interest.

It is recognised that monogenic disease with highly similar phenotypes can result from deleterious variation in different genes, the occurrence of polygenic conditions, such as typical forms of IBD, are also likely to have contribution from variation in a number of related genes, within an individual7. The NOD-signaling pathway, an innate immune signaling response, is highly implicated in Crohn’s disease pathogenesis, both through genomic analysis and transcriptomic analysis8,9. *NOD2* is the most implicated gene in Crohn’s disease pathogenesis and is a central player in the pathway, acting as an intracellular pattern recognition receptor for bacterial components, specifically muramyl dipeptide (MDP)9, figure 1. Variation within other genes within this pathway, including *XIAP, ATG16L1, CARD9* and *RIPK2,* have been implicated in both monogenic and polygenic forms of IBD10*.*

We hypothesise that genomic variation across the NOD-signaling pathway will have direct impact on transcription of related genes within the pathway. Therefore, we undertook systematic analysis of genes and complexes along the NOD pathway, to identify individual contributions to dysregulation of transcriptional programmes. We assess whether variation in single genes, and across protein complexes, are associated with altered gene expression within the NOD-signaling pathway to identify patterns in immune pathway transcription across a cohort.

## Materials and Methods

Paediatric IBD patients were recruited through from the Paediatric Gastroenterology service at the Southampton Children’s Hospital. Patients were recruited prior to diagnosistic endoscopy and are termed ‘treatment-naïve’ IBD patients. All patients have a terminal ileal biopsy and blood for whole exome sequencing data taken at the time of endocopy. These analyses included patients with a subsequent diagnosis of Crohn’s disease, ulcerative colitis or IBD-unclassified. No patients with monogenic IBD were included. Patient details are described in results. Whilst *NOD2* variation is associated with Crohn’s disease, ulcerative colitis and IBD-unclassified patients were included to ensure a spectrum of genomic mutation burden.

### Whole exome sequencing analysis

Patient DNA was extracted from peripheral venous blood samples collected in EDTA using the salting-out method, or from saliva, as previously described11. DNA concentration was estimated using the Qubit ® 2.0 Fluorometer. Approximately 20ug of each patient DNA was extracted for next generation sequencing.

Raw fastq sequencing data from patients with treatment-naive IBD were processed using our in-house pipeline (<https://github.com/UoS-HGIG>). VerifyBamID was utilised to check the presence of DNA contamination across the cohort12. Alignment was performed against the human reference genome (hg38 assembly) using BWA-mem13 (version 0.7.15). Aligned BAM files were sorted, and duplicate reads were marked using Picard Tools (version 2.9.2). Following GATK v3.814 best practice recommendations15 variants were called using GATK HaplotypeCaller to produce a gVCF file for each sample and later jointly genotyped. HaplotypeCaller default settings were utilised corresponding to variants with a minimum Phred base quality score of 20 being called.

Annotation of this composite file for variant-based analysis applied Annovar v2018Apr16 using default databases refSeq gene transcripts (refGene), deleteriousness scores databases (dbnsfp35c, CADD v.1.5 and DANN), dbSNP147 and the human genetic mutation database (HGMD Pro 2018) flat file16. Variant allele frequencies were sourced through the genome aggregation exome database (gnomAD)17, v2.1.1.

Application of GenePy

Whole exome sequencing data were transformed in a per gene, per individual GenePy score for integration with transcriptomic analysis6. Before GenePy scores were calculated, the joint called cohort VCF was quality controlled by implementing methods from Carson *et al*18. The VCF was then annotated with the databases refGene and gnomAD exome v2.1.117, using Annovar v2018Apr16. The VCF was additionally annotated with deleteriousness metric CADD (v.1.6)19. The GenePy score algorithm was then applied to the annotated cohort VCF (<https://github.com/UoS-HGIG/GenePy>). Demonstration of the contributions of individual variants to GenePy scores can be seen in supplementary data 1.

Normalisation and application of LOEUF score to GenePy

Some genes can accrue very high GenePy scores (due to length or mutability), and when summing GenePy scores across a complex we first needed to normalise the values in order to make each gene comparable. We normalised all GenePy scores between 0 and 1, where 0 represented the lowest GenePy score for that gene across all patients in the analysis, and 1 represented the highest score for that gene across all patients in the analysis.

The LOEUF score is a metric developed as part of the Genome Aggregation Database (gnomAD) which assigns a score to each gene based on the gene’s intolerance to inactivation20. This score can then be applied to determine which genes are able to accrue variation whilst maintaining activity, compared to those in which variation will be highly damaging. Higher LOEUF scores are associated with increased tolerance to variation. Genes within the NOD-pathway which are highly conserved and key to multiple inflammatory processes, such as *RIPK2* or *TAK1*, have very low LOEUF scores, whereas genes in which variation is more commonly seen, such as *NOD2* have higher scores.

We integrated the LOEUF score into the GenePy score in order to upweight the importance of variation in genes predicted to be intolerant to inactivation. LOEUF scores for genes in this analysis can be seen in supplementary data 2. The normalised GenePy score was divided by the respective LOEUF score for each gene.

### Ileal biopsy RNA extraction

Terminal ileal biopsies were obtained during routine endoscopy and frozen at -80 in 1ml of RNAlater (Sigma Aldrich) less than 30 minutes from collection. RNA was extracted from ileal biopsies using Maxwell processing, as previously described21. A single biopsy was used for each patient.

Targeted RNA sequencing

The HTG EdgeSeq Autoimmune Panel was used to measure mRNA expression levels in 2002 genes associated with autoimmune disease, including inflammatory bowel disease. The targeted sequencing was performed as previously described21. Briefly, an RNA sample for each patient was thawed, diluted and loaded onto the HTG EdgeSeq instrument. Following a standardised PCR clean-up procedure, and quantification, each sample underwent a dilution-based normalisation process. Individual libraries were pool at equimolar quantities into the final library for sequencing. The library was denatured and prepared for sequencing in line with Illumina and HTG practice guidelines. Sequencing was performed on the Illumina NextSeq platform.

RNA data processing

Output files from the NextSeq run were converted from BCL format to FASTQ files. Barcode mismatch filter was set to 0 to ensure reads were linked to the correct patient. Gene expression count matrix was constructed for each gene and each patient. These were merged to form a single output file containing all genes and all counts. Quality control was performed as previously described, in line with HTG recommendations21. Gene transcript counts were normalised using quantile normalisation for downstream integration with genomic data. Gene expression using the targeted RNA panel has previously been validated utilising single cell RNA sequencing21. We determined if histological inflammation, regardless of underlying genomic variation, was associated with differential gene expression using DEseq222.

### Key genes and protein complexes within the NOD-signaling pathway

We proposed that summing variation across key protein complexes within the NOD-signaling pathway would enable us to discern the impact of underlying genomic variation on gene expression. We utilised a pre-collated list of 95 genes in the NOD-signaling pathway, produced by HTG as part of their autoimmune panel product23.

The *NOD*-signaling pathway genes were cross-referenced with genes known to be implicated in IBD by either GWAS, or as a monogenic IBD gene23,24. Invariant genes or complexes were not assessed. Single genes within the pathway, and implicated in IBD, were included in the analysis. Molecular complexes in the NOD-signaling pathway, with two or more constituent proteins, containing genes with variation, were included in the analysis. Molecular complexes activated because of multiple inflammatory pathways, including the IKK complex and MAPK complexes, were excluded due to the lack of specificity to *NOD*-signaling.

#### Summing GenePy scores for protein complexes

We summed GenePy scores across the genes in key protein complexes. For each gene within a molecular complex the LOEUF corrected, normalised, GenePy scores for each gene were summed to create a GenePy score for the molecular complex (supplementary data 2). This provides a quantitative score reflecting the cumulative sum of variation within that complex.

### Genomic and transcriptomic integration

We utilised a stepwise linear regression model to determine the impact of genomic variation on transcription. The analyses were conducted to determine whether genomic variation within the NOD-signaling pathway impacted on gene expression within the same pathway. We utilised the GenePy scores for an individual gene, or summed scores for a complex, as the dependant variable. The quantile normalised (QN) expression values for the 95 NOD-signaling genes were entered as independent variables. These data assume an additive model of deleteriousness.

To assess whether patients with higher burden of variation within key *NOD2*-pathway genes and complexes exhibited similar expression patterns across all NOD-signaling genes we performed principal component analysis (PCA) and hierarchical clustering (HC). Patients within the top 10% of GenePy scores for *NOD2, NOD2-RIPK2* and *TAK1-TAB* were annotated within the analyses, the top 10% of patients were chosen as per previous evidence25. These 12 patients were also utilised to determine if high GenePy scores were driving inflammation within the terminal ileal, and therefore altered transcription related to inflammation, rather than altering pathway transcription *per se*.

#### Statistical analysis

Statistical analysis was performed using SPSS (v25, IBM). PCA and HC (euclidean distance, average clustering) were performed using Reveal (HTG, v3.1). Weighted gene coexpression network analysis was performed as previously described using the WGCNA ‘R’ package21,26.

Ethical approval

The study has category A ERGO II ethics approval (30630) and a REC approval from Southampton and South West Hampshire Research Ethics Committee (09/H0504/125). All patients and families provided informed consent at recruitment.

## Results

Thirty-nine patients were included in the analysis, 27 had a diagnosis of Crohn’s disease, 9 had a diagnosis of ulcerative colitis and 3 had a diagnosis of IBD-unclassified. The mean age at diagnosis was 13.2 years (range 2.9-16.8 years). All patients were naïve to any immunomodulating, anti-inflammatory or monoclonal treatment at the time of biopsy acquisition. The core NOD-signaling pathway was interrogated, figure 1. Three genes and three sets of genes encoding protein complexes were selected for analysis based on the criteria described above, table 1. The IRAK-TRAF6 complex acts synergistically with the core NOD2-signaling pathway and was included, in addition to the core NOD2-RIPK2 and TAK1-TAB complexes. The IKK complex, consisting of IKKa, IKKb and NEMO, was excluded as it is the target of multiple activators including TNF-a and IL1 signaling, and is not specific for NOD-dependant bacterial recognition and response27. Similarly, MAPK complexes were excluded due to the large number of MAPK genes that may be included and the lack of specificity to NOD-signaling, supplementary figure 1A. Further data can be seen in supplementary results, including phenotype-*NOD2* genotype correlation (supplementary table 2).

*NOD2* GenePy score is not associated with specific gene expression modules across all autoimmune genes

Given the established role of *NOD2* in Crohn’s disease pathogenesis we hypothesised that subgroup(s) of patients with accumulation of pathogenic *NOD2* variation would be characterised by similar gene expression. To test this hypothesis, we determined gene co-expression modules, identified through WGCNA of all 39 patients, and correlated these modules with *NOD2* GenePy score. Patients were clustered by the similarity of gene expression for all 2002 autoimmune gene transcripts, supplementary figure 2. NOD2 GenePy scores were entered as a continuous variable. Hierarchical clustering did not demonstrate patients with similar *NOD2* GenePy scores in the same clusters.

#### Gene co-expression modules in treatment naïve patients

Expression modules were identified within the AI transcripts, based on the 39 treatment naïve patients. Two large co-expression signatures emerged, the turquoise and blue modules, alongside several smaller co-expression modules. Supplementary figure 3. In order to determine whether specific sets of co-expressed genes were associated genomic variation in *NOD2* we analysed whether the co-expression modules were correlated with *NOD2* GenePy score. None of the 8 modules were significantly correlated with *NOD2*, supplementary figure 4.

Patients harbouring deleterious *NOD2* gene variation have reduced *NOD2* gene expression and increased expression of *NFKB* inhibitor-α

We expected *NOD2* (Figure 1.1) gene variation to impact on downstream gene expression, specifically within the signaling pathway. We examined the effect of variation in *NOD2* through stepwise linear regression, with *NOD2* GenePy score as the dependant variable and all 95 gene transcript levels as the independent variables. Increased *NOD2* GenePy scores, reflecting increased deleterious variation, was associated with a decrease in *NOD2* transcripts (figure 2A) and increased expression of *NFKBIA*, encoding a key inhibitory protein preventing NFKB signaling (table 2). The effect appears to reflect a decrease in NFKB signaling as a result of deleterious *NOD2* variation (supplementary figure 1B). *CCL5*, a T-cell chemokine, was also downregulated in patients with high variant deleteriousness in the *NOD2* gene.

#### 1007fs *NOD2* variant impacts on transcription but is not the only driver of reduced expression

*NOD2* harbours a protein truncating variant, 1007fs, commonly seen in Crohn’s disease patients. We assessed whether this specific nonsense variant within *NOD2* was driving the inverse relationship between GenePy score and transcript number. Four of the 39 patients were heterozygous for the 1007fs *NOD2* variant. A T-test demonstrated there was no significant difference in the *NOD2* expression level between those with the 1007fs variant (mean QN transcripts 42.9) and those without (mean QN transcripts 59.5), p= 0.12, although only 4 patients with the frameshift variant were seen in the patient group, figure 2A. No other protein truncating variants within *NOD2* were identified in the 39 patients.

We assessed if *NOD2* transcription was impacted by any of the three common variants associated with CD- 1007fs (4 patients = heterozygote), R702W (1 patient = heterozygote) and G908R (5 patient = heterozygote). Comparing these 10 patients to the remaining 29 patients, who did not harbour any of these variants demonstrated a significantly reduced NOD2 transcript count, 45.9 vs 61.8, respectively, p=0.049.

### Deleterious variation in *CARD9* is associated with decreased expression of *PYCARD and TANK*

*CARD9* interacts directly with *NOD2* (Figure 1.3), resulting in activation of downstream pro-inflammatory signaling through MAPK activation alongside activation of the *NLRP3* inflammasome. It also functions independently as a signal transduction complex alongside *BCL10* and *MALT1,* largely in response to fungal infection*,* which then activates the IKK complex triggering NFKB activation, or through *PYCARD*, activating the NLRP3 inflammasome28. Regression analysis, with *CARD9* GenePy score as the dependant variable demonstrated increased deleterious variation in *CARD9* was associated with a decrease in *PYCARD*, an upstream activator of the *NLRP3* inflammasome, figure 2B, table 2. Additionally, we observe an increase in increased transcription of *TRAF2*, and a decrease in *TANK,* the TRAF Family Member Associated NFKB Activator. *TANK* directly inhibits *TRAF2* and these data reflect a potential reduced ability to inhibit downstream NFKB signaling, through *TRAF2* and *IKK,* within the NOD-signaling cascade29.

Deleterious variation in the *NOD2-RIPK2* complex is associated with increased expression of *BIRC2, TXN and NLRP3*

Following activation by MDP, *NOD2* forms a complex with *RIPK2* (Figure 1.3). *XIAP, BIRC2, BIRC3* and *ITCH* all positively associate with the complex promoting downstream *RIPK2* kinase activity. Regression analysis, using the summed *NOD2-RIPK2* complex as the dependant variable, revealed increased deleterious variation within the complex was related to an increased expression of the *NLRP3* and *TXN*, table 2. Both of these genes are involved in the pro-inflammatory *NLRP3* inflammasome, however we also observe a decrease in PYCARD expression, encoding a key protein in the inflammasome activation pathway (Figure 1.6). *BIRC2* was highly significantly upregulated in patients harbouring deleterious variation in this complex, which includes *BIRC2*. The impact can be seen in supplementary figure 1C.

Genomic variation within the TAK1-TAB complex leads to reduced *MAPK14* expression

The *TAK1-TAB* complex, including *TAK1*, *TAB2* and *TAB3*, is a key signal transducer, both from the *NOD2-RIPK2* complex and toll-like receptors (TLRs) (Figure 1.4). Summed GenePy scores for the TAK1-TAB complex were used as the dependant variable. Expression of *MAPK14* and *BIRC3* was reduced in the presence of deleterious variation within the *TAK1-TAB* complex. There was also increased expression of *IFNA1*, reflecting activation of a type 1 interferon response pathway figure 2C and table 2. *MAPK14* autophosphorylates in the presence of the TAK1-TAB complex, leading to downstream activation of pro-inflammatory and anti-microbial gene expression, supplementary figure 1D.

### Deleterious variation in *ATG16L1* increases expression of *IKBKB*

*ATG16L1* encodes for a protein key in autophagy pathways. Activated *NOD2* works synergistically with *ATG16L1* to promote autophagy (Figure 1.7). Variation within *ATG16L1* is an established risk for Crohn’s disease development. Regression analysis, utilising *ATG16L1* GenePy score as the dependant variable demonstrated increased *ATG16L1* GenePy score was associated with an increase in *IKBKB*, an activator of NFKB-signaling, table 2. Autophagy represents a different pathway to core *NOD2*-signaling and variation in *ATG16L1* may be pro-inflammatory via NKFB signaling, whilst reducing autophagy.

Variation in the IRAK-TRAF6 complex, within the toll-like receptor (TLR) signaling pathway, results in decreased expression of the NFKB activating protein *IKBKG (NEMO)*

The IRAK-TRAF6 complex (*IRAK1, IRAK2, IRAK4, TRAF6, MYD88)* plays a key role in signal transduction from TLRs through interaction with *MYD88*. It acts synergistically with *NOD2* activation and one result of IRAK-TRAF6 activation is downstream activation of the TAK1-TAB and subsequently activation of NFKB signaling. Variation in the IRAK-TRAF6 complex was associated with a decrease in a single gene transcript, *IKBKG*, a component of the NFKB activating complex, figure 2D.

Patients with deleterious variants in *NOD2, RIPK2* and *TAK1-TAB* have variable NOD-signaling expression clustering

Patients with high burden of variation (top 10% of GenePy scores) in the central *NOD2* signaling complexes were assessed for similarity of all NOD-signaling gene expression. PCA revealed variable clustering, figure 3A. The four patients with highest burden variation in the *NOD2-RIPK2* complex did cluster together, along with a patient with high mutation burden in *NOD2* and a patient with high burden in the *TAK1-TAB* complex.

Using HC we were able to discern that several of the patients with high *NOD2-RIPK* or *NOD2* mutation were characterised by increased expression of *CXCL8* (encoding for the IL-8 neutrophil chemokine) and low expression of *IFNA1,* figure 3B. Interestingly, whilst deleterious variation in the *NOD2-RIPK* complex appears to negatively impact on downstream *NFKB* signaling there is still an overall inflammatory response, mediated through alternative inflammatory pathways. Further analysis of the patients with high genomic burden in *NOD2*, *NOD2-RIPK2* and *TAK1-TAB* revealed that of the 12 patients with the highest burden, 7 had histological inflammation whilst 5 were non-inflamed. This indicates that transcription differences noted through regression are not-solely driven by inflammation status of the biopsy.

Differential gene expression within the NOD-signaling pathway is observed between inflamed and non-inflamed tissue

Within the NOD-signaling pathway there were 17 genes with significant differential expression (corrected p value <0.05), 8 were upregulated in inflamed biopsies and 9 were downregulated. Further analysis of the patients with the top 10% of genomic burden in *NOD2*, *NOD2-RIPK2* and *TAK1-TAB* revealed that of the 12 patients with the highest burden, 7 had histological inflammation whilst 5 were non-inflamed. This indicates that transcription differences noted through regression are not-solely driven by inflammation status of the biopsy but by altered transcription related to underlying genomic differences.

## Discussion

The impact of genomic variation across key genes, and complexes, within the NOD-signaling pathway appears to be associated with a hypoinflammatory response, with reduced activation (or increased inhibition) of NFKB-signaling or reduced upstream activation of pro-inflammatory signaling. Variation in *NOD2*, and directly related complexes, appears to act synergistically to reduce *NOD2* transcription, whilst simultaneously increasing transcription of alternative inflammatory pathways including the *NLRP3* inflammasome and interferon signaling. We identify variation across the *TAK1-TAB* complex directly correlated with reduced *MAPK14* transcription. Variation in *NOD2*-synergistic activators of autophagy (*ATG16L1*), or NOD-signaling (*CARD9*), impact on downstream transcription, either increasing or decreasing NFKB signaling, respectively.

Previous data has pointed to a hypoimmune response in both Crohn’s disease patients and murine models harbouring deleterious *NOD2* variants3,9. *NOD2* variants are thought to be loss-of-function, leading to impaired *NOD2* activation30. Studies detailing the direct effect of deleterious *NOD2* variants repeatedly identify a decrease in pro-inflammatory cytokine response after MDP stimulation in peripheral blood mononuclear cells, specifically reduced NFKB production9. Overall, the mechanism whereby *NOD2* variants increase susceptibility to Crohn’s disease appears to be through impaired bacterial recognition/response leading to reduced bacterial clearance and increased chronic inflammation through non-*NOD2* proinflammatory pathways31. In addition to *NOD2*, several additional risk susceptibility genes, or monogenic IBD genes, lie within the NOD-signaling pathway including *XIAP, CARD9* and *TAB2*1. Loss-of-function variants within these genes are associated with severe monogenic forms of Crohn’s-like IBD (*XIAP, CARD9*), or increased risk of ‘classical’ Crohn’s disease (*TAB2*). Several studies have described the function of these genes in downstream NFKB signaling, including variants in XIAP32 and CARD933 leading to reduced NFKB production. Whilst the mechanism by which these genes lead to disease may not have the evidence base seen with *NOD2*, it appears that a hypo-inflammatory response is implicated, potentially alongside activation of additional aberrant pathways. It was important to include patients with ulcerative colitis is this analysis as UC patients should have no impact of *NOD2* variation on transcription, and it is therefore vital to include these patients to demonstrate a full spectrum of genomic variation and transcription levels. In addition, variation in the wider *NOD2*-signaling pathway may impact on ulcerative colitis patients and moving IBD towards a precise molecular diagnosis would equate to patient’s being categorised by molecular perturbation, rather than endoscopic/histological disease location.

Within this study we hypothesised that variation across the NOD-signaling pathway, with a focus on *NOD2*-signaling, would result in transcription level defects associated with a hypoinflammatory response. Previous data has inferred that disruption at any step on the *NOD2*-signaling cascade will result in decreased downstream NFKB or MAPK activation, although direct impact of genetic variation at each step has not been previously identified34. Through single gene, and whole complex, deleteriousness scoring we identify a consistent pattern of defects associated with decreased transcription of downstream *NFKB* or *MAPK* genes. Importantly, by summing deleteriousness across a complex we were able to observe cohort-level effects that may be missed if assessing a single variant or a single gene within an individual patient. As individual variant effects on gene transcription are likely to be very mild, or private to an individual, the ability to sum the effects of interacting genes allows a statistical association to emerge across a cohort. It is clear that for most patients with IBD, the effect of multiple genomic variants leads to disease, rather than a strong effect from a single gene25.

At an individual gene level, we reveal a striking decrease in *NOD2* transcripts associated with increased *NOD2* deleteriousness. This is not only driven by patients harbouring the nonsense 1007fs variant, with only four of the 39 patients being heterozygote, and no patients being homozygote for this protein truncating variant. It is not possible to determine whether several of the more common variants, harboured by patients with low *NOD2* transcript levels are in linkage disequilibrium with non-coding variants in the promotor region of *NOD2*. *ATG16L1* synergistically acts with *NOD2* to promote antibacterial autophagy and also has a role in negative regulation of proinflammatory MAPK and NFKB activation cascades35. Our results demonstrate the direct impact of this additional *ATG16L1* role, with *ATG16L1* variation leading to an increase in the NFKB activation transcript *IKBKB.* Where deleterious variation in the *NOD2* canonical pathway appears to lead to reduced NFKB/MAPK activation of downstream inflammatory signaling, variation in *ATG16L1* may lead to increased inflammation through impaired autophagy, or directly through the inability to negatively regulate *NOD2* activity. Interestingly, recent ileal transcriptomic data integrating with *NOD2* genotyping of the patients has identified an activated and dysregulated fibroblast cell signature in those with higher mutation burden36. Within this paper the authors identify an apparent *STAT3* driven regulation of downstream transcription, specifically related to mutant *NOD2* alleles.

We identify summed variation in the TAK1-TAB complex directly associated with reduced *MAPK14* transcription. Alongside this we observe an increase in gene transcription associated with alternative inflammatory pathways (*TXN* and *NLRP3*) seen with *NOD2-RIPK2* complex variation. Previous data has indicated a key role for *TAK1* in MDP-stimulated NOD2-signaling, with absence of TAK1 completely admonishing downstream NFKB and MAPK signaling37. Variation in the toll-like receptor transduction complex, *IRAK-TRAF6*, was associated with a decrease in IKBKG transcription, a potent activator of NFKB signaling. This membrane receptor-triggered pathway, acts in parallel to intracellular *NOD2* signaling, also leads to NFKB activation in response to bacterial recognition and response. These data imply that variation across this related complex also impairs antimicrobial response. The activation of the *NLRP3* inflammasome has been reported as the key inflammatory pathway leading to colitis in *NOD2* knockout murine models38. Here we present data suggesting that transcripts in the pathway are upregulated in IBD patients with genomic variation in a number of *NOD2*-signaling genes. In contrast, we identify variation in *CARD9* resulting in a decrease in *PYCARD*, a key signal transduction protein in the *NLRP3* inflammasome, also leading to caspase-1 activation and IL-1B/IL-18 processing. It is possible that reduced *NLRP3* activity is the key hypoimmune response for a subset of individuals, as observed in murine models, with concurrent increase in *NFKB*-signaling in these patients39. Previous data has indicated both a protective and antagonistic role for *NLRP3*, with the potential for related molecular variation within related pathways leading to activation or suppression in some patients, with a subsequent hyper, or hypoinflammatory, response leading to disease40.

We hypothesise that for many patients, IBD appears to arise due to multiple ‘hits’ across complexes/genes contributing to impairment of inflammatory pathways. These patients then fail to clear bacteria allowing invasion and chronic inflammation to develop due to perturbation of a number of innate immune responses. The precise immune impairment within an individual is likely to lead to commonality between subgroups of patients, with many Crohn’s disease patients having disease attributable to impaired *NOD2*-signaling4. This provides an opportunity to target novel molecules with therapeutics, including the potential for *NOD2* activators. Primarily these data provide the framework for molecular profiling of patients. Long-term follow-up data for these individuals will allow integration of treatment response and disease course and may allow prediction of outcomes at the point of diagnosis.

This study has several strengths. We provide a level of validation of genomic findings without the need for time-consuming functional assays, although caution should be exercised in interpretation of these results. Through use of targeted sequencing, we enable identification of lowly expressed transcripts, which are key in many of these analyses. Supplementary results indicate that *NOD2* does not drive a specific gene expression signature across all genes targeted by the AI panel, with a more select analysis revealing transcription differences. Additionally, whilst inflamed and non-inflamed tissue demonstrated a degree of differential gene expression the main analyses included patients with high variant burden but no ileal inflammation, demonstrating that genetic variation may be driving altered gene transcription independent of inflammation status. Additionally, we have previously determined that ileal transcriptomic profiles are not driven by inflammatory status21.

We acknowledge several limitations of this study. The curated list of *NOD*-signaling genes could include either fewer, or more, genes. This gene list was specifically designed for use with the targeted autoimmune gene panel, which itself has the potential to limit wider findings related to WGCNA. We chose to not expand the list to retain statistical power to determine the impact of genomic variation on the most closely related transcripts. In addition, the inability of WES to capture promotor/regulatory variation and the dependence of GenePy on *in silico* deleteriousness metrics and inability to determine compound heterozygosity. Our group has previously studied cytokine-induction in the context of *NOD2* variation, focused on MDP-induced responses3. Within this study we did not study cytokine levels in relation to NOD-signaling gene variation. Based on previous data we hypothesise that these effector cytokines (IL-18, IL-1, TNF) are the common endpoint of numerous immune pathways, and the hypoimmune response seen in specific IBD-related pathways, such as NOD-signaling, is overwhelmed by alternative activation leading to high levels of effector cytokines. Finally, whilst gene transcript number is frequently related to immune function, there is also post-translational regulation of signaling molecules, meaning inference of functional consequences of altered transcription must be cautious.

Conclusion

These data demonstrate a pathway-wide a of genomic variation in NOD-signaling genes, associated with reduced proinflammatory gene transcription within this pathway. Integration of genomic and transcriptomic data allows for statistical association of genomic variation with downstream transcription. We observe variation at each stage of the *NOD2*-signaling pathway resulting in broadly reduced *NFKB* signaling, with frequent upregulation of other inflammatory genes including *NLRP3* and interferons. Expanding these analyses to additional pathways implicated in IBD may allow for precise ‘immuno-typing’ of patients, identifying defects in specific immune pathways and paving the way for personalised therapy.

**References**

1. Graham DB, Xavier RJ. Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature*. 2020;578:527–539. Available at: http://www.nature.com/articles/s41586-020-2025-2 [Accessed March 1, 2020].

2. Ashton JJ, Mossotto E, Ennis S, et al. Personalising medicine in inflammatory bowel disease-current and future perspectives. *Translational pediatrics*. 2019;8:56–69.

3. Coelho T, Mossotto E, Gao Y, et al. Immunological Profiling of Paediatric Inflammatory Bowel Disease Using Unsupervised Machine Learning. *Journal of Pediatric Gastroenterology & Nutrition*. 2020;70:833–840. Available at: https://journals.lww.com/10.1097/MPG.0000000000002719 [Accessed July 15, 2020].

4. Coelho T, Mossotto E, Gao Y, et al. Immunological Profiling of Paediatric Inflammatory Bowel Disease Using Unsupervised Machine Learning. *Journal of Pediatric Gastroenterology and Nutrition*. 2020;70:833–840.

5. Naranbhai V, Fairfax BP, Makino S, et al. Genomic modulators of gene expression in human neutrophils. *Nature Communications*. 2015;6:1–13.

6. Mossotto E, Ashton JJ, O’Gorman L, et al. GenePy - a score for estimating gene pathogenicity in individuals using next-generation sequencing data. *BMC Bioinformatics*. 2019;20:254. Available at: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2877-3 [Accessed May 24, 2019].

7. Uhlig HH, Muise AM. Clinical Genomics in Inflammatory Bowel Disease. *Trends in Genetics*. 2017;33:629–641.

8. Andreoletti G, Shakhnovich V, Christenson K, et al. Exome Analysis of Rare and Common Variants within the NOD Signaling Pathway. *Scientific reports*. 2017;7:46454.

9. Caruso R, Warner N, Inohara N, et al. NOD1 and NOD2: signaling, host defense, and inflammatory disease. *Immunity*. 2014;41:898–908.

10. Crowley E, Warner N, Pan J, et al. Prevalence and Clinical Features of Inflammatory Bowel Diseases Associated with Monogenic Variants, Identified by Whole-exome Sequencing in 1000 Children at a Single Center. *Gastroenterology*. 2020;158:2208–2220.

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

12. Jun G, Flickinger M, Hetrick KN, et al. Detecting and Estimating Contamination of Human DNA Samples in Sequencing and Array-Based Genotype Data. *The American Journal of Human Genetics*. 2012;91:839–848.

13. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics*. 2013:1303.3997.

14. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*. 2010;20:1297–303.

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

16. Stenson PD, Mort M, Ball E V., et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human Genetics*. 2017;136:665–677.

17. Lek M, Karczewski KJ, Minikel E V., et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291.

18. Carson AR, Smith EN, Matsui H, et al. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics*. 2014;15:125.

19. Shihab HA, Rogers MF, Gough J, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*. 2015;31:1536–1543.

20. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. 2019:531210.

21. Ashton JJ, Boukas K, Davies J, et al. Ileal transcriptomic analysis in paediatric Crohn’s disease reveals *IL17-* and *NOD-* signalling expression signatures in treatment-naïve patients and identifies epithelial cells driving differentially expressed genes. *Journal of Crohn’s and Colitis*. 2020. Available at: https://academic.oup.com/ecco-jcc/advance-article/doi/10.1093/ecco-jcc/jjaa236/6000068 [Accessed November 27, 2020].

22. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology 2010 11:10*. 2010;11:1–12. Available at: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-10-r106 [Accessed August 2, 2021].

23. Anon. *Gene List HTG EdgeSeq Autoimmune Panel*. 2020.

24. Ashton JJ, Mossotto E, Stafford IS, et al. Genetic Sequencing of Pediatric Patients Identifies Mutations in Monogenic Inflammatory Bowel Disease Genes that Translate to Distinct Clinical Phenotypes. *Clinical and Translational Gastroenterology*. 2020;11:e00129.

25. Ashton JJ, Mossotto E, Stafford IS, et al. Genetic sequencing of paediatric patients identifies mutations in monogenic inflammatory bowel disease genes that translate to distinct clinical phenotypes. *Clinical and Translational Gastroenterology*. 2020;11:e00129.

26. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.

27. Israël A. The IKK complex, a central regulator of NF-kappaB activation. *Cold Spring Harbor perspectives in biology*. 2010;2.

28. Malik A, Sharma D, Malireddi RKS, et al. SYK-CARD9 Signaling Axis Promotes Gut Fungi-Mediated Inflammasome Activation to Restrict Colitis and Colon Cancer. *Immunity*. 2018;49:515-530.e5.

29. Xie P. TRAF molecules in cell signaling and in human diseases. *Journal of Molecular Signaling*. 2013;8:7.

30. Bonen DK, Ogura Y, Nicolae DL, et al. Crohn’s disease-associated NOD2 variants share a signaling defect in response to lipopolysaccharide and peptidoglycan. *Gastroenterology*. 2003;124:140–146.

31. Heel DA Van, Ghosh S, Butler M, et al. Muramyl dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn’s disease. *Lancet*. 2005;365:1794–1796.

32. Parackova Z, Milota T, Vrabcova P, et al. Novel XIAP mutation causing enhanced spontaneous apoptosis and disturbed NOD2 signalling in a patient with atypical adult-onset Crohn’s disease. *Cell Death and Disease*. 2020;11:1–11.

33. Bruyne M De, Hoste L, Bogaert DJ, et al. A CARD9 Founder Mutation Disrupts NF-κB Signaling by Inhibiting BCL10 and MALT1 Recruitment and Signalosome Formation. *Frontiers in Immunology*. 2018;9:2366.

34. Warner N, Burberry A, Franchi L, et al. A genome-wide siRNA screen reveals positive and negative regulators of the NOD2 and NF-κB signaling pathways. *Science Signaling*. 2013;6:rs3–rs3.

35. Sorbara MT, Ellison LK, Ramjeet M, et al. The protein ATG16L1 suppresses inflammatory cytokines induced by the intracellular sensors Nod1 and Nod2 in an autophagy-independent manner. *Immunity*. 2013;39:858–873.

36. Nayar S, Morrison JK, Giri M, et al. A myeloid–stromal niche and gp130 rescue in NOD2-driven Crohn’s disease. *Nature*. 2021:1–9. Available at: http://www.nature.com/articles/s41586-021-03484-5 [Accessed April 8, 2021].

37. Kim JY, Omori E, Matsumoto K, et al. TAK1 is a central mediator of NOD2 signaling in epidermal cells. *Journal of Biological Chemistry*. 2008;283:137–144.

38. Umiker B, Lee HH, Cope J, et al. The NLRP3 inflammasome mediates DSS-induced intestinal inflammation in Nod2 knockout mice. *Innate Immunity*. 2019;25:132–143.

39. Hirota SA, Ng J, Lueng A, et al. NLRP3 inflammasome plays a key role in the regulation of intestinal homeostasis. *Inflammatory Bowel Diseases*. 2011;17:1359–1372.

40. Zhen Y, Zhang H. NLRP3 Inflammasome and Inflammatory Bowel Disease. *Frontiers in Immunology*. 2019;0:276.

**Tables and figures**

**Table 1-** Genes and complexes to be entered as dependant variables in regression analysis, and the constituent proteins (genes). All gene’s GenePy scores are scaled to between 0-1 and corrected by LOEUF score prior to being summed to form the ‘complex’s GenePy score’

|  |  |
| --- | --- |
| **Gene or complex**  | **Proteins comprising complex** |
| *NOD2* | *NOD2* |
| *ATG16L1* | *ATG16L1* |
| *CARD9* | *CARD9* |
| NOD2-RIPK2 complex | *RIPK2, NOD2, XIAP, BIRC2, BIRC3, ITCH* |
| TAK1-TAB complex | *TAK1, TAB2, TAB3* |
| IRAK-TRAF6 complex  | *IRAK1, IRAK2, IRAK4, TRAF6, MYD88* |

**Table 2-** Linear regression results of genes and complexes and the impact on NOD-signaling transcription.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dependant variable-Gene/Complex GenePy score** | **Independent variable- Quantile normalised gene expression** | **Beta coefficient** | **R squared value for linear regression model** | **P value** |
| *NOD2* | *NOD2* | -0.702 | 0.460 | 0.000043 |
| *NFKBIA* | 0.486 | 0.001 |
| *CCL5* | -0.414 | 0.008 |
| *ATG16L1* | *IKBKB* | 0.504 | 0.403 | 0.001 |
| *IRAK4* | -0.498 | 0.001 |
| *CYBA* | -0.288 | 0.041 |
| *CARD9* | *PYCARD* | -0.559 | 0.510 | 0.0007 |
| *IFNB1* | 0.333 | 0.006 |
| *TRAF2* | 0.371 | 0.003 |
| *TANK* | -0.294 | 0.025 |
| *NOD2-RIPK2 complex (RIPK2, NOD2, XIAP, BIRC2, BIRC3, ITCH)* | *BIRC2* | 0.800 | 0.687 | 3.1475 x 10-8 |
| *TXN* | 0.417 | 0.000084 |
| *NLRP3* | 0.245 | 0.014 |
| *PYCARD* | -0.278 | 0.014 |
| *IRAK4* | -0.345 | 0.001 |
| *UBA52* | -0.224 | 0.033 |
| *TAK1-TAB complex (TAK1, TAB2, TAB3)* | *MAPK14* | -0.677 | 0.438 | 0.000017 |
| *IFNA1* | 0.479 | 0.001 |
| *BIRC3* | -0.375 | 0.008 |
| *IRAK-TRAF6 complex (IRAK1, IRAK2, IRAK4, TRAF6, MYD88)* | *IKBKG* | -0.508 | 0.258 | 0.001 |

**Figure 1-** Schematic representation of*NOD2*-signaling cascade and directly related inflammatory signaling pathways (Adapted from “Detection of Bacterial Peptidoglycan by NOD Receptors”, by BioRender.com (2021). Retrieved from https://app.biorender.com/biorender-templates).

**Figure 2A-** Relationship between quantile normalised *NOD2* transcript levels and *NOD2* GenePy score. Four patients harbouring the 1007fs variant are seen in orange, **2B-** Relationship between quantile normalised *PYCARD* transcript levels and *CARD9* complex GenePy score**, 2C-** Relationship between quantile normalised *MAPK14* transcript levels and TAK1-TAB complex GenePy score**, 2D-** Relationship between quantile normalised *IKBKG* transcript levels and IRAK-TRAF6 complex GenePy score

**Figure 3A-** Principal component analysis using quantile normalised data from 95 NOD-signaling genes demonstrating clustering of patients with the top 10% (4 patients) of deleterious genetic variation within *NOD2-RIPK2*, but less defined grouping for *NOD2* and *TAK1-TAB,* **3B-** Heatmap analysis using quantile normalised data from 95 NOD-signaling genes and average linkage clustering demonstrates highly variable clustering of patients. Two patients with top 10% of deleterious genetic variation in *NOD2* and one with top 10% in *NOD2-RIPK2* do cluster (boxed area), with high transcription of *CXCL8* (IL-8) and low expression of *TFNA1.*

**Supplementary data-**

Supplementary data 2 can be found at the following link- <https://www.dropbox.com/s/mr7xwgow3q5macc/Supplementary%20data%202-%20Genomic%20and%20RNA%20integration%20GenePy%20scores%20QN%20transcripts.xlsx?dl=0>

**Supplementary figure 1A-** *NOD2*-signaling cascade and directly related inflammatory signaling pathways. Kinases, demonstrated in yellow, are S/T kinase domains formed as part of that protein complex, **Supplementary 1B-** Impact of deleterious *NOD2* genetic variation within the *NOD2*-signaling cascade. Decreased transcription of *NOD2* and *NFKBIA*, inferring reduced downstream pro-inflammatory signaling. Blunted response to MDP from bacterial breakdown may result in poorer bacterial clearance and activation of alternative chronic inflammatory pathways**, Supplementary 1C-** Impact of deleterious *NOD2-RIPK2* complex (RIPK2, NOD2, XIAP, BIRC2, BIRC3, ITCH) genetic variation within the *NOD2*-signaling cascade. Increased transcription of *NLRP3*, inferring activation of alternative inflammatory pathways due to inability of signal transduction through the major *NOD2*-signaling cascade. Lack of synergistic activity between *NOD2-RIPK2* complex is associated with decreased *IRAK4*, part of the Toll-like receptor signal transduction pathway**, Supplementary 1D- -** Impact of deleterious *TAK1-TAB* complex (TAK1, TAB2, TAB3) genetic variation within the *NOD2*-signaling cascade. Reduction in *MAPK14*, a principle signal transduction kinase is associated with deleterious variation in the *TAK1-TAB* complex. Alternative type 1 interferon pathway activation is observed, with increased *IFNA1* transcription through *MAVs* and *IRF3* signaling.

**Supplementary figure 2-** Clustering of patients by WGCNA utilising all 2002 autoimmune gene transcripts. Annotation of patients with NOD2 GenePy score did not reveal clusters of patients with similar gene expression also harbouring similar NOD2 deleteriousness.

**Supplementary figure 3-** Gene coexpression modules determined using WGCNA performed on 39 treatment naïve IBD patients. Turquoise and blue modules represent large clusters of similarly expressed patterns of genes across the cohort.

**Supplementary figure 4-** Correlation coefficient values (p values) between gene expression modules and NOD2 GenePy scores